Instytut Fizyki Polskiej Akademii Nauk
Zespół Fizyki Biologicznej

Joanna Ida Sułkowska

ZWIJANIE I ROZWIJANIE BIAŁEK W MODELACH GRUBOZIARNISTYCH

rozprawa doktorska
wykonana pod kierunkiem
prof. dra hab. Marka Cieplaka

Warszawa, 2007

Wstęga Möbiusa (M. C. Escher)

Dla Piotra
i całej mojej Rodziny

Lista publikacji

Niniejsza praca powstała na podstawie wyników opisanych także w następujących publikacjach, do których będędziemy odwoływałać się w tekście poprzez ich rzymską numerację. Prace [X] oraz [XI] powstały także w trakacie doktoratu, aczkolwiek nie są one ściśle związane z zagadnieniami omawianymi w niniejszej rozprawie.

I. J. I. Kwiecińska* & M. Cieplak, „Chirality and protein folding", *J. Phys. Cond. Mat.* **17**, S1565-S1580 (2005) (praca zaliczona do jednej z 50 najczęściej czytanych prac w *J. Phys. Cond. Mat.* w roku 2005).

II. M. Cieplak & J. I. Sułkowska, „Thermal unfolding of proteins", *J. Chem. Phys.*, **123**, 194908 (2005).

III. M. Cieplak & J. I. Sułkowska, „Rozciąganie molekuł białek — porównanie ich własności mechanicznych", *Kosmos*, **55**, 4, (2006).

IV. J. I. Sułkowska & M. Cieplak, „Mechanical Stretching of proteins — A theoretical survey of the Protein Data Bank", *J. Phys. Cond. Mat.* **19** 283201 (2007) — artykuł przeglądowy.

V. J. I. Sułkowska, A. Kloczkowski, T. Z. Sen, M. Cieplak & R. L. Jernigan, „Predicting the Order in Which Contacts Are Broken during Single Molecule Protein Stretching Experiments", *Proteins* (2007) — w druku.

VI. J. I. Sułkowska & M. Cieplak, „Stretching to understand protein - A survey of the Protein Data Bank", *Biophys. J.* (2007) — w druku.

VII. J. I. Sułkowska, P. Sułkowski, P. Szymczak & M. Cieplak, „Tightening of knots in proteins", w recenzji w *Phys. Rev. Lett.*, [q-bio/0706.2380].

VIII. J. I. Sułkowska & M. Cieplak, „Selection of optimal variants of Go-like models of proteins through studies of stretching", — wysłane do publikacji.

IX. J. I. Sułkowska, B. Witkowski & M. Cieplak, „Bio-molecule stretching database" — w przygotowaniu.

X. M. Atakhorrami, J. I. Kwiecińska*, K. M Addas, G. Koenderink, A. J. Levine, J. X. Tang, & C. F Schmidt, „Cross talk, trapping effects and solvent inertia in one- and two-particle microrheology", *Biophys. J.*, **86** (1): 487A-487A Part 2 Suppl. S (2004).

XI. M. Atakhorrami, J. I. Sułkowska, K. M Addas, G. Koenderink, , J. X Tang, A. J. Levine, F. C. MacKintosh & C. F. Schmidt, „Correlated fluctuation of microparticles in viscoelastic solutions: quantitative measurement of materials properties by microrheology in the presence of optical traps", *Phys. Rev. E*, **73**, 061501 (2006).

* nazwisko panieńskie J. I. Sułkowskiej

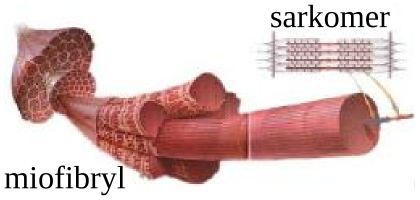
Spis treści

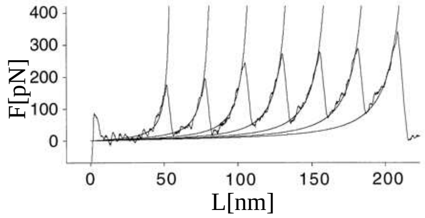
Wstęp

Białka są fascynującymi obiektami molekularnymi. Stanowią one podstawowy budulec całej przyrody ożywionej, składając się zaledwie z 20 aminokwasów o różnych kształtach i własnościach. Większość białek przyjmuje niepowtarzalną trójwymiarową strukturę, nazywaną strukturą natywną, konieczną do tego, aby mogły pełnić właściwą funkcję biologiczną. Ponadto, dla dużej liczby białek, stan natywny jest zaledwie punktem wyjścia do dalszych przemian strukturalnych pod wpływem sił mechanicznych. Dla tych białek elementem kluczowym dla wykonywanej funkcji biologicznej jest ich mechaniczna stabilność, czyli opór generowany w odpowiedzi na np. mechaniczne rozciąganie.

Procesy mechanicznego rozciągania występują niemalże na każdym etapie cyklu komórkowego. Do ważniejszych tego typu procesów należą: translokacja przez membranę (np. do i z mitochondriów) [1, 2], degradacja białek w proteozomie przez proteazy ATP [2], rozciąganie mięśni, w którym bierze udział gigantyczne białko tytyna [3, 4], procesy migracji i adhezji komórek związane z białkami o nazwie fibronektyny i tenascyny [5, 6], czy kontrola mechanicznych własności erytrocytów, za które są głównie odpowiedzialne białka nazywane spektrynami [7]. Zauważmy, że mechanizmy takie jak translokacja i degradacja wymagają przeciągania białka przez kanał, np. przez proteozom, podczas gdy rozciąganie mięśni nie zależy od czynników zewnętrznych. Procesy rozciągania zależą od miejsca przyłożenia siły. Wybór tych miejsc nie jest przypadkowy, ale jest determinowany przez proces biologiczny, w którym zachodzi rozciąganie. W przypadku naprężania mięśni tytyna jest rozciągana za swój pierwszy i ostatni aminkowas, podczas gdy w procesach naznaczania białek do degradacji, łańuch ubikwityn powstaje poprzez łączenie terminalu aminowego N z 46. aminokwasem kolejnej ubikwityny. Podobnie ubikwityna, w procesie wciągania innych biomolekuł do proteozomu, jest ciągnięta za inną parę aminokwasów: za terminal N i aminokwas 48.

Niesamowity postęp w ostatnich kilku latach w rozdzielczości urządzeń takich jak mikroskop sił atomowych *(Atomic Force Microscope)*, czy szczypce optyczne *(Optical Tweezers)* umożliwił doświadczalny pomiar siły oporu białka występującej podczas

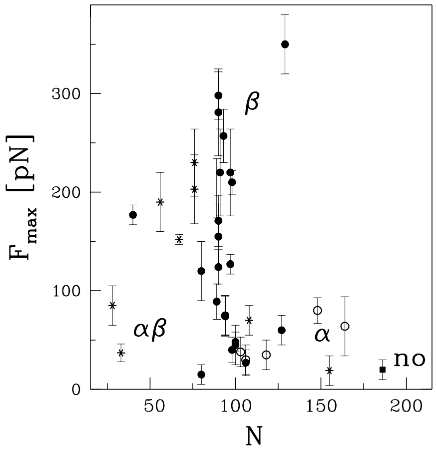




Rysunek 1: Lewy rysunek: schematyczny obraz równolegle ułożonych miofibryli, które są zbudowane z powtarzających się jednostek sarkomerowych. W skład sarkomeru wchodzi między innymi tytyna, która ulega rozciąganiu podczas naprężania mięśni. Prawy rysunek: wynik pomiaru mechanicznego oporu F ośmiu domen tytyny w funkcji rozciągnięcia L, otrzymany podczas rozciągania białka ze stałą prędkością [4].

jego mechanicznego rozciągania. Rysunek 1 przedstawia wyniki pierwszego pomiaru [4] mechanicznego oporu F w funkcji rozciągnięcia L dla tandemu kilku podjednostek tytyny. Kolejne maksima na wykresie $F(L)$ odpowiadają rozwijaniu kolejnej podjednostki. Największy opór z jakim białko przeciwstawia się rozciąganiu odpowiada największemu maksimum na wykresie $F(L)$ Maksimum to oznaczamy poprzez F_{max} i jest ono dobrze określoną wielkością fizyczną. Mimo ogromnego postępu, do tej pory doświadczalnie zbadano tylko 56 białek, kilka konformacji nukleotydów (DNA, RNA), czy elementy naprężeń dla białek motorycznych, z ponad 40000 dostępnych struktur umieszczonych w bazie danych *Protein Structure Data Bank (PDB)*. Rysunek 2 pokazuje skalę mechanicznego oporu F_{max} w funkcji wielkości białka i typu białka. Wyniki te pokazują, iż zbadane białka mają F_{max} pomiędzy 30 a 300 pN, przy czym F_{max} wydaje się nie zależeć od liczby aminokwasów w białku N. Na podstawie danych z rysunku 2 dochodzimy do wniosku, iż nietrywialna odpowiedź białka na mechaniczne rozciąganie wymaga analizy zdecydowanie większego zbioru danych, aby móc odpowiedzieć na pytanie, od czego zależą wykresy $F(L)$, w tym zwłaszcza maksymalny opór molekuły podczas rozciągania.

Wobec takiego stanu rzeczy pojawia się naturalne pytanie, czy istnieją techniki teoretyczne umożliwiające badanie dużych zmian konformacyjnych białek, kwasów nukleinowych, oraz kompleksów molekularnych. Obecnie istnieją dwa zasadnicze podejścia do tego problemu. Po pierwsze, konstruuje się tzw. modele pełnoatomowe, w których rozważa się ruch każdego atomu w białku [8, 9]. Do chwili obecnej zostały zbadane w ten sposób tylko 23 białka, gdyż metody te są bardzo czasochłonne. Ponadto wiele procesów mechanicznych wymaga badania kilku białek jednocześnie, np. tandemowo połączonych podjednostek tytyny, co dodatkow zwiększa czas obliczeń.



Rysunek 2: Wartość mechanicznego oporu F_{max} białek zmierzona eksperymentalnie w funkcji liczby aminokwasów N z których są zbudowane. Puste i pełne koła, gwiazdki i pełne kwadraty odpowiadają białkom o różnej strukturze.

Drugą klasą modeli teoretycznych do opisu zmian konformacyjnych białek stanowią modele gruboziarniste [10, 11], w których zazwyczaj rozważa się tylko ruch centralnego atomu C^{α} w każdym aminokwasie wchodzącym w skład białka. Tak duża redukcja stopni swobody znacznie skraca czas symulacji i w konsekwencji umożliwia rozważanie dużych układów i analizę powtarzalności badanych efektów.

Cele i treść rozprawy

Przedłożona praca jest poświęcona badaniu procesów zwijania i rozwijania białek przy użyciu modeli gruboziarnistych typu Go [12, 13] i modelu sieci gaussowskiej (GNM) [14, 15] z którymi wiąże się nadzieje, iż staną się podstawowym narzędziem do badania fundamentalnych praw fizycznych determinujących mechaniczne i kinetyczne własności biomolekuł.

Cele jakie sobie stawiamy w niniejszej pracy oraz jej konstrukcja są następujące:

- W rozdziale 1 wprowadzamy czytelnika w tematykę białek. Opisujemy mechanizmy rozciągania białek i charakteryzujemy ich własności mechaniczne zarówno z punktu widzenia wyników doświadczalnych jak i teoretycznych.
- W rozdziale 2 przedstawiamy podstawowe prawa fizyczne, którym podlegają molekuły biologiczne [16], opisujemy dynamikę molekularną jako najczęsciej stoso-

wane narzędzie do opisu cząstek w modelach gruboziarnistych, oraz charateryzujemy obecnie używane w symulacjach numerycznych modele gruboziarniste.

- Celem rozdziału 3 jest analiza dotychczas znanych oraz nowych — zaproponowanych przez nas — wariantów modelu Go do badania własności mechanicznych białek. Obecnie istnieje kilkanaście różnych wariantów modelu Go, które przede wszystkim różnią się doborem potencjałów [17, 18, 19, 20, 21], kryterium wyboru mapy natywnych kontaktów [22, 23, 24] oraz siłą ich oddziaływania [10, 25]. Taki stan rzeczy pokazuje, że konieczne jest wyłonienie optymalnej wersji modelu Go. Aby tego dokonać, w tym rozdziale najpierw charkteryzujemy istniejące warianty modelu Go, konstruujemy nowe jego formy typu C^{α} i uogólnioną formę typu $C^{\alpha-\beta}$. W wyniku dostajemy 504 możliwe warianty modelu Go.

- W rozdziale 4 dokonujemy wyboru optymalnego wariantu modelu Go do rozciągania białek na podstawie zgodności wyników teoretycznych z wartościami maksymalnego oporu mechanicznego białek wyznaczonego doświadczalnie dla 29 białek. Rozważamy 61 najbardziej realistycznych wariantów modelu Go spośród 504 skonstruowanych w rozdziale 3.

- W rozdziale 5 sprawdzamy czy optymalna wersja modelu Go do rozwijania jest także najlepszą formą modelu do symulacji procesu zwijania białek. Mimo że zdolność modeli typu Go do opisu zwijania białek jest kwestionowana, w przypadku białek, które mogą mieć kilka różnych dróg zwijania i dużych układów molekularnych, modele uproszczone są nadal jedynym dostępnym narzędziem teoretycznym. Tak więc wybór potencjałów, map kontaktów itp. do konstrukcji modelu Go, który mógłby być stosowany do symulacji procesów zwijania, jest bardzo ważnym krokiem w kierunku zrozumienia wieloskalowych procesów biologicznych.

- Celem rozdziału 6 jest scharakteryzowanie własności mechanicznych białek i wyznaczenie najsilniejszych białek na podstawie struktur zdeponowanych w *PDB*. Wykonanie tego zadania jest możliwe jedynie przy użyciu uproszczonych modeli, a wybór optymalnego modelu typu Go do rozciągania białek na podstawie danych doświadczalnych wykonany w rodziale 4 zapewnia nam wiarygodność otrzymanych wyników. W rozdziale tym opisujemy wykonany przez nas przegląd ok. 7500 kompletnych struktur z *PDB*, z których wyłaniamy 134 najsilniejszych białek. Na podstawie dokonanej analizy wyznaczamy ponadto najważniejsze czynniki decydujące o sile białek.

- W rozdziale 7 badamy własności mechaniczne białek z węzłami. Wykonany w ostatnich latach przegląd [26] struktur białkowych zawartych w *PDB* wyłonił białka o nietrywialnej topologii, zawierające węzły. Białka tego typu były dotychczas bardzo słabo poznane. Zatem o ich własnościach mechanicznych niewiele wiadomo, aczkolwiek wnikliwy przegląd literatury biologicznej pokazuje, że ich istnienie nie jest przypadkowe, a węzły które one tworzą mogą pełnić istotne biologiczne funkcje. W rozdziale tym opisujemy znalezione przez nas niestandardowe zachowanie się węzłów w białkach podczas procesu rozciągania; zachowanie to w szczególności daje wgląd w samą strukturę geometryczną białka.

- Niezwykle istotne jest udostępnienie środowisku naukowemu ogromnej ilości wyników dotyczących własności mechanicznych białek, które otrzymaliśmy i przedstawiliśmy w rozdziałach 6 i 7 — w szczególności, wyniki te mogą być podstawą przyszłych badań doświadczalnych i teoretycznych. W tym celu stworzyliśmy aplikację do rozciągania białek połączoną z bazą danych *Bio-molecule Stretching Data Base* (*BSDB*), która nie tylko zawiera informacje o mechanicznych własnościach biomolekuł, ale także umożliwia samodzielne badanie przez jej użytkowników nowych białek (także zawierających węzły). Ta aplikacja oraz baza danych *BSDB* opisane są w rozdziale 8.

- W rozdziale 9 pokazujemy, iż duże zmiany konformacyjne układów biologicznych mogą być także badane poprzez połączenie wyników z dwóch modeli gruboziarnistych: modelu Go i modelu sieci gaussowskiej GNM. Używając chwilowych konformacji białek z rozciągania modelem Go, stosujemy je do modelu GNM, w którym zgodnie z naszą teorią przewidujemy kolejność pękania kontaktów. Technika ta umożliwa badanie różnego rodzaju zmian konformacyjnych w dużych układach biologicznych, których struktury w stanie przejściowym nie są dostępne, a otrzymanie ich w modelach pełnoatomowych jest wręcz niemożliwe.

- W rozdziale 10 badamy procesy termicznego rozwijania białek. Zjawisko to jest często wykorzystywane w badaniach teoretycznych do interpretacji stanów przejściowych w białku podczas procesu zwijania [27]. Ponadto wydaje się, iż można połączyć zachowania się układu w stanie przejściowym z dużymi zmianami konformacyjnymi podczas rozwijania w ewolucji białek [28, 29, 30]. Tym niemniej analiza procesu termicznego rozwijnia wymaga obserwacji zachodzących zmian w dużych okresach czasu, które nie są osiągalne w konwencjonalnych pełnoatomowych symulacjach. Ponadto takie procesy jak termiczne zwijanie czy rozwijanie wymagają odpowiedniej statystyki do określenia prawidłowości zjawiska. Tak więc mechanizmy te mogą być jedynie wydajnie badane w ramach modeli gruboziarnistych.

- Rozdział 11 stanowi krótkie podsumowanie rozprawy.

Rozdział 1

Białka i ich własności

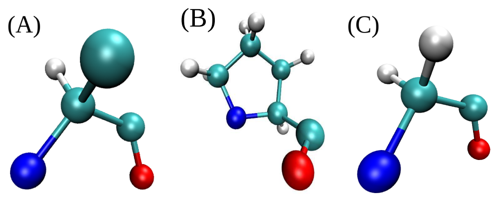
Białka są molekułami biologicznymi obecnymi w każdym żywym organizmie. Służą one do budowy układów molekularnych (rybozom, proteozom, kapsydy), jak też same pełnią wiele funkcji biologicznych, będąc odpowiedzialnymi np. za katalizę (enzymy), transport jonów (hemoglobina, chlorofor), napinanie mięśni, aktywację systemu immunologicznego, transport informacji pomiędzy komórkami itd. Białka wchodzą też w skład włosów, paznokci oraz skóry.

1.1 Skład, budowa i klasyfikacja białek

W podrozdziale tym przedstawiamy skład i budowę białek. Scharakteryzujemy poszczególne poziomy strukturalnej klasyfikacji białek. Podamy definicje pojęć: klasa, architektura, topologia i homologia, zgodnie z klasyfikacją hierarchiczną $CATH$ [31].

Budowa białek

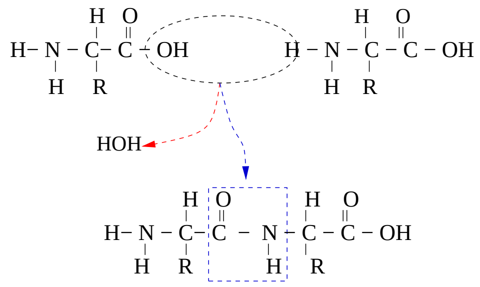
Białka są zbudowane z α-aminokwasów, których kolejność łączenia się z sobą i ich wzajemne przestrzenne ułożenie warunkuje trójwymiarową strukturę i właściwości biologiczne. W naturalnych białkach występuje 20 różnych aminokwasów. Każdy α-aminokwas jest zbudowany z centralnego tetraedrycznego węgla [32], oznaczanego jako C^{α}, do którego jest dołączona grupa aminowa NH_3^+, grupa karboksylowa COO^-, i specyficzna grupa boczna, oznaczana symbolem R, rysunek 1.1. We wszystkich aminokwasach (poza glicyną Gly) pierwszym atomem w grupie bocznej jest atom węgla, oznaczany jako C^{β}. Glicyna jest jedynym aminokwasem, który w miejscu grupy bocznej ma atom wodoru. Miejsce w białku, w którym znajduje się glicyna, wykazuje dużą giętkość. Nietypowym aminokwasem jest też prolina Pro, której grupa boczna jest wygięta w podkowę i połączona z atomem azotu, tworząc charakterystyczny pierścień.



Rysunek 1.1: Model budowy przestrzennej aminokwasu wokół centralnego węgla C^{α} (A), aminokwas prolina (B), aminokwas glicyna (C). Rysunek ten i dalsze prezentowane w tej pracy zostały wykonane przy użyciu programu graficznego VMD [33].

Struktura ta powoduje, iż białko w miejscu proliny tworzy charakterystyczny sztywny zakręt. Pozostałe aminokwasy różnią się chemiczną strukturą grupy bocznej.

Aminokwasy łączą się z sobą w procesie polimeryzacji, który prowadzi do powstania wiązania peptydowego pomiędzy węglem (C) grupy karboksylowej jednego aminokwasu, a azotem (N) grupy aminowej kolejnego aminokwasu. W trakcie tego procesu zostaje uwolniona jedna cząsteczka wody na jedno wiązanie peptydowe (rysunek 1.2). Wiązanie peptydowe jest płaskie i ma charakter wiązania częściowo podwójnego, tak że uniemożliwia ono rotację wokół osi wiązania N-C. Koniec polipeptydu z grupą aminową jest nazywany końcem (lub terminalem) N, a koniec z grupą karboksylową jest nazywany końcem C.



Rysunek 1.2: Schemat tworzenia się wiązania peptydowego.

Struktura przestrzenna białek

Sztywność wiązania peptydowego determinuje przestrzenne uporządkowanie (ułożenie) łańcucha polipeptydowego, ograniczając liczbę możliwych do przyjęcia konformacji. Dodatkowe ograniczenie stanowią zawady steryczne w postaci rozbudowanych grup bocznych niektórych aminokwasów, które ograniczają rotację wokół wiązania peptydowego. Ilość możliwych uporządkowań przestrzennych polipeptydów jest ponadto ograniczona przez otaczający roztwór wodny i aktywność fizjologiczną. Na strukturę białka w roztworze wodnym wpływa wiele czynników, między innymi: oddziaływanie elektrostatyczne, możliwość utworzenia wewnętrznych wiązań wodorowych oraz dwusiarczkowych, czy oddziaływnia hydrofobowe. Z tego powodu polipeptydy w białkach przyjmują konkretną specyficzną konformację nazywaną *strukturą natywną*.

Wymienione powyżej czynniki powodują, iż struktura natywna białka jest ściśle upakowana i charakteryzuje ją kilka odrębnych poziomów, zwanych też strukturami. Są to:

- **Struktura pierwszorzędowa:** jest to liniowa sekwencja aminokwasów połączonych wiązaniami peptydowymi. W jej składzie zawarte są też wszystkie inne wiązania kowalencyjne, np. mostki dwusiarczkowe między resztami cysteiny.

- **Struktura drugorzędowa:** jest to regularne wygięcie łańcucha polipeptydowego, które maksymalizuje liczbę możliwych wiązań wodorowych między tlenem karbonylowym CO a wodorem grupy aminowej NH [32]. Najczęstszymi tego typu wygięciami są helisy α (nazywane też spiralami α) i nici β (struktury β), z których są zbudowane białka przedstawione na rysunku 1.3. W helisie α wiązania wodorowe są ułożone równolegle do osi spirali. Na jeden obrót helisy przypada 3.6 aminokwasu, i typowa jej wielkość to około 5 obrotów. Wiązania wodorowe pomiędzy nićmi β są prostopadłe do łańcucha lub łańcuchów polipeptydowych i tworzone są między wiązaniami peptydowymi. Ponadto zdarzają się układy aminokwasów, które tworzą dodatkowe wiązania wodorowe pomiędzy grupami bocznymi i łańcuchem głównym. Nici β sąsiadujące ze sobą w strukturze mogą być równoległe lub anty-równoległe, zależnie od tego czy przebiegają w tym samym kierunku czy w przeciwnym. Dwie nici β połączone zakrętem nazywamy spinką β. Natomiast kilka równoległych nici β tworzy kartkę β, o typowej wielkość trzech nici i długości około 8 aminokwasów. W strukturze białka helisy α i nici β są połączone poprzez różnego kształtu i długości pętle.

- **Strukura trzeciorzędowa:** jest to sposób przestrzennego ułożenia i połączenia ze sobą struktur drugorzędowych. Zazwyczaj struktura ta określa już całkowitą, trójwymiarową, aktywną biologicznie stukturę natywną białka.

- **Struktura czwartorzędowa:** określa przestrzenne ułożenie i sposób oddziaływania polipeptydowych jednostek, nazywanych domenami, w białku np. w hemoglobinie.

Trójwymiarowa struktura białka może być wyznaczona głównie przy użyciu dwóch technik: jądrowego rezonansu magnetycznego [32] i krystalografię rentgenowską [32].

Struktury białek, jak też ich kompleksy z DNA i RNA, są zdeponowane w bazie danych *Protein Data Bank (PDB)* [1] [34]. Dotychczas zostało tam zdeponowanych ponad 40 000 struktur, co stanowi około 25% poznanych sekwencji białek. Każda zdeponowana struktura w bazie posiada unikalny czteroznakowy kod alfanumeryczny, nazywany *kodem PDB*, np. 27. domena tytyny jest oznakowana kodem 1tit. Takimi kodami *PDB* będziemy się posługiwać w całej rozprawie do identyfikacji białek.

Klasyfikacja białek

Klasyfikacja białek polega na pogrupowaniu ich na podstawie podobieństwa motywów i cech strukturalnych, podobieństwa sekwencyjnego i pokrewieństwa w pochodzeniu.

Jedną z możliwych metod hierarchicznej klasyfikacji domen białkowych jest podział według *CATH, Protein Structure Classification Database (UCL)* [2] [31], która dzieli białka na cztery główne poziomy: klasę (z ang. Class (C)), architekturę (Architecture (A)), topologię (Topology (T)) i homologię (Homology (H)).

- **Klasa** określa procentowy udział aminokwasów należących do tego samego typu struktury drugorzędowej. Rozróżniamy 4 klasy białek:
  - α (1), ponad 65% aminokwasów przyjmuje strukture różnego typu helis,
  - β (2), ponad 65% aminokwasów przyjmuje strukturę nici β,
  - α/β (3), ponad 35 % aminokwasów przyjmuje strukturę nici β i tyle samo różnego typu helis,
  - nie ustrukturyzowane (4), nie występują w nich większe elementy struktury drugorzędowej.

  Numery podane w nawiasach oznaczają liczbę porządkową w klasyfikacji *CATH*.

- **Architektura** określa ogólny kształt struktury domeny, wyznaczony poprzez wzajemną orientację stuktur drugorzędowych, ale przy zignorowaniu połączeń pomiędzy tymi strukturami. Najbardziej typową strukturą w klasie α jest wiązka prostopadła o symbolu (1.20); w klasie β — kanapka (2.60), baryłka (2.40) oraz wstążka (2.10); w klasie α/β — rulon (3.10), dwuwarstwowa kanapka (3.30) oraz trzywarstwowa kanapka (3.40) (rysunek 1.3).

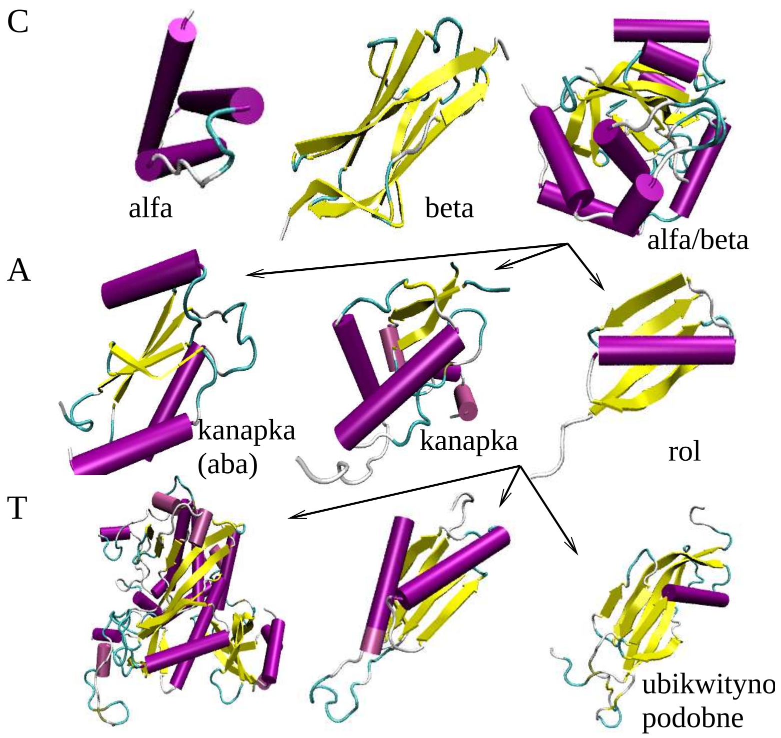
[1] *ProteinDataBank* - jest to kompleksowa baza danych poznanych struktur białkowych, www.rcsb.org

[2] cathwww.biochem.ucl.ac.uk

- **Topologia** określa grupę foldów, która powstała z pogrupowania struktur biorąc pod uwagę naraz ogólny kształt i połączenia pomiędzy strukurami drugorzędową w domenie.

- **Homologia** określa grupę domen białkowych, posiadających pewne szczególne wspólne cechy, które pochodzą od tej samej cechy ich wspólnego przodka.

Klasyfikacje białek są stosowne w różnym celu. W niniejszej pracy klasyfikacja *CATH* jest zastosowana do identyfikacji danych doświadczalnych zawartych w dalszej części tego rozdziału, np. tych przedstawionych w tablicach 1.1 i 1.2, oraz do scharakteryzowania własności mechanicznych białek uzyskanych z symulacji teoretycznych opisanych w rozdziale 6.



Rysunek 1.3: Podział białek według klasyfikacji *CATH*. Ciemne fioletowe cylindry reprezentują helisy α, jasne purpurowe cylindry helisy 3_{10}, żółte wstęgi nici β, szafirowe odcinki zakręty.

Białka z węzłami

Węzłem nazywamy zamknięta, nie przecinająca się jednowymiarową krzywą zanurzoną w trzech wymiarach. Przy pewnych założeniach, które opiszemy w dalszej części pracy, pojęcia węzła można uogólnić także na przypadek krzywych otwartych, które fizycznie mogą odpowiadać obiektom takim jak lina, polimer, lub dowolny łańcuch aminokwasów.

Węzeł w stanie natywnym białka został po raz pierwszy zidetyfikowany w roku 1994 przez Mansfielda [35]. Obecnie znanych jest ok. 300 białek z nietrywialnymi węzłami [36, 37, 38, 39], co stanowi mniej niż 1% całego zbioru PDB. Pokazuje to, że węzły bardzo rzadko występują w białkach. Funkcja biologiczna białek z węzłami nie została jeszcze wyjaśniona, chociaż wiadomo że biorą one udział np. w procesie degradacji w proteozomie [40]. Sposób zaplątania i kształt węzłów w białkach nie wydaje się być przypadkowy [38].

Wspomnijmy ponadto, że dotychczas dużo bardziej zostały zbadane węzły występujące w innych obiektach biologicznych, np. w DNA i RNA. Ich własności zostały stosunkowo dobrze poznane doświadczalnie dzięki zastosowaniu mikroskopu sił atomowych, oraz teoretycznie poprzez symulacje numeryczne [41, 42, 43]. Symulacje numeryczne pozwoliły także na poznanie wielu innych aspektów dotyczących fizyki węzłów np. na polimerach i linach idealnych [44, 45, 46, 47].

Reasumując, węzły często występują w otaczającym nas świecie. Tym niemniej, ich własności w białkach nie zostały do tej pory wyjaśnione. W rozdziale 7 niniejszej pracy, na podstawie symulacji rozciągania wyznaczymy główne własności węzłów w białkach i porównany je do własności węzłów na polimerach.

1.2 Mechaniczne rozwijanie białek

W tej części opiszemy mechaniczne siły występujące w układach biologicznych. Scharakteryzujemy zasady eksperymentalnego pomiaru oporu białka oraz sposób interpretacji wyników. Opis sposobu przeprowadzania eksperymentu jest niezbędny do konstrukcji teoretycznego modelu do rozciągania białek i porównania wyników teoretycznych z doświadczalnymi. Następnie scharakteryzujemy obecny stan wiedzy dotyczącej własności mechanicznych białek na podstawie danych doświadczalnych i symulacji teoretycznych. Podamy teoretyczną interpretację wyników doświadczalnych.

1.2.1 Mechaniczne siły w układach biologicznych

Proces mechanicznego rozwijania w układach biologicznych należy rozważać na dwóch poziomach: pojedynczej molekuły charakteryzującej się pewną stabilnością mechaniczną, oraz układu makromolekuł, który determinuje charakter mechanicznego rozwijania.

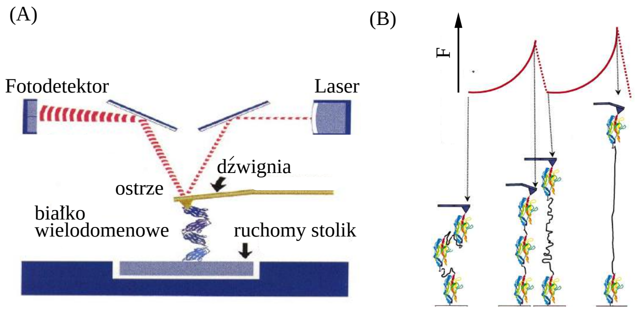
Wiele „pojedynczych" molekuł biologicznych ma ściśle określoną mechaniczną funkcję. Opór tych molekuł podczas rozwijania za pomocą mechanicznej zewnętrznej siły — ich mechaniczna stabilność — ma kluczowe znaczenie fizjologiczne. Przykładem tzw. *białka mechanicznego* jest tytyna [48]. Tytyna jest odpowiedzialna za pasywną elastyczność sarkomeru w szkieletowych i sercowych mięśniach, gdzie zachowuje się jak molekularna sprężyna i zapewnia powrót sarkomeru do jego wyjściowej konformacji po rozciągnięciu mięśni [49, 50]. Fibronektyny i tenascyny są składnikami macierzy zewnątrzkomórkowej, w której są rozciągane i skracane, tak aby ułatwić np. migrację komórek i adhezję [5, 6]. W przypadku adhezji, mechaniczna siła powoduje otwieranie się fibronektyn i udostępnianie miejsc aktywnych. Wspólną cechą tych mechanicznych białek jest tandemowe ułożenie pojedynczych domen typu kanapek β. Jednakże taka forma białka nie jest wyznacznikiem *białka mechanicznego.* Białka nazywane spektrynami [50] składają się wyłącznie z helis α. Spektryny pełnią kluczową rolę w elastyczno-mechanicznych własnościach erytrocytów, które podczas swojej wędrówki w żyłach są ściskane i deformowane [7].

Powyższe przykłady demonstrują wagę mechanicznej stabilności białka jako jego indywidualnej własności. Jednocześnie istnieje wiele procesów komórkowych, gdzie rozwijaniu podlegają całe makromolekuły jak i też pojedyńcze przypadkowe białka. Na przykład, struktura kwasu nukleinowgo musi być rozpelciona, aby mogło nastąpić przemieszczenie się polimerazy DNA i RNA, rybozomu oraz helikazy DNA i RNA. Ponadto istnieje coraz więcej dowodów, iż ta „maszyna" molekularna wywiera mechaniczną siłę na biomolekuły, aby te mogły pełnić swoją komórkową funkcję. Kolejnym przykładem jest rozwijanie białek podczas transportu przez niektóre błony komórkowe. Kanały zawarte w błonie są zbyt wąskie i sztywne, aby białko mogło przez nie być przeciągnięte w stanie natywnym. Innym przykładem jest mechanizm znakowania białek przeznaczonych do degradacji poprzez dołączanie łańcucha ubikwityn. Łańcuch ten pełni funkcję „wędki", za którą białka są ciągnięte. Odkrycie tego procesu zostało nagrodzone nagrodą Nobla w roku 2002 [51]. Najmniej zbadanym procesem mechanicznego rozwijania białek jest degradacja białek w proteozomie [1].

Podsumowując, mechaniczne rozwijanie białek i makromolekuł jest obecne na każdym etapie cyklu komórkowego. Może ono zależeć od czynników zewnętrznych (otoczenia) tak jak w przypadku przeciskania przez kanały, lub też nie zależeć (rozciąganie tytyny). Ponadto należy jeszcze zauważyć, iż miejsce, za które białko jest rozciągane nie jest przypadkowe, ale wynika z procesu biologicznego w którym zachodzi. Tak więc, np. w przypadku naprężania mięsni, biomolekula jest rozciągana za końce białka, podczas gdy w procesach naznaczania białek do degradacji łańcuch ubikwityn jest połączony za koniec aminowy N i 46. aminokwas. Ta sama ubikwityna w procesie wciągania innych biomolekuł do proteosomu jest ciągnięta za inna parę: 48. aminokwas i koniec C.

1.2.2 Doświadczalne rozciąganie białek i ich charakterystyka

Do pomiaru mechanicznego oporu biomolekuły podczas rozciągania, w zależności od skali generowanego oporu, używa się szczypiec magnetycznych (maksymalna siła 20 pN), szczypiec optycznych (do 100 pN), lub mikroskopu sił atomowych (AFM) (od 20 pN). Ze względu na nieograniczony zakres pomiaru wielkości siły, najpowszechniej stosowany jest mikroskop sił atomowych [52]. Może on służyć do pomiaru rozciągania pod wpływem stałej prędkości, lub pod wpływem stałej siły [53, 54], kiedy to nazywany jest „imadłem siłowym”. Rozciąganie ze stałą siłą wyznacza zależność siły od prawdopodobieństwa rozwinięcia, podczas gdy rozciąganie ze stałą prędkością wyznacza bezpośrednio maksymalny opór białka. W niniejszej pracy będziemy testować gruboziarniste modele teoretyczne na podstawie wyników pomiarów rozciągania ze stałą prędkością.



Rysunek 1.4: Budowa mikroskopu sił atomowych. (A) Schematyczny obraz zachowania się tandemu białek podczas rozciągania wraz z wykres oporu mechanicznego F w funkcji jego rozciągnięcia L otrzymany w pomiarze przy użyciu mikroskopu sił atomowych. (B)

Mikroskop sił atomowych pracujący w modzie ze stałą prędkością jest pokazany na rysunku 1.4. Jest on zbudowany z ruchomego stolika, na którym jest umieszczona próbka z badaną biomolekułą, dźwigienki zakończonej ostrzem i urządzenia optycznego wraz z laserem do pomiaru wychylenia dźwigni. Rozciąganie białka odbywa się poprzez zaczepienie ostrza za znacznik w białku i przesuwanie stolika w przeciwnym kierunku. W eksperymencie tym zakłada się, iż opór molekuły jest proporcjonalny do wychylenia dźwigienki w funkcji rozciągnięcia molekuły, zgodnie z prawem Hooke'a.

Mimo że idea eksperymentu jest prosta, sam pomiar i przygotowanie próbki nie należy do trywialnych. Przygotowanie próbki polega na skonstruowaniu łańcucha, w skład którego wchodzi badane białko, a następnie zaczepienie go bezpośrednio lub za pomocą łącznika do podłoża i do ostrza. Początek pomiaru jest zazwyczaj obarczony dużym błędem wynikającym z ułożenia badanej molekuły prostopadle do kierunku rozciągania.

Rysunek 2(B) z wstępu pokazuje typowy wykres siły oporu F w funkcji długości rozciągniętej molekuły L. Jest to tzw. wykres $F(L)$, który bynajmniej nie odpowiada prostej zależności liniowej. Zależność liniowa jest obserwowana na małych odległościach, jednakże bardzo szybko odpowiedź molekuły przechodzi w kształt złożony z wielu maksimów. Ten kształt jest unikalny dla każdego białka i jest nazywany *mechanicznym śladem* białka. Maksima i minima we wzorze są determinowane przez silniejsze i słabsze elementy struktury. Innymi słowy, maksima wyznaczają siłę potrzebną do rozerwania pewnej grupy kontaktów.

Największe wzniesienie na krzywej $F(L)$ odpowiada największej sile F_{max} z jaką białko przeciwstawia się rozciąganiu, zanim zostanie całkowicie rozwinięte do postaci liniowej sekwencji aminokwasów. F_{max} wyznacza najbardziej stabilny element białka. Molekułę białka zazwyczaj rozciąga się do długości L_{max}, którą można oszacować jako

$$L_{max} = N \cdot d_0,$$

gdzie N jest liczbą aminokwasów pomiędzy miejscami do których przyłożona jest siła rozciągająca, a długością pomiędzy kolejnymi węglami C^{α}, $d_0 = 3.8$Å. Gwałtowny wzrost siły przy zbliżaniu się do tej odległości oznacza pomiar naprężeń w wiązaniach kowalencyjnych między kolejnymi atomami ciężkimi w łańcuchu polipeptydowym białka. Wiązania kowalencyjne oraz wiązania siarczkowe nie podlegają rozerwaniu przez mikroskop siły atomowej.

1.2.3 Własności mechaniczne białek

Dane doświadczalne

Pierwszy eksperymentalny pomiar siły oddziaływań wewnątrzmolekularnych został przeprowadzony w roku 1994 dla komplementarnych nici DNA przez Lee i in. [55]. W tym samym roku Moy i in. [56] dokonali pomiaru siły rozrywania układu receptor-ligand, a już w dwa lata później miał miejsce pierwszy klasyczny eksperyment mechanicznego rozciągania przeprowadzony dla kompleksu streptawidyna-biotyna [57] i kwasów nukleinowych [58]. W roku 1997 ruszył lawinowy pomiar własności mechanicznych białek, DNA i RNA. Rief i in. [59] jako pierwsi zaobserwowali zmiany konformacyjne pod wpływem zewnętrznej siły dla pojedynczego łańcucha polisacharydów. W tym samym czasie Mitsui i in. [60] oraz Rief i in [59] wykazali odwracalność procesu mechanicznego rozwijania dla pojedynczej domeny tytyny za pomocą AFM. Następnie, w roku 1998, Oberhauser i in. [61] oraz Rief i in. [62] po raz pierwszy pokazali wykazali zależność charakterystyką mechanicznego oporu białek a icj mechaniczną funkcją, także dla tytyny. W międzyczasie wykonano pomiary między innymi dla fibronektyn, spektryn, ubikwityny jak też enzymów. Do chwili obecnej własności mechaniczne białek zostały zbadane dla 56 struktur. Pełna lista białek z wartością maksymalnej siły oporu jest podana w tabeli 1.1. Inne miejsce przyłożenia siły niż w terminalnych aminokwasach jest podane w nawiasach, np. ubikwityna rozciągana za 46. aminokwas i koniec C jest oznaczona 1ubq(46-C). Podsumwanie dotychczasowego stanu wiedzy na temat własności mechanicznych białek jest przedstawione na rysunku 1 we wstępie, gdzie maksymalny opór białek F_{max} jest pokazany w funkcji długości N, jak też klasy białek. Wartość mechanicznego oporu wraz z opisem białka, wartość prędkości w jakiej został wykonany pomiar i dodatkowo kilka struktur nie białkowych są przedstawione w tablicy 1.1.

Tablica 1.1: Mechaniczny opór F_{max} białek wyznaczony doświadczalnie. Symbol PBD oznacza kod białka, N liczbę aminokwasów, v_p prędkość doświadczalną użytą do rozciągania biomolekuły. W przedostatniej kolumnie podajemy oznaczenie białka zgodnie klasyfikacją hierarchiczną $CATH$. Niepełna numeracja lub puste miejsca oznaczają brak odpowiednich danych w tej bazie.

białko	PDB	N	F_{max}[pN]	v_p [μm/s]	$CATH$	odnośniki
Tytyna						
$I1_{I27}$	1g1c	97	127	600	2.60.40.10	[63]
I4		90	171+/26		2.60	[64]
I5		~90	155+/-33		2.60	[64]
I4-I11		~90	150-200+/30		2.60	[64]
I27	1tit	89	204+/-30	0.2-1.5	2.60.40.10	[4, 48, 64]
C6,				zazwyczaj 0.6		[65, 66, 67]
				zazwyczaj 0.6		[68, 69, 70]
I28		93	257+/-27		2.60.40.10	[68, 64, 65]
I27-I28		~91	211-306	1	2.60.40.10	[68]
I27-I30		~90	230		2.60.40.10	[4]
I27-I34		~90	150-330+/-20	1	2.60.40.10	[4, 68, 64]
I27-I34		~90	231+/-6	0.5	2.60.40.10	[71, 72]
I32		90	298+/-24		2.60.40.10	[64]
I34		90	281		2.60.40.10	[64]
Sk47-Sk53		~96	210	0.5	2.60	[71]
A60-A65		~99	180	0.5	2.60	[71]
Fibronektyny FN						
pFN		~90	80-200	0.6	2.60	[73, 74, 75]
$^{1}FNIII_{2FNIII}$		97	220+/-44	0.6	2.60	[73]

białko	PDB	N	F_{max}[pN]	v_p [μm/s]	$CATH$	odnośniki
^{2}FNIII$_{1FNIII}$		91	220+/-44	0.6	2.60	[73]
^{1}FNIII$_{I27}$		97	$\leq$ 20	0.6	2.60	[73]
^{10}FNIII	1fnf	$\sim$92	74+/-20	0.6	2.60	[70, 73]
^{11}FNIII	1ttg	94	74	0.6	2.60.40.30	[74]
^{12}FNIII$_{13FNIII}$	1fnh	92	124+/-18	0.6	2.60.40.30	[73]
^{13}FNIII$_{I27}$	1fnh	89	89+/-18	0.6	2.60.40.30	[73]
$^{2-14}$FNIII		$\sim$90	145	0.6	2.60	[73]
Inne FN						
TN		$\sim$91	137+/-12	0.3-0.5	2.60.40.30	[61, 73]
TNFN*15		$\sim$91	113	0.2-0.6	2.60.40.30	[71]
TNFNAll		$\sim$90	138+/-12	0.3-0.5	2.60.40.30	[61]
TNFNA-D		$\sim$91	138	0.3-0.5	2.60.40.30	[61]
I54-I59	1nct	$\sim$98	210	0.5-1	2.60.40.10	[72, 76]
I48-I54,ConFN		$\sim$100	200	0.5	2.60	[71]
Spektryny						
-monomery						
natywana	1u4q	$\sim$106	30+/-5	0.3	1.20.58.60	[77, 7]
α-spektryny R16	1aj3	$\sim$106	54+20	0.3	1.20.58.60	[78]
α-spektryny$_{13-18}$	1u4q	$\sim$106	26+15	0.3	1.20.58.60	[77, 79]
α-spektryny$_{18-21}$	1u4q	$\sim$106	26+15	0.3	1.20.58.60	[7]
β-spektryny$_{1-4}$	1s35	$\sim$106	27-13	0.3	1.	[77, 80, 79]
α-aktyny$_{1-4}$	1hci	$\sim$106	38	0.3	1.20.58.60	[79, 7]
-dimery						
α,β-spektryny*			54+/-30			[79]
α-aktyny$_{1-4}$			50+/-20		1.20.58.60	[79]
Inne						
PEVK$_{I27}$		186	<20	0.4	4	[64, 81, 82, 83]
N2$_{I27}$		572	<20		4	[64, 76]
rybonucleaza H	1rnh	155	19		3.30.420.10	[67]
E2lip3$_{I27}$	1qjo	80	15+/-10	0.7	2.40.50.100	[84]
E2lip3(N-41)$_{I27}$	1qjo	40	177+/-3	0.7	2.40.50.100	[84]
ankiryna*1	1n11	33	37		3	[85, 86]
ankiryna*24	1n11	792	450		3	[85]
Mel-CAM		$\sim$ 100	30		2.60.40.10	[87]
Mel-CAM $^{+DTT}$		$\sim$ 100	41		2.60.40.10	[87]
VACM1 $^{+DTT}$	1vcs	$\sim$ 98	40		2.60.40.10	[88]
$^{1-5}$DdFLN	1wlh	100	40-100		2.60	[89]
$_{I30}{}^{4}$FLN$_{I31}$	1ksr	100	45+/-20	0.2-0.4	2.60	[89, 90, 91]
C2A	1dqv	127	60		2.60.40.180	[92]
T4 lizozym	1b6i	164	64		1.10.530.40	[93]
barnaza$_{I27}$	1bnr	108	70	0.1-0.5	3.10.450.30	[94]
kalmodulina	1cfc	148	80		1.10.238.10	[92]
ubikwityna(48-C)	1ubq	28	85+/-20	0.3	3.10.20.90	[95, 96]
ubikwityna(N-C)	1ubq	76	203	0.2-0.4	3.10.20.90	[95, 96]
ubikwityna	1ubq	76	230+/-34	1	3.10.20.90	[96]
GFP$_{DdFLN}$	1b9c	238	104+/-40		2.40.155.10	[97]

białko	*PDB*	*N*	F_{max}[pN]	v_p [μm/s]	*CATH*	odnośniki
GFP_{Ig}	1b9c	238	104+/-40		2.40.155.10	[97]
GFP(3-212)	1emb	219	130+/-30		2.40.155.10	[98]
GFP(132-212)	1emb	80	120+/-30		2.40.155.10	[98]
GFP(3-132)	1emb	129	350+/-40		2.40.155.10	[98]
białko L	1hz6	67	152+/-5	0.7	3.10.20.10	[99]
nić pajęcza		∼608	176+/-73	0.2-1.5	2	[100]
filamin A* 24		96	50-220	0.37	2.60	[101]
bovin	1v9e	259	1100		3.10.200.10	[102]
bakterio-rodopsyna	1at9	231	350		1.20.1070.10	[103, 104]
streptavidyna-biotyna			350			[105]
makroglobuła			750			[60]
DNA			13			[58, 102]
P5abc			19			[106]

Największy mechaniczny opór, około 400 pN, jest generowany przez białko ankirynę [85, 86], jednakże tak duża siła wynika tutaj z przejścia topologicznego między wyjściową strukturą białka a strukturą powstałą w wyniku rozciągania. Rozciąganie pojedynczej domeny ankiryny generuje siłę zaledwie około 50 pN. W klasie białek rozciąganych w kierunku wyznaczonym przez terminale z największą siłą około 200 pN przeciwstawiają się rozciąganiu 27. domena tytyny (1tit) i ubikwityna (1ubq). Mimo że generują one podobną siłę, *mechaniczny ślad* białka jest inny. Tytyna ma architekturę kanapki β a ubikwityna α/β rulonu. Kolejne dwa białka, jedno o budowie podobnej do 1tit a drugie o zupełnie innej C2A (1dqv) (trzywarstwowa kanapka α/β) i polikalmodulina (1cfc), wykazują podobną siłę 60 pN, jednakże dużo słabszą niż dla tytyny czy ubikwityny. *Mechaniczny ślad* dla C2A jest wyraźny, a dla 1cfc rozmazany. Bardzo interesujący przypadek stanowią białka ubkiwityna, E2lip3 oraz GFP, w których opór mechaniczny zmienia się ponad trzykrotnie w zależności od wyboru miejsca przyłożenia siły.

Ta krótka charakterystyka miała na celu pokazanie, iż liczba dotychczas przeprowadzonych doświadczeń nie jest wystarczająca do określenia czynników determinujących własności mechaniczne białek. Ponadto pomiar eksperymentalny określa raczej w ogólny sposób mechanizm rozciągania, niż wskazuje na konkretne elementy struktury odpowiedzialne za największy mechaniczny opór.

Dane teoretyczne

Istotny wkład w zrozumienie własności mechanicznych białek miały symulacje teoretyczne przy użyciu modeli pełnoatomowych [107, 108]. W technikach tych rozważa się ruch każdego atomu, jak i dość realistyczny charakter oddziaływań. W przełomowej pracy z roku 1998 [8] Schulten pokazał, iż głównym czynnikiem determinującym duży opór 27. domeny tytyny jest 6 wiązań wodorowych pomiędzy równolegle ułożonymi terminalnymi nićmi β, A i G, które także są ułożone równolegle do kierunku rozciągania. Takie ułożenie terminalnych nici β w stosunku do kierunku rozciągania zostało nazwane *motywem ścinającym*. Natomiast prace [109, 110] pokazały, że fibronektyny, które są bardzo podobne strukturalnie do 27. domeny tytyny a charakteryzują się małą siłą oporu, nie mają tych właśnie wiązań wodorowych. Jednakże w obu tych przypadkach końcowe rezulaty uzyskano tylko na podstawie jednej lub dwóch trajektorii, które nie muszą odzwierciedlać całkowitego krajobrazu energetycznego białka. Mimo że modele pełnoatomowe pozwalają wyjaśnić daną sytuację, zostały one zastosowane tylko do około 23 białek, gdyż rozciągnięcie nawet średniej wielkości białka wymaga kilku miesięcy obliczeń numerycznych. Dodatkowym ich mankamentem jest o kilka rzędów większa siła niż wartości otrzymywane eksperymentalnie. Pełna lista maksymalnej siły mechanicznego oporu białek wyznaczona na podstawie modeli pełnoatomowych jest przedstawiona w tablicy 1.2.

Tablica 1.2: Mechaniczny opór białek F_{max} wyznaczony teoretycznie na podstawie modeli pełnoatomowych.

Białko	PDB	N	F_{max}[pN]	v_p[Å/ps]	odnośniki
Immunoglobuliny					
I1	1gcg	97	2397	0.5	[111]
I1 zredukowana	1gcg	97	2090	0.5	[111]
I27	1tit	89	2479	0.5	[111, 96]
I27	1tit	89	2040	0.5	[8, 9, 112, 113, 109]
I28		93	2082	0.5	[8]
Fibronektyny typ III					
^{1}FNIII	1oww	97	1500	0.01	[114]
^{2}FNIII		91	1600	0.01	[114]
^{7}FNIII	1fnf	93	1638	0.5	[115, 113]
^{9}FNIII	1fnf	91	2000	0.1	[115, 109, 110, 116]
^{10}FNIII	1fnf	94	1580	0.5	[115, 109, 110, 113, 116]
^{9}FNIII	1fnf	91	^{9}F<^{10}F	0.1	[117, 118]
Inne					
ubikwityna (N-C)	1ubq	76	2000	0.1	[119]
ubikwityna (48-C)	1ubq	28	1200	0.1	[119]
bovine	1v9e	259	3000	0.5	[102]

Białko	PDB	N	F_{max}[pN]	v_p[Å/ps]	odnośniki
barnaze	1bnr	108	500	0.01	[94]
cad1	1edh	211	1850	0.5	[113]
cad2	1edh	211	1970	0.5	[113]
VCAM1	1vsc	89	2050	0.5	[113]
VCAM2	1vsc	108	1620	0.5	[113]
T4 lizozym	1b6i	164	75	10^4	[116]
cytochrom C6 cc6	1cyi	89	no peak	0.5	[113]
białko wiążące igb	1bdd	60	no peak	0.5	[113]
synaptotagmin (c2)	1rsy	125	no peak	0.5	[113]

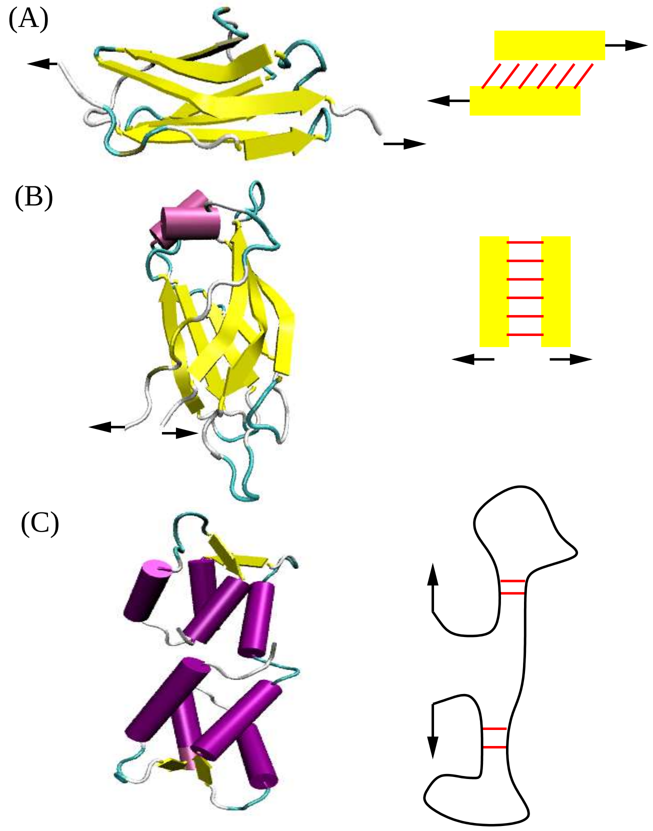
Alternatywnym podejściem do badania własności mechanicznych białek jest zastosowanie teoretycznych modeli gruboziarnistych. Kluczowym elementem ich konstrukcji jest bardzo duża redukcja stopni swobody, np. poprzez reprezentowanie każdego aminokwasu w białku jednym atomem C^{α}. Szczegółowy opis modeli gruboziarnistych znajduje się w rozdziale 2.2, a w szczególności modelu Go w rozdziale 3. Modele te pozwoliły na teoretyczne przeanalizowanie ponad 60% białek z biomolekuł wcześniej zbadanych doświadczalnie, i wykonanie wstępnej charakterystyki własności mechanicznych białek [84, 99, 120, 121].

Czynniki determinujące mechaniczny opór białek

Zestawienie rezultatów doświadczalnych 1.1 i teoretycznych 1.2 pozwala wyodrębnić kilka cech i motywów w białku, które decydują o sile oporu. Wydaje się, że stawiany podczas rozciągania opór mechaniczny białka zależy od jego klasy, topologii i od kierunku rozciągania. Przypuszcza się, iż białka helikalne lub białka nieustrukturyzowane charakteryzują się małą siłą oporu (np. 1n11, 1a35, 1cfc). Mechaniczny opór białek klasy β lub α/β zależy od ich wewnętrznej topologii — charakteryzującej się motywem *ścinającym* lub *suwakowatym*.

- Motyw *ścinający* determinuje duży opór w białku podczas mechanicznego rozciągania. Jest tworzony przez dwie równoległe, terminalne nici β, połączone wiązaniami wodorowymi, podczas gdy całość jest ułożona równolegle do kierunku rozciągania. Ułożenie motywu ścinającego np. w tytynie jest pokazane na rysunku 1.5.

- Motyw *suwakowaty* determinuje małym opór mechaniczny w białku podczas rozciągania. Jest zbudowany z równoległych terminalnych nici β (np. pierwsza domena białka C2A 1dqv), lub helis α (np. białko poly-kalmodulina 1cfc), jednakże kontakty łączące te elementy są ułożone prostopadle do kierunku rozciągania, rysunek 1.5.



Rysunek 1.5: Motyw ścinający, odpowiedzialny za duży opór białka podczas rozciągania (na przykładzie tytyny) (A), motyw suwakowaty w klasie białek β — mały mechaniczny opór (na przykładzie C2A) (B), motyw suwakowaty w klasie białek α — mały mechaniczny opór (na przykładzie 1cfc) (C).

Należy zauważyć, że bardzo ważnym elementem dla prawidłowego porównania wyników doświadczalnych z teoretycznymi przewidywaniami jest określenie czynników zewnętrznych i stopnia ich udziału w mechanicznym oporze. Mechaniczny opór białka zależy przede wszystkim od dwóch czynników:

- Prędkości; im większa prędkość rozciągania tym większy stawiany opór, jednakże charakter wzrostu wartości stawianego oporu jest indywidualny dla każdego białka [72], np. domeny tytyny charakteryzują się słabą zależnością wartości oporu od prędkości w granicach 100 — 1000 nm/s, podczas gdy białka fibronektyny typu III są bardzo czułe na zmianę prędkości nawet w małym zakresie.

- Historii rozciągania, która ogólnie mówi o otoczeniu badanej molekuły, czyli o: ilości, jednorodności i sztywności domen przed i po badanym białku, postępującym wzroście uległości wobec siły rozciągającej wraz ze wzrostem liczby rozwiniętych domen (ta zależność jest bardzo ważna dla łatwo rozwijalnych białek), sposobie oddziaływania między badanym białkiem a otoczeniem [122].

Rozdział 2

Teoretyczny opis biomolekuł

Praca ta poświęcona jest teoretycznemu opisowi biomolekuł przy pomocy modeli gruboziarnistych. W tym rozdziale przedstawiamy podstawowe prawa fizyczne opisujące molekułę biologiczną [16], opiszemy dynamikę molekularną jako najczęsciej stosowane narzędzie do opisu cząstek w modelach gruboziarnistych, oraz scharateryzujemy obecnie stosowane modele gruboziarniste.

Najdokładniejszy opis molekuły można uzyskać przy użyciu równania Schrödingera [123]. Równanie Schrödingera opisuje zachowanie funkcji falowej i pozwala wyznaczyć przestrzenny rozkład prawdopodobienstwa w systemie kwantowym. Rozwiązanie tego równania w pełnej ogólności nie jest jednak możliwe dla układów o takim stopniu złożoności jak białka. Aby możliwe było przeprowadzenie symulacji numerycznych, stosuje się zatem dwa istotne przybliżenia. Pierwsze z nich to przybliżenie Borna-Oppenheimera [124], zgodnie z którym ruchy jąder oraz elektronów można rozseparować. Drugie przybliżenie polega na przyjęciu, że jądra są cząstkami punktowymi i ich ruch może byc opisany równaniami Newtona dla punktowych mas. Tym niemniej, takie równania Newtona muszą zawierać pewien dodatkowy potencjał, który uwzględnia uśredniony wpływ elektronów, efekty kwantowe itp. Przykładem takiego potencjału jest potencjał Lennarda-Jonesa, z którego wielokrotnie korzystamy w niniejszej pracy. W efekcie molekułę można opisać jako układ mechanicznych atomów połączonych sprężynami i obdarzonych ładunkiem, znajdujących się w dodatkowym polu potencjału.

Podkreślmy ponadto, że obraz molekuł uzyskany dzięki eksperymentom (np. krystalograficznym) nie jest wystarczający do poznania mechanizmu ich funkcjonowania. Dostarcza on jedynie „zamrożonego” obrazu systemu molekularnego odpowiadającego temperaturze 0 K, podczas gdy rzeczywiste procesy biologiczne zachodzą w temperaturze pokojowej około 300K.

Powiązanie struktury, zmian konformacyjnych i wynikającej z tego funkcji białek wymaga obserwacji molekuły w przestrzeni i czasie. Śledzenie zmian strukturalnych pod wpływem różnych czynników zewnętrznych, np. takich jak siła, dla dużego zbioru danych dostarcza niezbędnej infomacji o dynamicznych własnościach. Osiągnięcie tych celów jest możliwe przy użyciu *dynamiki molekularnej.*

2.1 Dynamika molekularna

Pojęcie dynamiki molekularnej odnosi się do numerycznego rozwiązywania klasycznych równań Newtona dla układu atomów. Symulacje dynamiki molekularnej są wykorzystywane do opisu procesów równowagowych układów zbyt skomplikowanych by móc je opisać analitycznie, procesów które wymagają dużej statystyki, i przede wszystkim do badania dynamiki układu.

Równania ruchu

W dynamice molekularnej ruch jąder atomowych opisywany jest klasycznymi równaniami ruchu w których uwzględniony jest dodatkowy potencjał reprezentujący ruch elektronów, efekty kwantowe itd. Rówania Newtona [125] dla układu N oddziałujących atomów o masach $m_i, i = 1, \ldots, N$, przyjmują postać

$$m_i \ddot{\mathbf{r}}(t) = \mathbf{F}(\mathbf{r}(t)) = -\nabla E_p(\mathbf{r}(t)) + ... \tag{2.1}$$

gdzie $\mathbf{r} \in \mathbb{R}^3$ jest wektorem położenia i-tego atomu. Wektor całkowitej siły $\mathbf{F}$ jest zbudowany z pochodnych cząstkowych energii potencjalnej E_p opisującej oddziaływania atomów, oraz wszystkich dodatkowych sił występujących w układzie oznaczonych poprzez $\ldots$, na które mogą się składać siły zewnętrzne (np. mechaniczne rozciąganie), oraz siły niezachowawcze (tarcie).

Analityczne rozwiązanie równań ruchu możliwe jest tylko dla bardzo prostych sił i małych układów. W związku z tym analizę skomplikowanych procesów zachodzących w białkach przeprowadzamy w oparciu o metody numeryczne. Istnieje kilka sposobów numerycznych rozwiązania równań ruchu, np. żabiego skoku [126] czy też przewidywania-poprawiania [127], których dobór zależy od charakteru opisywanego układu. W tej pracy używamy algorytmu przewidywania-poprawiania piątego rzędu. W wyniku rozwiązania numerycznego równań ruchu otrzymuje się zespół par położeń i prędkości atomów $\{\mathbf{R}^n, \mathbf{V}^n\}$, dla liczb naturalnych n reprezentujących kolejne chwile czasu $t = n\Delta$, gdzie Δ jest krokiem czasowym symulacji. Zespół $\{\mathbf{R}^n, \mathbf{V}^n\}$ dla n=1,2,3... opisuje dynamiczną trajektorię układu.

Geometria układu i protokół symulacji

W przeprowadzanych symulacjach każdą molekułę reprezentujemy poprzez zbiór mas punktowych skupionych w jądrach atomów. Położenia atomów w stanie natywnym zazwyczaj pochodzą z danych eksperymetalnych lub modeluje się je „de novo" teoretycznie. Symulacja dynamiki molekularnej w zależności od doboru zespołu statystycznego, w którym się ona odbywa, składa się z trzech etapów. Pierwszy etap polega na określeniu początkowych współrzędnych i prędkości atomów układu. Początkowa prędkość jest przydzielana atomom losowo, tak aby wypadkowy pęd układu był zerowy i rozkład prędkości odpowiadał zamierzonej temperaturze symulacji, zgodnie z rozkładem Maxwella. Drugim etapem jest doprowadzenie układu do równowagi, podczas którego dochodzi do wymiany między energią potencjalną i kinetyczną systemu. Układ modelowy osiąga stabilność, gdy całkowita energia układu oraz jej składowe – energia potencjalna i kinetyczna – fluktuują wokół ustalonych wartości średnich. Osiągnięcie stabilności oznacza gotowość modelu do przeprowadzenia symulacji dynamiki molekularnej. Ostatni trzeci etap polega na gromadzeniu informacji o zachowaniu się układu. Dobry opis zachowania układu w symulacjach dynamiki molekularnej wymaga dużej powtarzalności. Zatem protokół ten odpowiada symulacjom, w których całkowita energia układu jest stałą ruchu, co odpowiada zespołowi mikrokanonicznemu.

2.1.1 Dynamika molekularna w zespole kanonicznym

Opisany powyżej protokół dynamiki molekularnej w układzie mikrokanoniczym nie jest interesujący pod względem opisu rzeczywistego układu biologicznego, w którym procesy biologiczne zależą od temperatury czy ciśnienia. Z tego powodu opracowano wiele algorytmów, w których istnieje możliwość prowadzenia symulacji w zespole kanonicznym (stała liczba cząstek, objętość, temperatura; NVT), izotermicznym (stała liczba cząstek, ciśnienie, temperatura; NPT) czy przy stałej liczbnie cząstek, ciśnieniu i entalpi (NPH). Z punktu widzenia tej pracy najbardziej interesujący jest zespól NVT, którego częściowa implementacja w dynamice molekularnej jest przedstawiona poniżej.

Równanie Langevina

W rzeczywistym układzie molekuła białka jest zanurzona w ośrodku wodnym o temperaturze pokojowej. Z tego powodu podlega ona nieustanym ruchom termicznym. Siły naśladujące wpływ roztoworu wodnego o zadanej temperaturze mogą byc wprowdzone do równania 2.1 poprzez szum Langevina. Zatem w układzie atomów o jednakowej masie m, równanie ruchu 2.1, przyjmuje postać

$$m\ddot{\mathbf{r}}_i = -\gamma\dot{\mathbf{r}}_i + \mathbf{F}_{c,i} + \Gamma_i \ . \tag{2.2}$$

gdzie $\mathbf{F}_c = -\nabla_r E_p$ jest wypadkową siłą wynikającą z potencjałów opisujących układ, a γ oznacza współczynnik tarcia. Γ_i odzwierciedla przypadkowe fluktuacje (np. zderzenia z cząsteczkami wody), które balansują dysypację energii spowodowaną tarciem. Dyspersja Γ jest dana rozkładem Gaussa w funkcji temperatury T i opisana relacją

$$\langle\Gamma(0)\Gamma(t)\rangle = 2\gamma k_B T\delta(t), \tag{2.3}$$

gdzie k_B oznacza stałą Boltzmanna, t czas i $\delta(t)$ deltę Diraca.

2.2 Modele gruboziarniste

Modelami gruboziarnistymi nazywamy modele, w których dany układ jest reprezentowany nie przez wszystkie występujące w nim atomy, lecz jedynie poprzez najważniejsze z nich, też lub poprzez „pseudoatomy" reprezentujące całe grupy atomów.

Zastosowanie modeli gruboziarnistych do opisu białek opiera się na przybliżonym opisie aminokwasów z których są one złożone. Każdy aminokwas może być reprezentowany poprzez pojedynczy atom C^{α} [128, 129], lub też — bardziej dokładnie — od dwóch do kilku atomów. W tym drugim przypadku stosuje się kilkanaście podejść. Najczęściej uwzględnia się explicite atomy w łańcuchu głównym, natomiast łańcuch boczny reprezentuje się pojedyńczym atomem C^{β} lub pseudoatomem, którego rozmiar zależy od ilości atomów w grupie bocznej. Rzadko natomiast reprezentuje się grupę boczną kilkoma atomami, gdyż czas obliczeniowy znacznie się wtedy wydłuża i staje się zbliżony do czasów symulacji w modelach pełnoatomowych.

Ponadto, konstruując model gruboziarnisty należy dokonać wyboru wybór pomiędzy ciągłą a dyskretną reprezentacją białka. Modele dyskretne zbudowane są na sieciach, których stopień dokładności pozwala na mniej [130, 131] lub bardziej [132, 133, 134] precyzyjny opis białka. Główną wadą podejścia dyskretnego jest zaburzenie lokalnej geometrii łańcucha głównego, której opis jest ograniczony rozdzielczością wybranej sieci. Opis geometrii białka poprzez sieć wymaga także wykonania symulacji technikami takimi jak Monte Carlo. Natomiast w przypadku modeli ciągłych istnieje możliwość zaadoptowania do opisu ruchu układu typowych technik dynamiki molekularnej, z których korzystamy też w niniejszej pracy.

Użycie modelu gruboziarnistego ogranicza możliwość opisu istniejących oddziaływań w białku, reprezentowanych przez odpowiednie potencjały. W przypadku modeli jednokulkowych stosuje się potencjały, które muszą jednocześnie uwzględniać oddziaływania wodorowe, średnią wartość oddziaływań Van der Waalsa, efekty hydrofobowe itp. Najczęściej w takim przypadku stosuje się potencjał Lennarda-Jonesa. Modele, które w bardziej precyzyjny posób opisują łańcuch główny, umożliwiają zastosowanie różnego rodzaju potencjałów do opisu różnego typu oddziaływań. W przypadku modeli na sieci używa się potencjałów, których kształt w minimum jest raczej płaski i odpowiednio dobrany do parametrów sieci.

Ostatnim elementem konstrukcji modeli gruboziarnistych jest wybór techniki przeszukiwania przestrzeni konformacyjnej. Tutaj także istnieje duża dowolność w wyborze parametrów użytej procedury, z tym że techniki dynamiki molekularnej mogą byc stosowane tylko do modeli ciągłych.

Ze względu na znacznie krótszy czas obliczeniowy i dużą ilość zastosowań (między innymi do symulacji procesów zwijania i rozwijania białek, przewidywania struktur, itd.) istnieje ogromna ilość różnych wariantów modeli, zarówno ciągłych jak i na sieci. Opisanie ich wszystkich nie jest w tym miejscu możliwe, i nie to jest celem niniejszej pracy. Zainteresowanego czytelnika odsyłamy do bardzo dobrej pracy przeglądowej [135].

W niniejszej rozprawie będziemy rozważać dwa rodzaje modeli ciągłych: model Go wraz z dynamiką molekularną i model sieci gaussowskiej, który jest modelem stacjonarnym.

Modelem gruboziarnistym typu Go nazywamy model, w którym minimum energetyczne białka jest determinowane tylko przez kontakty istniejące w białku w stanie natywnym. Obecnie istnieje kilkanaście różnych wariantów modelu Go, które przede wszystkim różnią się doborem potencjałów [17, 18, 19, 20, 21], kryterium wyboru mapy natywnych kontaktów [22, 23, 24] oraz siłą ich oddziaływania [10, 25]. W przypadku modeli zbudowanych tylko z węgli C^{α} najpowszechniej przyjmuje się, że wszystkie oddziaływania w białku są równoważne. Jednakże istnieje też spora grupa modeli, w których rozważa się efektywną energię oddziaływania. Pierwszą efektywną energię oddziaływania wyznaczyli Miyazawa i Jernigan [136] poprzez rozważenie statystycznego prawdopodobieństwa występowania oddziałujących par aminokwasów w białkach. Potencjał ten został wyprowadzony z założenia, iż siła oddziaływań między parą aminokwasów jest proporcjonalna do częstości ich występowania, f, w strukturach białek, danej rozkładem Boltzmana $f \sim e^{-E}$. Rozważenie wszystkich struktur białek z PDB doprowadziło ich do skonstruowania macierzy o rozmiarze 20×20, określającej energię oddziaływań dla par i, j o 210 różnych parametrach ϵ_{ij}^{MJ}. Alternatywny sposób konstrukcji efektywnego potencjału został zaproponowany przez [137].

Rozdział 3

Modele gruboziarniste typu Go

Z przytoczonych w poprzednim rozdziale przykładów gruboziarnistych modeli białek najpowszechniej stosowane są modele typu Go, ze względu na ich prostą implementację oraz specyficzność względem białek. Ponadto niesamowity postęp przy wyznaczeniu struktur natywnych coraz większych i jeszcze bardziej złożonych układów molekularnych, których badanie modelami pełnoatomowymi jeszcze długo nie będzie możliwe, uzasadnia rozwijanie tej dziedziny.

W niniejszyn rozdziale, korzystając z obecnie istniejących wariantów modelu Go, konstruujemy jego nowe formy, tak aby jak najdokładniej opisać krajobraz energetyczny białek podczas rozciągania. W sumie otrzymujemy wyniki dla ponad 61 różych modeli. W ogólności model Go konstruuje się tak, aby minimum globalne zastosowanego w nim potencjału było zgodne z eksperymentalnie wyznaczoną strukturą natywną białka [13, 12]. Zgodnie z dyskusją z rozdziału 2.2, nie istnieje jedna oczywista droga konstrukcji modelu. Jednakże, opierając się na założeniu, że energia molekuły w modelu Go jest zbudowana na podstawie stanu natywnego, należałoby przyjąć, iż ten model powinien najlepiej odzwierciedlać rzeczywisty ruch właśnie w jego pobliżu. Z tego powodu podczas konstrukcji modelu należy przede wszystkim brać pod uwagę własności białka w stanie przejściowym, szybko osiągalnym podczas rozwijania. Konstruując nowe formy modelu Go, skupimy się na własnościach białek podczas rozciągania.

W pierwszej części tego rozdziału rozważamy najprostszą formę redukcji stopni swobody, reprezentując każdy aminokwas w białku jego centralnym atomem węgla. Następnie konstruujemy nowe warianty modelu Go o parametrach: mapy kontaktów (4 warianty), potencjały dla oddziaływań natywnych (6 wariantów), potencjały związane z geometrią łańcucha głównego (2 warianty), i skale energii (7 wariantów) które mogą być jednorodne i niejednorodne. Potencjał w skalach niejednorodnych wywodzi się ze specyficzności lub własności związanych z geometrią grup bocznych. W drugiej części konstruujemy model Go z grupami bocznymi $C^{\alpha-\beta}$, w którym każdy aminokwas (oprócz glicyny) jest reprezentowany dwoma atomami. W następnym rozdziale, na podstawie danych doświadczalnych, dokonamy wyboru optymalnej wersji modelu Go do rozciągania białek, a następnie ocenimy jego stosowalność do zwijania białek.

Nowe formy potencjałów wprowadzone w tym rozdziale zostały także opisane w pracach [I] i [VIII].

3.1 Modele Go typu C^α

W modelu Go typu C^α, który nazywamy także modelem jednokulkowym, każdy aminokwas wchodzący w skład białka jest reprezentowany jedynie poprzez jego węgiel C^α. Współrzędne tych atomów są określone doświadczalnie i odczytujemy je z bazy danych PDB w której są one zapisane. Następnie wyznaczamy *mapę kontaktów*, czyli podział oddziaływań w białku na natywne i nienatywne. Efektywną energię E_p molekuły konstruuje się z sumy potencjałów sił fizycznych określających oddziaływania wiążące (atomowe) i niewiążące (np. van der Waalsa, elektrostatyczne czy mechaniczne wynikające z odchylenia od zmierzonych wartości eksperymentalnych kątów płaskich i torsyjnych). Dla tak zbudowanej energii potencjalnej rozwiązuje się następnie równania ruchu dla węgli C^α, w wyniku których otrzymujemy dynamiczny obraz ruchu białka w zadanym polu siłowym.

W modelu Go energia potencjalna E_p molekuły zbudowanej z N węgli C^α o położeniach $\{\mathbf{r}_i\}$, $i = 1, 2 \ldots N$, dana jest wzorem:

$$E_p(\{\mathbf{r_i}\}) \;=\; V^{BB} + \; V^S \; + \; V^{NAT} \; + \; V^{NON}. \tag{3.1}$$

W powyższym równaniu człon V^{BB} oznacza potencjał harmoniczny utrzymujący kolejne atomy C^α we wzajemnej odległości $d_0 = 3.8$ Å i dany wzorem

$$V^{BB} = \sum_{i=1}^{N-1} k_1 (r_{i,i+1} - d_0)^2, \tag{3.2}$$

gdzie $r_{i,i+1} = |\mathbf{r}_i - \mathbf{r}_{i+1}|$ oznacza odległość pomiędzy atomami C^α o numerach i oraz $(i + 1)$, natomiast $k_1 = 100\epsilon$ Å^{-2}. ϵ oznacza parametr energii, którego dokładna definicja będzie podana poniżej. Człon V^S kształtuje lokalną sztywność łańcucha głównego. Rozważamy tu dwie jego formy: jedną uwzględniającą kąty płaskie i dwuścienne [138, 24, 139] i drugą dużo szybszą numerycznie, zależną od natywnej chiralności aminokwasów [I] i [23]. Ostatnie dwa człony reprezentują pozostałe oddziaływania pomiędzy atomami C^α i zależą od wyboru mapy kontaktów. Rozważamy siedem sposobów konstrukcji mapy kontaktów i pokazujemy, że uzyskiwane wyniki są czułe na wybór mapy. Potencjał V^{NON} zapewnia efekt wyłączonej objętości określony parametrem

d_{cut}, który w szczególności nie dopuszcza by atomy się wzajemnie przenikały. Człon ten opisuje oddziaływania aminokwasów, które nie zostały uwzględnione w mapie kontaktów i dany jest członem potencjału Lennarda-Jonesa,

$$V^{NON} = \sum_{i<j} V_{ij}^{NON}, \tag{3.3}$$

$$V_{ij}^{NON} = \begin{cases} 4\epsilon^o \left[\left(\frac{\sigma_0}{r_{ij}}\right)^{12} - \left(\frac{\sigma_0}{r_{ij}}\right)^{6} \right] + 1 & , r_{ij} < d_{cut} \\ 0 & , r_{ij} \geq d_{cut}, \end{cases} \tag{3.4}$$

gdzie dla odległości mniejszej niż d_{cut}=4Å potencjał jest czysto odpychający, zaś dla większych odległości wynosi zero. Ponadto $\sigma_0 = 2^{-1/6} \cdot d_{cut}$. V^{NAT} opisuje oddziaływania pomiędzy parami aminokwasów z mapy kontaktów, których siła zależy od parametrów energii E. Paremetry energii wyznaczają głębokość potencjału i mają zasadniczy wpływ na własności układu. W niniejszej pracy rozważamy 6 form potencjału natywnego i 6 skali energii, uwzględnając także jednorodną skalę energii $\mathcal{E}_{ij} = \epsilon$.

Sposób oznaczenia modeli

Rozważane w niniejszej pracy modele określone są przez kilka podstawowych atrybutów, które oznaczamy w następujący sposób:

$$model = \left\{ V^{NAT}, S, \mathcal{M}, E \right\}.$$

Wymienione człony oznaczają wybór funkcjonalej formy potencjałów V^{NAT} dla oddziaływań natywnych, potencjałów S do opisu lokalnej sztywności łańcucha, sposobów konstrukcji mapy kontaktów $\mathcal{M}$, i zbioru wartości parametrów energii do konstrukcji skali energii, E. Dyskutowane formy tych atrybutów są wyszczególnione w tabeli 3.1, a ich konstrukcję opisujemy poniżej. Umożliwiają one skonstruowanie 504 różnych modeli.

3.1.1 Potencjały V^{NAT} dla oddziaływań natywnych $C^\alpha - C^\alpha$

W niniejszej pracy rozważamy 6 rodzajów oddziaływań V^{NAT} pomiędzy natywnymi kontaktami. W szczególności,

$$V^{NAT} = \sum_{i,j} V_{ij}^{NAT},$$

gdzie V_{ij}^{NAT} oznacza potencjał pomiędzy atomami C_i^α i C_j^α, natomiast sumowanie przebiega po wszystkich parach (i, j), które są określone przez mapę kontaktów $\mathcal{M}$. W

Tablica 3.1: Oznaczenia parametrów użytych do konstrukcji modeli Go: mapy kontaktów $\mathcal{M}$, potencjały natywne V^{NAT}, potencjały określające sztywność łańcucha głównego S, skale energii $\mathcal{E}$.

parametr	
V^{NAT}	$\in \{\ 6-12,\ 6-12_{const},\ 10-12,\ (6-12, 10-12),\ 6-10-12,\ 6-12_{exp}, Morse\ \}$
S	$\in \{\ C,\ A\ \}$
$\mathcal{M}$	$\in \{\ M2,\ M3,\ M4,\ M2c75,\ M3c7,\ M3c75,\ CSU\ \}$
$\mathcal{E}$	$\in \{\ E^{o},\ E^{L1,\,L2},\ E^{T1,\,T2,\,T3},\ E^{\nu 1,\,\nu 2},\ E^{HB-MJ},\ E^{HB-HH}\ \}$

każdym przypadku V_{ij}^{NAT} jest dobrany tak, aby jego minimum odpowiadało odległości r_{ij}^{n} pomiędzy atomami C_i^{α} i C_j^{α} w stanie natywnym, zapisanej w pliku PDB. Ponadto V_{ij}^{NAT} jest proporcjonalne do parametru E_{ij}, czyli tzw. skali energii. Przypomnijmy, iż mapa kontaktów dzieli wszystkie oddziaływania na natywne i nienatywne. Podczas symulacji natywny kontakt może być zrywany i tworzony. Istnienie dynamiczne kontaktu w danym V^{NAT} podczas dynamiki molekularnej jest określone parametrem odległości progowej $l_{e,ij}$. Jeśli $r_{ij} < l_{e,ij}$ przyjmujemy, że kontakt natywny jest obecny dynamicznie. Wykresy potencjałów są pokazane na rysunku 3.1.

Potencjał $6-12$

Symbol V^{6-12} określa potencjał Lennarda-Jonesa, dany równaniem

$$V_{ij}^{6-12} = 4E_{ij}\Big[\Big(\frac{\sigma_{ij}}{r_{ij}}\Big)^{12} - \Big(\frac{\sigma_{ij}}{r_{ij}}\Big)^{6}\Big], \tag{3.5}$$

gdzie r_{ij} jest odległością pomiędzy C^{α} atomami z aminokwasów i oraz j. Parametr odległości σ_{ij} wynosi $\sigma_{ij} = r_{ij}^{n}/\sqrt[6]{2}$, gdzie r_{ij}^{n} jest odległością między atomami w stanie natywnym. W tym przypadku $l_{e,ij} = 1.5 \cdot \sigma_{ij}$.

Potencjał $6-12$ ze stałą szerokością

Alternatywnym spojrzeniem na charakter oddziaływania kontaktów natywnych jest przyjęcie potencjału, którego kształt nie zależy od natywnej odległości pomiędzy $C^{\alpha}-C^{\alpha}$ a jedynie od położenia minimum. Potencjał ten jest dany wzorem [140]

$$V_{const,ij}^{6-12} = 4E_{ij}\Big[\Big(\frac{s}{r_{ij}-\sqrt[6]{2}(r_{ij}^{n}-s)}\Big)^{12} - \Big(\frac{s}{r_{ij}-\sqrt[6]{2}(r_{ij}^{n}-s)}\Big)^{6}\Big], \tag{3.6}$$

gdzie parametr s wyzncza stałą szerokość potencjału. Do wyznaczenia wartości s użyliśmy 500 różnej klasy niehomologicznych białek. Otrzymaliśmy, iż średnia odległość dla natywnych kontatków pomiędzy C_i^α i C_j^α wynosi $s \simeq 4$Å. Bardziej szczegółowe podejście do wyznaczenia wartości s polega na podzieleniu natywnych kontatków na oddziaływania wodorowe i pozostałe, dla których otrzymujemy odpowiednio dwie wartości $s' \simeq 2.4$Å i $s'' \simeq 5.6$Å. W tym przypadku $l_{e,ij} = 1.5 \cdot s + \sqrt[6]{2}(r_{ij}^n - s)$.

Potencjał $6 - 10 - 12$

Powyższe potencjały opisują sytuacje gładkiego wejścia i wyjścia z obszaru natywnego kontaktu. Jednakże czasami podczas rozciągania białek wiązanie wodorowe przed ostatecznym rozerwaniem przechodzi przez *stan pośredni*. Stan ten charakteryzuje się nadal istnieniem kontaktu, jednakże za pośrednictwem cząsteczki wody. Istnienie stanu pośredniego uwzględniamy poprzez wprowadzenie dodatkowego minimum, i w konsekwencji dodatkowej bariery energetycznej. Taka forma potencjału jest obecnie często używana [25, 141, 11, 142] i dana jest wzorem

$$V_{ij}^{6-10-12} = 4E_{ij}\Big[13\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - 18\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{10} + 4\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\Big], \tag{3.7}$$

gdzie $\sigma_{ij} = r_{ij}^n/\sqrt[6]{2}$. Mankamentem tego potencjału jest fakt, iż położenie drugiego minimum zależy od pierwszego, tak więc i szerokość bariery jest zmienna. W symulacjach zwijania białek bariera ta ma interpretację dodatkowego kosztu energii, jaką para oddziałujących aminokwasów musi „zapłacić" przed ostatecznym powstaniem kontaktu [142], jednakże w przypadku jego ustalenia się jest on chroniony przed rozerwaniem. W przypadku tego potencjału zakładamy, iż natywny kontakt istnieje, gdy odległość r_{ij} jest mniejsza niż położenie maksimum bariery, $l_{e,ij} = 1.4 \cdot \sigma$.

Potencjał $6 - 12$ **z dodatkowym minimum w odległości 3 Å**

Kolejny dyskutowany przez nas potencjał jest podobny do poprzedniego, z tym zasadniczym wyjątkiem, iż stan pośredni jest realizowany przez funkcję o stałej szerokości i jej minimum jest w odległości odpowiadającej dokładnie rozmiarowi jednej cząsteczki wody pomiędzy donorem a akceptorem. Wcześniejsze sposoby realizcji tych własności były modelowane dość skomplikowanym potencjałem, nie adekwatnym do prostego modelu Go. Prostszą formę zaproponowano w pracy [140], którą my następnie zmodyfikowaliśmy tak, aby była odpowiednia dla rozciągania białek. Funkcja ta dana jest wzorem

$$V_{exp,ij}^{6-12} = V_{const,ij}^{6-12} + A\, exp(-(r_{ij} - \sigma_{ij} + D\sqrt[6]{2} - 9)^2/D^2B), \tag{3.8}$$

gdzie parametry $A = 0.3, B = 0.005, D = 6$ kontrolują kształt potencjału. W szczególności potencjał ten tworzy barrierę energetyczną której maksimum wynosi -0.01ϵ

oraz drugie minimum w odległości r'_{ij}=σ_{ij}+3 Å. Podobnie jak w powyższym potencjale, przyjmujemy $l_{e,ij}$=σ_{ij}+2.32, co odpowiada odległości r_{ij} mniejszej niż położenie maksimum bariery.

Potencjał $10 - 12$

Kolejnym potencjałem jest potencjał $10 - 12$ dany wzorem

$$V_{ij}^{10-12} = E_{ij}\Big[5\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - 6\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{10}\Big], \tag{3.9}$$

gdzie σ_{ij} odpowiada odległości natywnej. Potencjał ten jest często stosowany do opisu oddziaływań wodorowych [24], ze względu na to, iż człon przyciągający jest wyraźnie silniejszy na małych odległościach niż w przypadku oryginalnego potencjału Lennarda-Jonesa. Porównanie kształtu funkcji V^{10-12} z V^{6-12} jest pokazane na rysunku 3.1. Zgodnie z pracą [24] przyjmujemy $l_{e,ij} = 1.2 \cdot r_{ij}^n$. Podkreślmy jednakże, iż wybór konkretnej wartości $l_{e,ij}$ nie wpływa zasadniczo na proces zwijania w tym przypadku, zgodnie z wynikami pracy [143].

Potencjał Morse'a

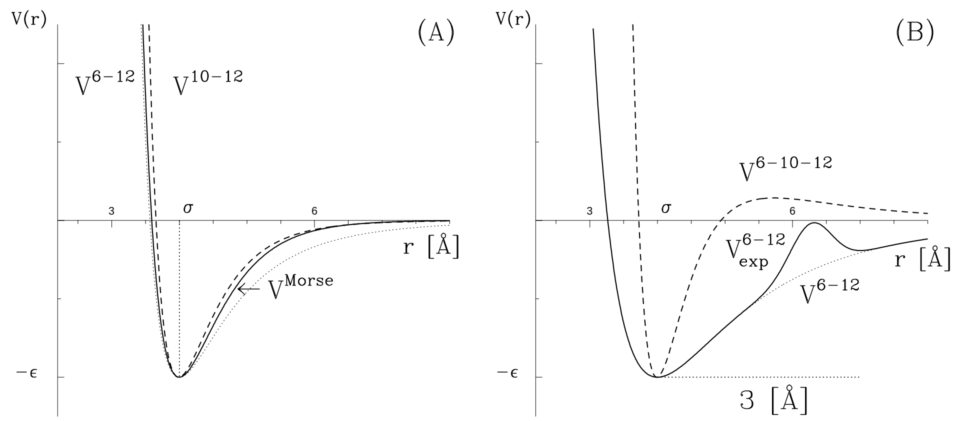
Potencjał Morse'a dany jest równaniem,

$$V_{ij}^{Morse} = E_{ij}\Big[\Big(1 - e^{-\alpha(r_{ij}-\sigma_{ij})}\Big)^2 - 1\Big], \tag{3.10}$$

gdzie α determinuje kształt potencjału. Istnieje kilka sposobów wyznaczania α, których wybór zależy od własności badanego układu. W niniejszej pracy parametr ten został dobrany tak, aby kształt potencjału Morse'a w minimum był zgodny z oryginalnym potencjałem Lennarda-Jonesa dla najbardziej typowej odległości C_i^α-$C_j^\alpha \simeq$ 4Å. $\alpha = 1.7 r_{ij}^n$. Zauważmy, iż tak dobrany kształt potencjału Morse'a jest podobny do potencjału 6-12 ze stałą szerokością, z wyjątkiem tego, iż pierwszy charakteryzuje się silniejszym przyciąganiem. W tym przypadku natywny kontakt jest obecny dla $l_{e,ij} = 3.4 \cdot r_{ij}^n$ i odpowiada wartości potencjału Lennarda-Jonesa w punkcie przegięcia.

3.1.2 Potencjały S związane ze sztywnością łańcucha głównego

Jak opisaliśmy w rozdziale 1, aminokwasy w stanie natywnym są względem siebie ułożone w szczególny sposób, nazywany strukturą drugorzędową. Istnienie struktury drugorzędowej narzuca więzy geometryczne na łańcuch główny, takie jak prawoskrętna chiralność oraz odpowiednie wartości kątów płaskich i dwuściennych. W niniejszej



Rysunek 3.1: Kształty potencjałów użyte do opisu natywych kontaktów dla odległości natywnej σ=4Å. Panel (A) potencjał V^{10-12} (linia przerywana), potencjał V^{Morse} (linia ciągła). Panel (B) potencjał $V^{6-10-12}$ (linia przerywna), potencjał V_{exp}^{6-12} (linia ciągła) z drugim minium w odległości 3Å od pierwszego. Oryginalny potencjał Lennarda-Jonesa V^{6-12} jest zaznaczony kropkowaną linią na obu panelach. Minimum dla potencjałów V^{6-12} i $V^{6-10-12}$ jest przesunięte do r=4Å.

pracy rozważamy dwa sposoby opisu lokalnej sztywności łańcucha, uwzględniając własności chiralne lub kątowe, odpowiednio poprzez potencjały chiralnościowy (C) lub potencjał kątowy (A).

Potencjał kątowy

Potencjał ten dany jest wzorem [138]

$$V^A = \Upsilon\Big[\sum_{i=1}^{N-2} K_\theta(\theta_i-\theta_i^n)^2+\sum_{i=1}^{N-3}\Big(K_\phi^1(1-cos(\phi_i-\phi_i^n))+K_\phi^3(1-cos3(\phi_i-\phi_i^n))\Big)\Big]\ , \quad (3.11)$$

gdzie wartości stałych $K_\theta = 20\epsilon/(rad)^2, K_\theta^1 = \epsilon, K^3 = 0.5\epsilon$ przyjęliśmy takie same jak w pracy [24]. Stała Υ jest parameterem, który kontroluje całkowitą siłę potencjału. O ile wprost nie stwierdzamy inaczej, w niniejszej rozprawie przyjmujemy Υ=1, zgodnie z wieloma innymi pracami, np. [24]. Wartości kątów płaskich i dwuściennych wyznczane są na podstawie trzech i czterech kolejnych reziduów wzdłuż sekwencji. Typowym podejściem stosowanym przy konstrukcji mapy natywnych kontaktów, gdy użyty jest V^A, jest nieuwzględnienie w niej kontaktów $i, i+2$ oraz $i, i+3$, gdyż mogłyby one przeciwdziałać V^A.

Potencjał chiralnościowy

Zauważmy, iż użycie omówionego powyżej potencjału V^A podczas dynamiki molekularnej wymaga numerycznego liczenia pochodnych. Dużo szybszym numerycznie sposobem nałożenia więzów na konformację łańcucha głównego (zwłaszcza dla dużych układów) jest potencjał chiralnościowy V^C, który faworyzuje natywną wartość lokalnej chiralności. Istnieje kilka sposobów uwzględnienia chiralnej natury białek [23, 144]. Jednakże na podstawie naszych wyników z pracy [I] chiralność wydajniej jest przedstawić jako „karę” za wychylenie poza stan równowagowy odpowiadający natywnej konformacji. Postać ta jest dana wzorem

$$V^C = \sum_{i=2}^{N-2} \frac{1}{2}\kappa(C_i \ - \ C_i^n)^2 \ , \ \ C_i = \frac{(\mathbf{w}_{i-1} \times \mathbf{w}_i) \cdot \mathbf{w}_{i+1}}{d_0^3}, \tag{3.12}$$

gdzie parametr C_i^n określa chiralność aminokwasu i w stanie natywnym, a $d_0 = |\mathbf{w}_i|$ jest odległością między kolejnymi C$^\alpha$ atomami, , $\mathbf{w}_i = \mathbf{r}_{i+1} - \mathbf{r}_i$. Przyjmujemy ponadto κ=1, zgodnie z wynikami pracy [I]. Ponadto zauważmy, iż V^C jest używany tylko wtedy, gdy chiralność aminokwasów zmienia znak, tak więc w tym przypadku krótkozasięgowe kontakty typu $i, i+3$ są nadal obecne w mapie kontaktów (przeciwnie do potencjału V^A).

Analityczne porównanie potencjałów C i A

Wybór optymalnego potencjału chiralnościowego jest bardzo ważnym elementem prostych modeli Go, gdzie cały aminokwas jest reprezentowany tylko jednym węglem C^α. W rozdziale 5 przedyskutujemy bezpośredni wpływ potencjałów na własności kinetyczne białek, a obecnie skupimy się na analitycznym wykazaniu, iż potencjały te powinny działać podobnie, z tym wyjątkiem, iż V^A zapewnia dodatkową sztywność lokalną.

Zauważmy, iż potencjał V^C jest funkcją trzech kolejnych wektorów jednostkowych $\mathbf{w}_i$, które łączą cztery kolejne atomy C^α. Następnie oznaczmy przez ϕ kąt dwuścienny pomiędzy płaszczyzną wyznaczoną przez wektory $\mathbf{w}_{i-1}$ and $\mathbf{w}_i$ i płaszczyzną powstałą z wektorów $\mathbf{w}_i$ and $\mathbf{w}_{i+1}$. Wychylenie od wartości natywnej ϕ^n oznaczmy przez $\delta\phi = \phi - \phi^n$. W ogólnej sytuacji kąt dwuścienny można wyrazić poprzez

$$\operatorname{tg}\phi = \frac{C_i}{D_i} \ \ , \tag{3.13}$$

gdzie C_i dane jest równaniem (3.12) oraz

$$D_i = (\mathbf{w}_{i-1} \times \mathbf{w}_i) \cdot (\mathbf{w}_i \times \mathbf{w}_{i+1})/d_0^4 \ \ . \tag{3.14}$$

Rozważmy obecnie potencjał V^A dany wzorem (3.11). Założenie małych wychyleń od położenia równowagi $\delta\phi$ pozwala nam rozwinąć cosinusy zawiązane z kątami dwuściennymi w potencjale V^A dla danego i-tego reziduum, jako

$$V_i^{dwu} \simeq \frac{\Upsilon(K_\phi^1 + 9K_\phi^3)}{2}\delta\phi^2 + \mathcal{O}\big((\delta\phi)^3\big). \tag{3.15}$$

Człon V^{dwu} z kolei także możemy wyrazić jako funkcję $\mathbf{w}_i$, zauważając iż dla małych wartości $\delta\phi$

$$\delta\phi \simeq \operatorname{tg}\delta\phi = \frac{\operatorname{tg}\phi - \operatorname{tg}\phi^n}{1 + \operatorname{tg}\phi \cdot \operatorname{tg}\phi^n} \simeq \frac{C_i - C_i^n}{D_i^n + C_i C_i^n / D_i} + \ldots, \tag{3.16}$$

gdzie D_i^n oznacza natywną wartość D_i i kropki oznaczają kolejne człony wynikające z rozwinięcia D_i wokół jego wartości natywnej. Wstawiając otrzymaną postać $\delta\phi$ do (3.15) znajdujemy, że V_i^{dwu} składa się także z członu, który w pierwszym przybliżeniu możemy zidentyfikować jako przeskalowany potencjał chiralnościowy (3.12)

$$V_i^{dwu} \simeq \frac{\Upsilon(K_\phi^1 + 9K_\phi^3)}{2\big(D_i^n + (C_i^n)^2/D_i^n\big)^2}\Big(C_i - C_i^n\Big)^2 + \ldots$$

W ten sposób pokazaliśmy, iż potencjał V^C jest głównie odpowiedzialny za część dwuścienną potencjału V^A, lecz nie uwzględnia dodatkowgo członu związanego z kątami płaskimi. Tak więc V^A możemy zapisać jako

$$V^A \simeq V^C + \Upsilon K_\theta \sum_i (\delta\theta_i)^2 \,, \tag{3.17}$$

przynajmniej dla małych odchyleń kątów. Podsumowując, użycie potencjału V^A prowadzi do większej lokalnej sztywności niż V^C. Powyższy wniosek może sugerować, iż optymalny numerycznie potencjał mógłby powstać z potencału V^C po uwzględnieniu analitycznej formy potencjału płaskiego danej wyrażeniem (3.17).

3.1.3 Mapy kontaktów $\mathcal{M}$

Wybór mapy kontaktów $\mathcal{M}$ jest podstawowym i kluczowym elementem w modelu Go. W niniejszej pracy rozważamy trzy techniki wyznaczenia mapy kontaktów.

Technika I

Najprostsza technika polega na wprowadzeniu jednorodnego promienia odcięcia R_c dla odległości r_{ij}, węgli C_i^α, które nie są sąsiadmi wzdłuż sekwencji. Korzystając z wcześniejszych prac [21, 145, 146, 10], które pokazały iż najlepsze wyniki otrzymuje się dla R_c=7 Å lub R_c=7.5 Å, my także rozważamy te dwie wartości. Odpowiadające tym parametrom mapy kontaktów oznaczamy przez $Mc7$ i $Mc75$. Zauważmy, iż technika ta ma wady i zalety. Zaletą jest jej stosowalność do niekompletnych struktur białek. Wadami są: możliwość uwzględnienia w mapie kontaktów, które są w mniejszej odległości niż zasięg R_c, a nie tworzą faktycznie kontaktu; nie uwzględnienie oddziaływań dalekozasięgowych, które mogą istnieć nawet do odległości 12 Å [23] (np. bywają one tworzone przez grupy boczne dużych aminokwasów, poprzez oddziaływania hydrofobowe).

Technika II

Nieco bardziej realistyczny sposób konstrukcji mapy kontaktów został zaproponowany przez Tsai [147]. Technika ta polega na określeniu natywnego kontaktu na podstawie wszystkich ciężkich atomów, których promienie van der Waalsa są dodatkowo powiększone o czynnik 1.24 [147]. Powiększenie uwzględnia istnienie oddziaływań przyciągających. Przyjmujemy, iż dwa aminokwasy i, j tworzą kontakt natywny, jeśli istnieje chociaż jedno przykrywanie się pomiędzy efektywnymi kulami tych dwóch aminokwasów. Mapę kontaktów zbudowaną zgodnie z tą procedurą oznaczamy poprzez M. Mapa kontaktów wyznaczona techniką II była także użyta w pracach [129, 120, 148].

Technika III

Zastosowanie trzeciej techniki wymaga posiadania pełnej informacji o strukturze białka, włącznie z położeniem atomów wodoru. Kostruuje się ją korzystając ze specyficznych własności geometrycznych, które determinują różnego rodzaju wiązania, rozważane na poziomie elektronowym. Pozwala to na identyfikację kontaktów natywnych i przede wszystkim określenie rodzaju kontaktu (np. wiązanie wodorowe, hydrofobowy-hydrofobowy, aromatyczny-aromatyczny, aromatyczny-polarny etc.), który najbardziej stabilizuje strukturę. Technika ta jest między innymi użyta w oprogramowaniu *Contacts of structural unit, CSU* [149], w którym ponadto zastosowano automatyczną analizę przestrzeni międzyatomowej do stwierdzenia istnienia i rodzaju kontaktu. W niniejszej pracy nie konstruujemy oddzielnie mapy kontaktów techniką III, lecz korzystamy z wyników oprogramowania CSU.

Powyżej pokazaliśmy, iż technika II daje lepsze wyniki niż I. Obecnie porównujemy technikę II z III. W tym celu wyznaczyliśmy mapy kontaktów technikami II i III dla czterech różnych klas białek i nałożyliśmy odpowiednie ich pary na siebie. W wyniku tego uzyskaliśmy bardzo dobre przykrywanie się pomiędzy mapami wyznaczonymi technikami efektywnych sfer van der Waalsa i techniką *CSU*, z wyjątkiem kontaktów $i, i+2$. Te krótkozasięgowe kontakty istnieją znacznie częściej w kryterium efektywnego przykrywania. Natura tych nadmiarowych kontaktów jest różna od pozostałych kontaktów natywnych, gdyż są to oddziaływania van der Waalsa, które są o rząd wielkości słabsze niż wiązanie wodorowe. Z tego powodu wprowadzamy dodatkowe kryterium dla technik I i II, oznaczane poprzez dodatkową cyfrę 2, 3 lub 4 umieszczaną bezpośrednio po M w symbolu mapy kontaktów, np $M2c75$. W mapach z cyfrą 2, wszystkie kontakty wykryte daną technika istnieją. Cyfra 3 oznacza, że kontakty $i, i+2$ są usunięte z mapy. W przypadku map z cyfrą 4, kontakty $i, i+2$ oraz $i, i+3$ są wyłączone z grona natywnych oddziaływań; ten przypadek rozważamy tylko w obecności potencjału kątowego A, który opisuje siły związane z lokalną sztywnością łańcucha głównego. Rozważane powyżej eliminacje kontaktów są ważne z punktu widzenia białek helikalnych, natomiast ich wpływ na wyniki symulacji dla białek klasy β jest znacznie mniejszy.

Podsumowując, rozpatrywane przez mapy kontaktów $\mathcal{M}$ oznaczone są jednym z następujących symboli:

$$\mathcal{M} \in \{\ M2,\ M3,\ M4,\ M2c75,\ M3c7,\ M3c75,\ CSU\ \},$$

zgodnie z tabelą 3.1.

3.1.4 Skale energii $\mathcal{E}$

Skalą energii E_{ij} nazywamy zbiór parametrów określających siłę oddziaływania między parami aminokwasów i, j w białku. W niniejszej rozprawie rozważamy jednorodne oraz niejednorodne skale energii. W pierwszym przypadku E_{ij}=$const$. Natomiast niejednorodne skale energii konstruujemy na podstawie różnych własności geometrycznych, topologicznych, chemicznych i statystycznych aminokwasów, i w tym wypadku amplituda E_{ij} zależy od rodzaju natywnego kontaktu.

Jednorodne skale energii

W najprostszym przypadku zakładamy jednorodny parameter energii, tak więc każdy natywny kontakt oddziałuje z taką samą siłą, E_{ij}=ϵ i takim samym potencjałem. W tym przypadku energia w stanie natywnym ma wartość liczby natywnych kontaktów.

Kwazi-jednorodną formę jednorodnej mapę otrzymujemy po przyjęciu potencjału $10-12$ dla kontatków wodorowych i potencjału $6-12$ (lub $6-12_{const}$) dla pozostałych natywnych kontatków przy jednoczesnym utrzymaniu takiej samej skali energii E_{ij}=$const$ dla wszystkich potencjałów. Cechą tych skal energii jest faworyzowanie kontaktów wodorowych, które są silniej przyciągane niż pozostałe dla tej samej odległości.

Kwazi-jednorodne skale będziemy oznaczać $(10-12, 6-12)$ i $(10-12, 6-12_{const})$, gdzie pierwszy człon odnosi się do kontatków wodorowych, a drugi do pozostałych.

Wartość parametru energii ϵ jest proporcjonalna do średniej energii oddziaływania między dwoma aminokwasami w rzeczywistym białku, która wynosi około 2 kcal/mol. W wyniku zamiany jednostek 1kcal/mol na Kelvin otrzymujemy, iż 1ϵ/k$_B$=1000K. W temperaturze pokojowej zredukowana temperatura $\tilde{T} = k_B T/\epsilon$ wynosi 0.3.

Niejednorodne skale energii

W niniejszym paragrafie wprowadzamy niejednorodne skale energii, poprzez określenie wartości parametrów energii dla różnych grup kontaktów. Podkreślmy, że w przypadku kiedy nowa amplituda energii nie jest zdefiniowana explicite, kontakt natywny oddziałuje poprzez jednorodny parametr energii ϵ.

Skale energii E^L i E^G są skonstruowane na podstawie obserwacji wyników z symulacji rozciągania białek z jednorodną wartością energii. Skale te osłabiają oddziaływania dalekozasięgowe. Idea wprowadzenia skali E^ν wywodzi się z prac [150, 151], gdzie kluczowym elementem jest rzeczywista ilość oddziaływań między parą aminokwasów. Szablon skali E^{HH} zaczerpnęliśmy z pracy [152] i faworyzuje ona oddziaływania hydrofobowe. Skala E^{HO} faworyzuje oddziaływania wodorowe [153]. Skala E^{MJ} powstała na podstawie statystyki występowania danej pary aminokwasów w białku opracowanej przez S. Miyazawa i R. L. Jernigana [136]. Połączenie skal E^{HO} i E^{MJ} jest inspirowane pracą [25], a połączenie skal E^{HO} i E^{HH} wynikło z obserwacji zachowania się poszczególnych grup kontaktów podczas symulacji rozciągania.

Skala energii E^L_{ij}

Rozważenie w przypadku dalekozasięgowych kontaktów krzywych Lennarda-Jonesa dla odległości faktycznej, z której pochodzi kontakt i jej przeskalowanie do położenia węgli C^α pokazuje, iż w pierwszym przypadku siła oddziaływania między i, j jest znacznie słabsza niż w drugim. Skala E^L jest skonstruowana tak, aby gładko wytłumić dalekozasięgowe kontakty powyżej parametru odcięcia m. Poniżej m kontakty oddziałują parametrem energii E^L=ϵ. W przedziale pomiędzy m a 12Å energia jest dana wzorem [152]

$$E^L_{ij} = \epsilon^O_{ij} \cdot \left(1 - \frac{r_{ij} - m^2}{(m+n)^2 - m^2}\right), \tag{3.18}$$

gdzie n jest parametrem zależnym od doboru m. Rozważamy dwa przypadki: m=8.5Å, n=3.5Å i m=10.6Å, n=1.4 odpowiadające skalom energii o oznaczeniach E^{L1}_{ij}, E^{L2}_{ij}. Dla odległości natywnej $r^n_{ij} \geq$12Å, E^L_{ij}=0.

Skala energii E_{ij}^T

Skala E_{ij}^T osłabia oddziaływania dalekozasięgowe na podstawie geometrii aminokwasów i specyficzności w naturze przykrywania się efektywnych atomów w stanie natywnym. Rozważamy trzy sytuacje oznaczone $\alpha\alpha$, $\alpha\beta$ i $\beta\beta$ odpowiadające następującym rodzajom oddziaływań. Sytuacja typu $\alpha\alpha$ często odpowiada istnieniu kontaktu wodorowego i ma miejsce gdy istnieje chociaż jedno przykrywanie pomiędzy atomami w łańcuchu głównym z aminokwasów i i j. Jeśli takie przykrywanie nie istnieje, a występuje chociaż jedno oddziaływanie między atomami z łańcucha głównego a grupą boczną, to odpowiada to sytuacji drugiej, $\alpha\beta$. Kontakty typu $\beta\beta$ powstają gdy aminokwasy są w kontakcie tylko poprzez grupy boczne. Sytuację trzecią intrepretujemy jako kontakty dalekozasięgowe lub hydrofobowe.

Rozważamy trzy skale energii $E_{ij} = E_{ij}^{T1}, E_{ij}^{T2}, E_{ij}^{T3}$, na podstawie trzech sytuacji opisanych powyżej na różne sposoby.

$$E_{ij}^{T1,T2,T3} = \begin{cases} \epsilon & \epsilon & \epsilon & \text{dla } \alpha\alpha \\ \epsilon & \epsilon & \epsilon * 0.75 & \text{dla } \alpha\beta, \beta\alpha \\ \epsilon * 0.75 & \epsilon * 0.9 & \epsilon * 0.9 & \text{dla } \beta\beta \end{cases} \tag{3.19}$$

Trzy kolumny odpowiadają trzem skalom energii i oddziaływaniu między atomami $C_i^\alpha - C_j^\alpha$ atomami. Skale te redukują rolę grup bocznych w stosunku do oddziaływań wodorowych.

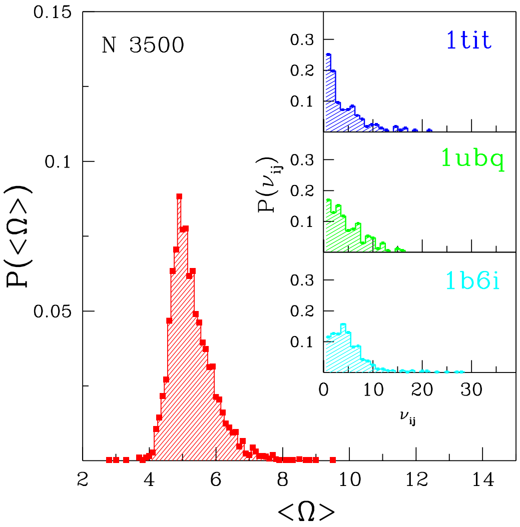
Na podstawie skal energii E_{ij}^L i E_{ij}^T konstruujemy kolejną skalę energii, w której łączymy naraz skalę E^{T1} lub E^{T2} lub E^{T3} ze skalą E^{L1}.

Skala energii E^ν

Skala energii E^ν jest najpowszechniej stosowaną techniką do wyznaczenia niejednorodnego zbioru parametrów energii. W modelu tym przyjmujemy wartości energii oddziaływania proporcjonalnie do liczby przykrywań ν_{ij} między ciężkimi atomami w parze aminokwasów z natywnego kontaktu i, j i odpowiednio normuje [150, 151]. Rysunek 3.2 pokazuje rozkłady ν_{ij} dla trzech białek. Głównym parametrem i zarazem niewiadomą tej techniki jest sposób unormowania ν_{ij}, który możemy zapisać w postaci proporcji

$$E_{ij}^\nu = \epsilon_{ij} \cdot \nu_{ij} / \Omega_k, \tag{3.20}$$

gdzie Ω_k oznacza różne sposoby normalizacji. W niniejszej pracy rozważamy trzy sposoby unormowania ($k = 1, 2, 3$). Ω_1 oznacza liczbę wszystkich par pomiędzy ciężkimi atomami w rozważanym białku i w konsekwencji całkowitą energię w stanie natywnym równą liczbie kontaktów. Ω_2=4.9 reprezentuje najbardziej typową liczbę oddziaływań między ciężkimi atomami dla pary reziduów, otrzymaną na podstawie zbioru



Rysunek 3.2: Rozkład średniej liczby kontaktów unormowany do liczby oddziałujących par i, j w białku, na podstawie 3500 białek. Panele po prawej stronie od góry do dołu, rozkład liczby oddziaływań ν_{ij}, między parami aminokwasów i, j w białkach o 1tit, 1ubq i 1b6i.

3500 białek, zgodnie z rozkładem pokazanym na rysunku 3.2. Ω_3 oznacza największą liczbę kontaktów między ciężkimi atomami dla danej pary aminokwasów w rozważanym białku. Panele wewnętrzne na rysunku 3.2 pokazują między innymi przykładowe wartości Ω_3 uzyskane dla trzech klas białek 1tit(Ω_3=26) , 1ubq (Ω_3=19), 1b6i(Ω_3=34). Liczby te są największe dla białek helikalnych jak 1b6i. Trzeci sposób normalizacji był wcześniej użyty w pracy [151], jednakże maksymalna wartość Ω_3 tam uzyskana jest inna od otrzymanej w tej pracy. Różnica wynika z użycia różnych technik do wyznaczenia mapy natywnych kontaktów, co jak pokazaliśmy już wcześniej zasadniczo wpływa na liczbę natywnych par.

Skala energii E_{ij}^{HH}

Opisane powyżej różne warianty skal energii powstały z rozważenia geometrii stanu podstawowego. Alternatywnym podejściem jest uwzględnienie własności chemicznych i wzięcie pod uwagę skali hydrofobowości i obecności wiązań wodorowych.

Skala energii uwzględniająca własności hydrofobowe E^{HH} powstała na podstawie pracy [152], gdzie oddziaływaniu dwóch hydrofobowych aminokwasów i, j przypisano energię $E_{ij}^{HH} = 2\epsilon$, parze w skład której wchodzi hydrofobowy i alifatyczny aminokwas 1ϵ, i jeśli nie występuje żaden hydrofobowy $E_{ij}^{HH} = 0$. Do aminokwasów hydrofobowych zostały zaliczone Cys, Ile, Leu, Met, Phe, Trp, i Val, a do alifatycznych Ala, His, Thr i Tyr. Skalę energii E^{HH} stosujemy równocześnie ze skalą faworyzującą istnienie wiązań wodorowych E^{HB}.

Skala energii E^{HB}

Identyfikacja wiązań wodorowych została przeprowadzona metodą zaproponowaną przez Kabscha i Sandera [154]. Technika ta oszacowuje energię wiązania wodorowego na podstawie geometrii i elektrostatyki, które obie są funkcją długości wiązania wodorowego i odpowiedniego ułożenia względem siebie akceptora (atom N) i donora (atom C). Wiązanie wodorowe jest obecne jeśli energia wiązania leży poniżej pewnej progowej wartości. Zauważmy, iż wykonanie tego oszacowania wymaga obecności atomów wodoru, które jednakże są zawarte tylko w co drugim pliku PDB. Z tego powodu zmodyfikowaliśmy tę technikę tak, aby polegała ona sprawdzaniu odległości pomiędzy powiększonymi [147] sferami van der Waalsa dla atomów N i C, i mogła być zastosowana do każdego białka.

Dla tej skali energii przyjmujemy, iż kontakty zaliczone do grona wodorowych oddziałują poprzez $E_{ij}^{HB}=\epsilon$, dla $i, j-i > 2$ wzdłuż sekwencji. Ponadto niektóre aminokwasy w konformacji kartek β lub β spinek mogą oddziaływać poprzez dwa wiązania wodorowe, które to wymuszają dodatkową orientację łańcucha głównego. W przypadku kiedy aminokwasy i i j oddziałują poprzez dwa wiązania wodorowe lub wiązanie wodorowe i kontakt z grupą boczną, dodatkowo pary $(i, j-1), (i, j+1), (i-1, j), (i+1, j)$ zaliczamy do kontaktów natywnych o wartości energii oddziaływania $\frac{1}{4}\epsilon$. Podobny sposób wzmocnienia konformacji wokół aminokwasów oddziałujących dwoma wiązaniami wodorowymi był zastosowany w pracy [153, 25].

Skala energii E^{MJ}

Alternatywnym sposobem uwzględnienia różnych własności chemicznych 20 aminokwasów jest użycie efektywnej energii zaproponowanej przez Miyazawę i Jernigana [136], omówionej w rozdziale 2.2. W wyniku tego podejścia została skonstruowana symetryczna macierz o rozmiarach 20×20, określająca energię oddziaływań dla par aminokwasów i, j o 210 różnych parametrach ϵ_{ij}^{MJ}. W niniejszej pracy wartości te wymagają unormowania proporcjonalnie do energii oddziaływania dla pozostałych kontaktów natywnych w tym oddziaływ wodorowych (E_{ij}^{HB}). Korzystając z wyników pracy [25] parametry energii ϵ_{ij}^{MJ} normujemy przez wartość średnią wszystkich 210 parametrów. Parametry ϵ_{ij}^{MJ} stosujemy do kontaktów natywnych tworzonych wyłącznie poprzez grupy boczne, pozostałym kontaktom wodorowym przypisujemy skalę energii E_{ij}^{HB}, zgodnie z procedurą zaproponowaną w pracy [25]. Reasumując, skala energii E_{ij}^{MJ} jest złożeniem parametrów energii ϵ_{ij}^{MJ} i skali E_{ij}^{HB}. Alternatywny sposób konstrukcji efektywnego potencjału został zaproponowany w pracy [137].

Porównanie skal energii

Tabela 3.2 zawiera energie w stanie natywnym otrzymane na podstawie powyżej dyskutowanych skal energii dla mapy $M3$, na podstawie białek 1tit, 1ubq i 1crn. W celu porównania różnych skal energii dla danego białka, wartości w tabeli zostały podane jako bezwymiarowe, po unormowaniu ich poprzez ϵ=const z jednorodnej skali energii. Wyniki te pokazują, iż wypadkowe wartości energii są bardzo zbliżone do energii uzyskanej dla jednorodnej skali energii ϵ (różnica wynosi mniej niż 8%).

Znacząco powyżej jest skala $E^{MJ,HB}$ a poniżej E^{T3}. Jest to bardzo ważna obserwacja, gdyż pozwala ona na przyjęcie takiej samej wartości temperatury zredukowanej $\widetilde{T}$ =0.3. Wartości zredukowanej temperatury dla modeli skal $E^{MJ,HB}$ i $E^{L1,T3}$ będą podane w rozdziale 4. Ważną obserwacją jest też zmniejszenie wartości energii o ponad 15% przy zmianie mapy kontaktów z $M3$ na $M4$.

Tablica 3.2: Energia wyrażona w jednostkach bezwymiarowych w stanie natywnym dla różnych skal energii na podstawie mapy $M3$. Pierwsza kolumna dodatkowo podaje wartości uzyskane na podstawie $M4$. Każdy wiersz reprezentuje wartości uzyskane dla danego białka, 1tit,1crn i 1ubq.

białko	E^o/ϵ	E^{L1}/ϵ	E^{L2}/ϵ	E^{T1}/ϵ	E^{T2}/ϵ	E^{T3}/ϵ	$E^{\nu 1}/\epsilon$	$E^{\nu 2}/\epsilon$	$E^{HB,MJ}/\epsilon$
	$M3$ $M4$								
1tit	167 / 156	166.55	162.26	157.25	163.1	151.65	167	153.26	236.21
1crn	100 / 79	100	99.37	96.75	98.70	92.95	100	97.75	124.67
1ubq	171 / 150	170.23	164.02	159.25	166.3	154.95	171	167.95	243.30

3.2 Model Go z grupami bocznymi, $C^{\alpha-\beta}$

Rozważymy teraz model Go, w którym grupy boczne w każdym aminokwasie (oprócz glicyny) reprezentujemy dodatkowym węglem C^{β}. Jest to motywowane faktem, iż topologiczny i chemiczny charakter oddziaływań grup bocznych jest inny niż oddziaływań między łańcuchem głównym. Ponadto w modelach rozważanych w niniejszej pracy nie jest uwzględniona specyficzność oddziaływań woda-białko. Oddziaływania te ułatwiają zwijanie, kierują aminokwasy hydrofobowe do wnętrza białka a hydrofilowe na zewnątrz białka. Model, w którym aminokwas jest reprezentowany dwoma atomami C^{α} i C^{β}, będziemy oznaczać symbolem $C^{\alpha-\beta}$.

Do konstrukcji modelu $C^{\alpha-\beta}$ używamy zarówno niektórych potencjałów i map kontaktów omówionych na przykładzie modelu z C^{α}, jak też wprowadzamy nowe formy wynikające z geometrii modelu z grupami bocznymi. Szczegółowy opis dyskutowanego tutaj modelu przedstawiono także w pracach [IV] oraz [VIII].

3.2.1 Konstrukcja modelu

Grupy boczne w białku reprezentujemy poprzez uwzględnienie węgli C^{β}. Atom C^{β} jest pierwszym atomem z grupy bocznej i jest obecny w każdym aminokwasie poza glicyną. W stanie natywnym, położenie C^{β} jest określone w pliku PDB rozważango białka. Preferowane położenie dynamiczne grupy bocznej C^{β} określamy w następujący sposób. Każdy atom węgla C^{β} łączymy z C^{α} tego samego aminokwasu poprzez wiązanie harmoniczne, którego minimum odpowiada odległości ze stanu natywnego, a kierunek jest wyznaczony przez wektor $\vec{r}_i^{\beta}$ na podstawie położenia $\vec{r}^{\alpha}$ danego przez C_i^{α} i jego sąsiadów wzdłuż łańcucha. Kierunek wektora $\vec{r}_i^{\beta}$ wynika z rozważań nad geometrią wiązania peptydowego [155, 156] i dany jest wzorem

$$\vec{r}_i^{C^{\beta}} \ = \ l(\hat{a}cos\ \theta + \hat{b}sin\ \theta), \tag{3.21}$$

gdzie optymalna wartość kąta θ wynosi 37.6° i

$$\hat{a} = \frac{\hat{s}_{\ i,i-1} + \hat{s}_{\ i,i+1}}{|\hat{s}_{\ i,i-1} + \hat{s}_{\ i,i+1}|} \qquad \hat{b} = \frac{\hat{s}_{\ i,i-1} \times \hat{s}_{\ i,i+1}}{|\hat{s}_{\ i,i-1} \times \hat{s}_{\ i,i+1}|}. \tag{3.22}$$

Wektor jednostkowy $\hat{s}_{i,j}$ dany jest wzorem

$$\hat{s}_{ij} = \frac{\hat{r}_i^{C^{\alpha}} - \hat{r}_j^{C^{\alpha}}}{|\hat{r}_i^{C^{\alpha}} - \hat{r}_j^{C^{\alpha}}|} \quad . \tag{3.23}$$

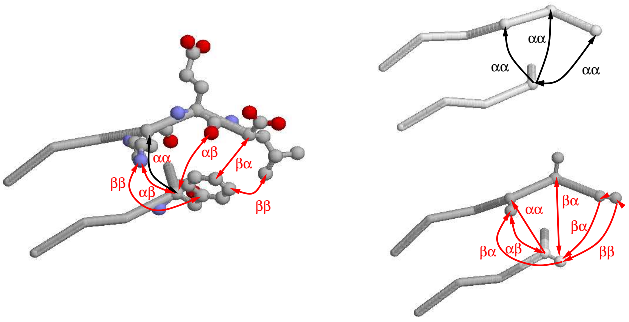
Energia potencjalna molekuły skonstruowanej z N węgli C^{α} i C^{β} w położeniach $\{\mathbf{r}_i\}$, $i = 1, 2 \ldots N$ dana jest wzorem:

$$E_p^{\alpha-\beta}(\{\mathbf{r_i}\}) \ = \ V^{BB} \ + \ V^{BS} + \ V^{NAT} \ + \ V^{NON} \ + \ V^{S} \ \ . \tag{3.24}$$

Człony V^{BB}, V^{S}, V^{NON} są takie same jak dla C^{α} modelu Go. V^{BS} reprezentuje potencjał harmoniczny utrzymujący węgiel C^{β} w odległości natywnej $|\vec{r}_i^{\beta}|l$=1.5 Å od C^{α}. Potencjały V^{NAT} opisuje oddziaływania dla czterech różnych par natywnych kontaktów. Konstrukcja modelu $C^{\alpha-\beta}$ wraz z nową mapą kontaktów (omówioną poniżej) jest pokazana na rysunku 3.3.

3.2.2 Mapa kontaktów

Mapa natywnych kontaktów dla modelu $C^{\alpha} - C^{\beta}$ jest skonstruowana według techniki drugiej z podrozdziału 3.1.3, gdzie obecność kontaktu jest sprawdzana na podstawie przykrywania się powiększonych sfer van der Waalsa [147] między ciężkimi atomami. Uwzględnienie dodatkowego atomu swobody umożliwia znacznie dokładniejszy opis oddziaływań między parą aminokwasów. Obecnie rozróżniamy cztery rodzaje



Rysunek 3.3: Rysnek lewy pokazuje częściowo pełnoatomowy obraz molekuły wraz z zaznaczonymi kontaktami między ciężkimi. Rysunek prawy pokazuje liczbę i położenie możliwych kontatków natywnych między parą aminokwasów w modelu typu C^{α} (górny rysunek) i modelu typu $C^{\alpha} - C^{\beta}$ (dolny rysunek).

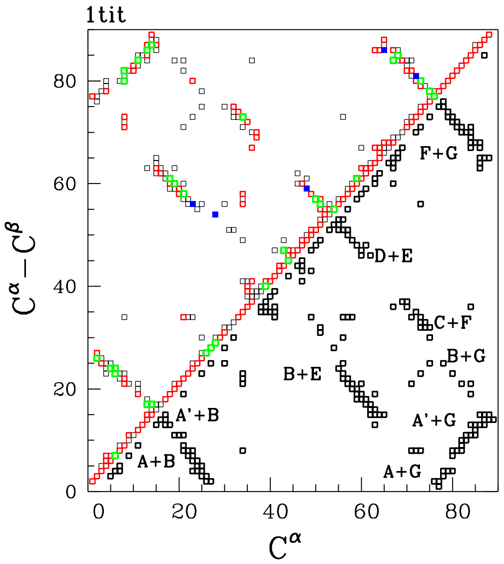
przykrywań atomowych i w konsekwencji cztery rodzaje kontaktów: a) C_i^{β}-C_j^{β}, jeśli istnieje chociaż jedno przykrywanie się efektywnych atomów z grup bocznych i, j; b) C_i^{β}-C_j^{α}, jeśli istnieje chociaż jedno przykrywanie między grupą boczną i a łańcuchem peptydownym j; c) C_i^{α}-C_j^{β} jeśli istnieje przykrywanie między łańcuchem głównym i a grupą boczną j; d) C_i^{α}-C_j^{α} jeśli istnieje przykrywanie między łańcuchami głównymi reziduuów i, j. Różne rodzaje kontaktów są pokazane na rysunku 3.3. Warto także podkreślić, że para aminokwasów w modelu $C^{\alpha-\beta}$ może być w kontakcie poprzez od jednego aż do czterech przyciągających potencjałów. Rysunek 3.4 pokazuje różnicę w liczbie kontaktów między parą aminokwasów w modelu zbudowanym z C^{α} i $C^{\alpha-\beta}$. Analizując zbiór 500 białek stwierdziliśmy, że średnia liczba kontatów dla pary aminokwasów wynosi około 1.5.

3.2.3 Potencjały dla oddziaływań natywnych

Do opisu oddziaływań między natywnymi kontaktami stosujemy oryginalny potencjał Lennarda-Jonesa V^{6-12} dany wzorem (3.5) oraz potencjał V^{10-12} dany wzorem (3.9). Na przykład dla modelu $C^{\alpha-\beta}$ potencjał V^{6-12} przyjmuje formę

$$V_{\alpha-\beta}^{6-12} = \sum_{i+1<j}^{NAT\ \alpha} 4\ E_{ij} \Big[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6} \Big] + \sum_{i+1<j}^{NAT\ \beta} 4\ E_{ij} \Big[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6} \Big], \quad (3.25)$$

gdzie pierwsze sumowanie przebiega po węglach C^{α} a drugie po węglach C^{β} i odpowiada czterem rodzajom kontaktów, C_i^{β}-C_j^{β}, C_i^{β}-C_j^{α}, C_i^{α}-C_j^{β} i C_i^{α}-C_j^{α}. Każde z oddziaływań jest reprezentowane przez odpowiadający mu parametr odległości σ_{ij}, wyznaczony tak, aby minimum potencjału zgadzało się z odległością natywną, np. dla kontaktu $C^{\beta} - C^{\beta}$



Rysunek 3.4: Mapa kontaktów dla tytyny, powyżej diagonali w modelu dwukulkowym, poniżej w modelu jednokulkowym. Kolor kwadratów powyżej diagonali odpowiada liczbie oddziaływań między parą aminokwasów: niebieskie pełne kwadraty oznaczają 4 oddziaływania, zielone 3, czerwone 2 i czarne puste kwadraty 1 oddziaływanie. Elementy struktury drugorzędowej będące w kontakcie oznaczone są literami zgodnie z notacją na rysunku 4.1.

jest to odległość między węglami $(C^\beta - C^\beta)$ pomnożona przez czynnik $\sqrt[6]{2}$. Kryterium istnienia kontaktu jest takie samo jak dla modelu jednokulkowego. W podobny sposób można zapisać potencjał V^{10-12} dla modelu dwukulkowego, gdzie każdy z czterech możliwych kontaktów jest reprezentowany przez odpowiadający mu parametr odległości r_{ij}^n wyznaczony w stanie natywnym, np. dla dla kontaktu $C^\beta - C^\beta$ jest to odległość między węglami $C^\beta - C^\beta$.

3.2.4 Skale energii

Jednorodna skala energii

W modelu $C^{\alpha-\beta}$ w pierwszym przybliżeniu zakładamy, że każdy z istniejących natywnych kontaktów jest równie ważny w białku. Bazując na tym założeniu przyjmujemy, iż cztery rodzaje możliwych oddziaływań pomiędzy parą dwóch aminokwasów mają taką samą siłę $E_{ij}{=}E^o{=}\epsilon$. Tak więc, w idealnym układzie energia w stanie natywnym byłaby dokładnie 4 razy większa niż dla modelu zbudowanego tylko z C^α.

Przy przełożeniu wartości energii używanej w modelu Go jednokulkowym na rzeczywistą wartość, założyliśmy, iż średnia energia oddziaływania między parą aminokwasów wynosi około 2 kcal/mol, tak więc ϵ/k_B odpowiadał 1000K. Z tego założenia korzystamy dalej, jednakże obecnie każdy aminokwas średnio oddziałuje z wartością 1.5 ϵ, jak wynika z rozkładu średniej ilości oddziaływań między kontaktami. Zatem dla temperatury pokojowej zredukowana temperatura w modelu dwukulkowym $\widetilde{T} = k_B T/\epsilon$ jest rzędu $300K/750K$ i wynosi 0.4.

Niejednorodna skala energii E_{ij}^{ν}

Konstrukcja tej skali została opisana powyżej w punkcie 3.1.4, jednakże przypomnijmy, że polega ona na uzależnieniu wartości energii oddziaływania od liczby atomowych przykrywań ν_{ij} między rozważaną parą aminokwasów i, j. W niniejszym modelu ν_{ij} jest podzielone na cztery przypadki odpowiadające czterem możliwym oddziaływaniom i dane jest wzorem

$$E_{ij}^{\nu} = \begin{cases} \epsilon_{ij} \cdot \nu_{ij}^{\alpha\alpha}/\Omega_{\alpha\alpha} & \text{dla } C^{\alpha-\alpha} \\ \epsilon_{ij} \cdot \nu_{ij}^{\alpha\beta}/\Omega_{\alpha\beta} & \text{dla } C^{\alpha-\beta} \\ \epsilon_{ij} \cdot \nu_{ij}^{\beta\alpha}/\Omega_{\beta\alpha} & \text{dla } C^{\beta-\alpha} \\ \epsilon_{ij} \cdot \nu_{ij}^{\beta\beta}/\Omega_{\beta\beta} & \text{dla } C^{\beta-\beta} \end{cases} \tag{3.26}$$

gdzie np. $v_{ij}^{\beta\beta}$ oznacza liczbę kontaktów atomowych wyłącznie między grupami bocznymi pary aminokwasów i, j, i np. Ω_{bb} oznacza normalizację do średniej liczby przykrywań atomowych między kontaktami typu $\beta\beta$ w danym białku.

Rozdział 4

Optymalny model Go do rozciągania białek

Obecnie rozważymy sposób wyboru optymalnego wariantu modelu Go do rozciągania białek na podstawie zgodności wyników teoretycznych z wartościami mechanicznego oporu białek wyznaczonymi eksperymentalnie dla 29 białek. Rozważamy 61 najbardziej realistycznych wariantów modelu Go z 504 zaprezentowanych w poprzednim rozdziale. Podkreślmy w tym miejscu istotny fakt, iż minimum globalne w modelu Go odpowiada stanowi natywnemu molekuły, a zatem poprawnie skonstruowany model powinnien najlepiej opisywać ruch molekuły przede wszystkim w pobliżu stanu natywnego, a nie z dala od niego. Sugeruje to, iż rozważenie wyników eksperymentalnych dla rozciągania białek powinno dostarczyć dużo cenniejszych wskazówke do konstrukcji optymalnego modelu Go, niż korzystanie z wyników dotyczących zwijania białek. Ponadto zauważmy, że generowana podczas mechanicznego rozciągania białka i przeciwstawiająca się jego rozciąganiu siła oporu, mimo iż w skomplikowany sposób zależy od czasu rozciągania, ma dobrze określone i mierzalne doświadczalnie maksimum F_{max}. Tak więc, wybór optymalnej formy modelu powinien być przeprowadzony na podstawie zgodności eksperymentalnie wyznaczonej wartości F_{max} z analogiczną wartością wyznaczoną z symulacji rozciągania białek.

Nasze rozważania rozpoczniemy od określenia testowego zbióru białek badanych eksperymentalnie i otrzymanych dla nich danych doświadczalnych, które będą stanowić wzorzec dla teoretycznie wyznaczonych wartości. Następnie opisujemy algorytm dynamiki molekularnej do rozciągania białek. Określamy parametry i technikę rozciągania białek. Podajemy statystyczny sposób pomiaru korelacji. Następnie określamy temperatury wykonania symulacji rozciągania. Przeprowadzamy symulacje rozciągania zbioru testowego w 61 modelach, z których wyznaczamy F_{max}, po czym przechodzimy do omówienia wyników. Wyboru optymalnego modelu dokonujemy na podstawie dwóch kryteriów. Pierwsze z nich polega na sprawdzeniu zgodności pomiędzy mechanicznym oporem wyznaczonym teoretycznie F^t_{max}, a jego odpowiednikiem F^e_{max} otrzymanym doświadczalnie; zgodność tę obliczamy jako współczynnik korelacji Pearsona. Drugie kryterium oparte jest o współczynnik Theil'a U. Model, dla którego otrzymamy najwyższy współczynnik Pearsona oraz najniższy współczynnik Theila, przyjmiemy jako optymalną formę modelu Go do rozciągania białek. Następnie dyskutujemy wpływ (a) różnych form potencjałów dla oddziaływań natywnych, (b) potencjału określającego sztywnośc łańcucha, (c) wyboru skal energii, na charakter i sposób rozwijania się białek, poprzez zgodność krzywych $F(d)$ i scenariusza pękania kontaktów z dostępnymi wynikami doświadczalnymi.

Szczegółowy opis wyników opisanych w niniejszym rozdziale można także znaleźć w pracy [VIII].

4.1 Metody

4.1.1 Dane doświadczalne

Tablica 1.1 w rodziale 1 zawiera zbiór wartości F^e_{max} dla wszystkich 55 białek, które dotychczas zbadano doświadczalnie. Jako że niezbędnym elementem konstrukcji modelu Go jest znajomość struktury natywnej białka zdeponowanej w bazie PDB, ze zbioru 55 białek wybraliśmy podzbiór 29 białek, których kompletna struktura jest dostępna w PDB. Podzbiór ten nazywamy zbiorem testowym i oznaczamy symbolem S29. W szczególności spośród 55 wyjściowych białek odrzuciliśmy krowią anhydrazę 1v9e (gdyż doświadczalnie była ona rozciągana z koordynacyjnie połączonym jonem cynku), oraz białka E2lip3 i rybonukleazę H (które są reprezentowane strukturami o bardzo różnej od siebie wartości średniego kwadratowego odchylenia $RMSD$ [1]). Tablica 4.1 zawiera zbiór wyselekcjonwanych 29 eksperymentalnie zmierzonych wartości F^e_{max} spośród 55 białek z tablicy 1.1 z rozdziału 1.

Symulację rozciągania przeprowadziliśmy dla każdego białka ze zbioru S29, ciągnąc za te aminokwasy, których użyto doświadczalnie. W większości przypadków kierunek ciągnięcia jest wyznaczony przez wekotr łączący terminalne atomy C^α. W przypadku innego kierunku, tak jak w przypadku zielonego białka GFP o kodzie PDB 1emb, rozciąganego doświadczalnie w trzech kierunkach, miejsce zaczepienia siły, które wyznacza kierunek rozciągnięcia, jest podane w nawiasach. Należy jednak zaznaczyć, iż

[1] Root Mean Square Deviation, jest miarą używaną do porównywania struktur, która wyraża się wzorem $\sqrt{\sum_{i=1}^{N}\frac{d_i^2}{N}}$, gdzie d_i jest odległością pomiędzy porównywanymi atomami o indeksie i, a N jest liczbą atomów w każdej z struktur.

dla 12 białek ze zbioru S29 doświadczalny pomiar siły był wynikiem rozciągania kilku modułów połączonych w tandem, przy użyciu sztucznych linkerów. Ponadto moduły te nie zawsze były takie same i mogły się rozwijać w różnej kolejności. Problemy powstające z interpretacją takich danych scharakteryzowaliśmy w części 1.2.1. Z grona 12 tandemowo rozciąganych białek dla 10 można założyć (korzystając z wyników z innych prac teoretycznych i doświadczalnych), iż rozciągnięcie pojedynczej domeny jest wystarczające dla porównania danych. W przypadku tandemowo rozciąganych FNIII dodatkowo wykonaliśmy symulacje wielodomenowe.

Niektóre z białek są ponadto reprezentowane kilkoma strukurami, które różnią się znacząco wartością RMSD, a w konsekwencji mają różne mapy kontaktów. Do tych białek należą białko L (2ptl, 1hz6, 1hz5), białko G (1pga, 1gbi), druga domena białka wiążącego wapń C2A (1dqv,1rsy,1byn), barnaza (1bni,1bnr,1byn), oraz 10 domen fibronektyny FNIII-10 (1fnf, 1ttf, 1ttg). Teoretycznie wyznaczona siła dla tych białek jest średnią z wartości siły dla każdej ze strukur.

Badania doświadczalne i teoretyczne [71, 91, 157] pokazały istnienie kilku ścieżek rozwijania, które zazwyczaj charakteryzują się inną wartością mechanicznego oporu. W zbiorze testowym do takich białek należą: zielone białko fluoryzujące GFP, białko L, białko G i białka z rodziny fibronektyn. Wartość mechanicznego oporu podana w tablicy 4.1 reprezentuje w tych przypadkach najbardziej typową trajektorię.

4.1.2 Protokół symulacji rozciągania białek

Równania ruchu

Ruch molekuły białka w stałej temperaturze jest badany metodą dynamiki molekularnej z termostatem reprezentowanym szumem Langevina, opisanym w rozdziale 2.1. W równaniu Langevina danym wzorem 2.2 wskaźniki i oznaczają obecnie atomy C_i^α reprezentujące aminokwasy, a m jest równe średniej masie aminokwasu w danym białku. $\vec{F}_{c,i}$ oznacza wypadkową siłę wynikającą z molekularnych potencjałów oraz zewnętrzne siły, które działają na kulkę o numerze i w położeniu r_i. Stała tłumienia γ wynosi $2m/\tau$, a temperatura T jest wprowadzona poprzez siły przypadkowe o dyspersji $\sqrt{2\gamma k_B T}$. W tym opisie dynamika każdej kulki jest przetłumiona, tak więc naturalna skala czasu ruchu kulki nie odpowiada okresowi oscylacji w minimum potencjału, lecz jest równa czasowi jaki jest potrzebny do przebycia odległości równej rozmiarowi aminokwasu (1.5-4.1Å) podczas swobodnej dyfuzji [138, 159, 22]; czas ten jest rzędu nanosekund. Równanie ruchu jest rozwiązywne numerycznie przy pomocy algorytmu przewidywania-poprawiania piątego rzędu [127], gdzie tarcie i przypadkowe fluktuacje pojawiają się jako szum zaburzający układ w każdym kroku całkowania.

Tablica 4.1: Wartość opór mechanicznego uzyskana w optymalnym modelu Go $\{6-12, C, M3, E^o\}$ i w doświadczeniu.

PDB	N	F $^{e}_{max}[\epsilon/\text{Å}]$	F $^{t}_{max}[\epsilon/\text{Å}]$		odnośniki
1tit	89	204 +/-30	2.15	I27*8	[4, 158]
1nct	98	210 +/-10	2.4 +/-0.2	I54-I59	[72, 76]
1g1c	97	127 +/-10	2.3 +/-0.2	M5 tytyny	[63]
1b6i	164	64 +/- 30	1.2	T4 lysozym(21-141)	[93]
1aj3	106	68 +/- 20	1.23	spektryny R16	[78]
1qjo	80	15 +/- 10	1.2	eE2lip3(N-C)	[84]
1qjo	40	177 +/- 10	2.0	E2lip3(N-41)	[84]
1dqv	127	60 +/- 15	1.5	C2A	[92]
1rsy	127	60 +/- 15	1.7 +/-0.2	C2A	[92]
1byn	127	60 +/- 15	1.4	C2A	[92]
1cfc	148	80 +/- 13	0.75	kalmodulina	[92]
1n11	33	37 +/- 9	0.4	ankiryna*1	[85, 86]
1bni	108	70 +/- 15	1.4, 1.7	barnaza$_{I27}$	[94]
1bnr	108	70 +/- 15	1.05	barnaza$_{I27}$	[94]
1bny	108	70 +/- 15	1.1, 1.3	barnaza$_{I27}$	[94]
1hz6	67	152 +/- 10	3.5	białko L	[99]
2ptl	67	152 +/- 10	2.2 +/-0.2	białko L	[99]
1ksr	100	45 +/- 20	2.0 +/-0.3	^{4}DdFLN	[89, 90]
2rn2	155	19 +/- 10	1.8 +/-0.2	ribonukleaza H	[67]
1ubq	76	230 +/- 34	2.32	ubikwityna	[96]
1ubq	76	203 +/- 35	2.32	ubikwityna(N-C)*9	[95, 96]
1ubq	28	85 +/- 20	0.9	ubikwityna(K48-C)*(2-7)	[95, 96]
1emb	129	350 +/- 30	5.15 +/-0.4	GFP(3-132)	[97]
1emb	219	130 +/- 30	2.3, 4.3	GFP(3-212)	[97]
1emb	80	120 +/- 30	2.2 +/-0.2	GFP(132-212)	[98]
1emb	235	104 +/- 40	2.3 +/-0.2	GFP(N-C)	[98]
1fnf	94	75 +/- 20	1.6, 1.8	^{10}FNIII	[70, 73]
1ttf	94	75 +/- 20	0.7, 1.2	^{10}FNIII	[74]
1ttg	94	75 +/- 20	0.7, 1	^{10}FNIII	[74]
1fnh	92	124 +/- 18	1.8	^{12}FNIII	[73]
1fnh	89	89 +/- 18	1.4, 1.7	^{13}FNIII	[73]
1oww	93	125 +/- 31	2.1 +/-0.2	^{1}FNIII	[73]
1ten	90	135 +/- 40	1.7	^{3}TNFNIII	[61, 73]
1pga	56	190 +/- 25	1.9	białko G	[157]
1gb1	56	190 +/- 25	1.65	białko G	[157]

Jako krok całkowania przyjmujemy Δt=0.005τ, gdzie $\tau = \sqrt{ma^2/\epsilon}$ a wielkość a=5Å jest dobrana tak, aby odpowiadała typowej wielkości sfery van der Waalsa aminokwasów.

Symulacje rozciągania

W symulacji rozciągania oba końce białka są przymocowane do sprężyn o stałej sprężystości k=0.06 ϵ/Å^2, która jest bliska właściwości elastycznej dźwigni używanej w doświadczeniu i dla ϵ=0.2eV wynosi około 48N/m. Swobodny koniec sprężyny zaczepionej za terminal N jest zakotwiczony, zaś swobodny koniec drugiej sprężyny jest przesuwany ze stałą prędkością $v_p = 0.005$ Å/τ wzdłuż pewnego wyznaczonego kierunku. Jeśli nie podajemy inaczej, kierunek ten jest równoległy do wektora łączącego końce białka w stanie natywnym. Zgodnie z założeniem, iż ruch w układzie jest przetłumiony i skala czasu τ jest rzędu 3 ns, prędkość v_p wynosi 10^5nm/s. Jest to zaledwie o rząd wielkości więcej niż maksymalna eksperymentalna prędkość. Przesuwanie sprężyn odpowiada przesuwaniu dźwigienki AFM, a odległość w stanie natywnym między terminalnymi aminokwasami odpowiada zerowemu przesunięciu oznaczanemu wielkością d. Przesunięcie sprężyny powoduje opór białka F w wyniku czego otrzymujemy wykres $F(d)$. W skończonej temperaturze pojawiają się fluktuacje siły, wynikające ze wzbudzenia ciągnięcia sprężyny i białka. Szum z wykresu siły eliminujemy poprzez uśrednienie $F(d)$ co 100τ, co odpowiada średniowaniu odległości d co 0.5Å.

4.1.3 Interpertacja danych

Statystyczny pomiar współczynnika korelacji teorii z danymi doświadczalnymi

Załóżmy, iż maksymalana siła F_{max} została wyznaczona dla zbioru D białek. Eksperymentalnie wyznaczoną wartość F_{max} oznaczamy jako $F^e_{max,\lambda}$, gdzie $\lambda = 1, ..., D$. Odpowiadającą jej teoretyczną wartość z symulacji oznaczamy jako $F^t_{max,\lambda}$. Stopień korelacji między zbiorami $F^t_{max,\lambda}$ i $F^e_{max,\lambda}$ może być zmierzony poprzez kwadrat współczynnika korelacji Pearsona

$$R^2 = 1 - \sqrt{\frac{1}{D}\sum_{\lambda=1}^{D}\big(\frac{F^t_\lambda - F^e_\lambda}{F^e_\lambda}\big)^2}\ . \tag{4.1}$$

Czym bliżej wartość R^2 znajduje się jedynki, tym bardziej dwa zbiory danych są podobne, natomiast wartość R^2 równa zero oznacza brak jakiejkolwiek korelacji. Optymalne dopasowanie danych do zależności

$$F^t_{max,\lambda} = a \cdot F^e_{max,\lambda} \tag{4.2}$$

jest otrzymane metodą najmniejszych kwadratów. Wadą wyznaczenia R^2 jest to, iż może ono zostać przeszacowane przez pojedyncze duże błędy, dając w wyniku słabą korelację i niesłuszne odrzucenie dobrego modelu. Dlatego rozważamy także inną statystyczną miarę korelacji nazywaną współczynnikiem U wprowadzonym przez Theil'a (z ang. Theil's U coefficient). Współczynnik U uwzględnia relatywne zmiany pomiędzy danym białkim a jego sąsiadem, tak więc wymaga ustalenia kolejności, względem której korelacja jest liczona. W moim przypadku zbiór danych został uprządkowany od najmniejszej do największej wartości $F^e_{max,\lambda}$. Niech $W_{\lambda+}$ i $w_{\lambda+1}$ oznaczają odpowiednio przewidywaną i aktualną pojedynczą zmianę w trendzie

$$W_{\lambda+1} = \frac{F^t_{\lambda+1} - F^e_\lambda}{F^e_\lambda}, \qquad w_{\lambda+1} = \frac{F^e_{\lambda+1} - F^e_\lambda}{F^e_\lambda}. \tag{4.3}$$

Współczynnik U dany jest relacją

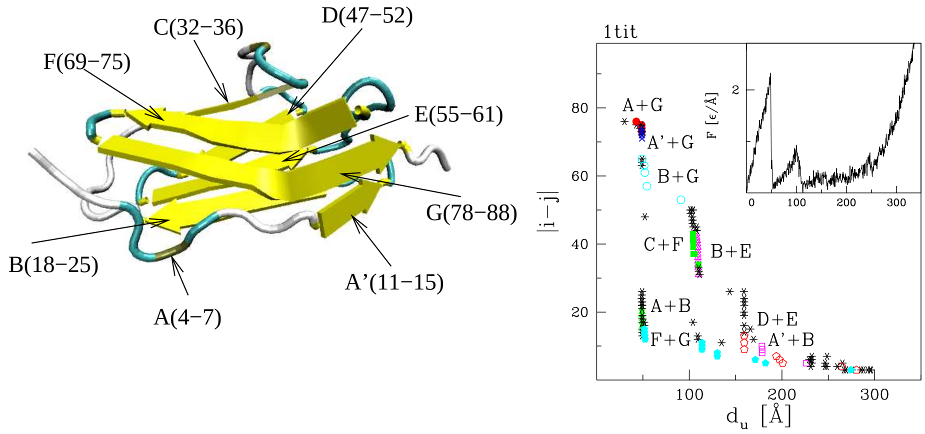
$$U = \sqrt{\frac{\sum_{\lambda=1}^{D-1}(W_{\lambda+1} - w_{\lambda+1})^2}{\sum_{\lambda=1}^{D-1}(w_{\lambda+1})^2}} \ . \tag{4.4}$$

W przypadku idealnej korelacji U=0. Wielkość U większa niż 1 odpowiada sytuacji, w której teoretyczne przewidywanie jest przeciwne do wartości zmierzonej, natomiast U=1 ($W_{\lambda+1}$=0) oznacza brak zmiany w przewidywaniu.

Interpretacja danych

Własności mechaniczne zostały najlepiej poznane doświadczalnie i teoretycznie (z pełnoatomowych symulacji) dla 27. domeny tytyny, która w niniejszej pracy będzie używana jako białko testowe do badania różnych procesów. 1tit jest białkiem klasy β, składającym się z ośmiu nici β oznaczonych A(4-7), A'(11-15), B(18-25), C(32-36), D(47-52), E(56-61), F(69-75) i G(78-88). Struktura trzeciorzędowa tytyny wraz z zaznaczonymi nićmi β jest pokazana na rysunku 4.1.

Jako przykład wyniku symulacji rozciągania białka tytyny modelem Go o parametrach $\{6-12, C, M3, E^o\}$, na rysunku 4.1 przedstawiamy wykres siły oporu w funkcji przemieszczania, gdzie dla rozciągnięcia d = 85Å indentyfikujemy maksimum siły F^t_{max}=2.1ϵ/Å. Podkreślmy, że do opisu wyników teoretycznych stosujemy parametr d, określający rozciągnięcie białka w stosunku do odległości między końcami białka w stanie natywnym (tzn. w stanie natywnym przyjmujemy $d = 0$) — w przeciwieństwie do parametru L mierzonego doświadczalnie, pokazanego np. na rysunku 1.



Rysunek 4.1: Lewy panel: wstążkowa reprezentacja 27 domeny tytyny (1tit). Symbole oznaczają osiem kolejnych nici β, wraz z pozycją sekwencyjną aminokwasów w każej z nich. Prawy panel: przykładowy scenariusz rozwijania wraz z wykresm siły w funkcji rozciągnięcia w modelu Go o parametrach $6-12, C, M3, E^o$ dla 1tit, dla $\tilde{T}$=0.3.

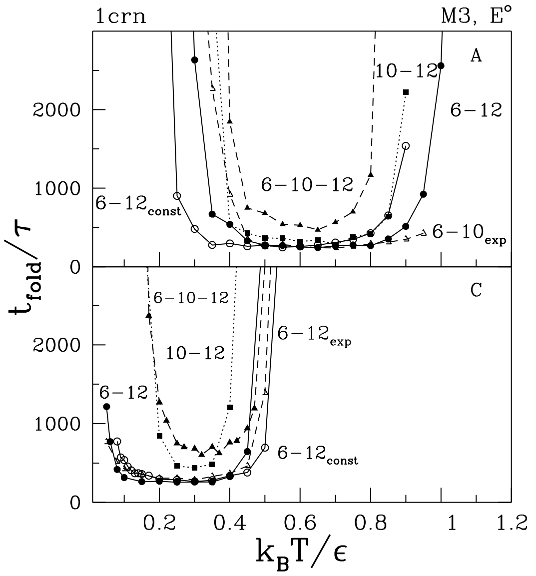
Szczegółową analizę mechanizmu rozciągania przeprowadzamy używając diagramów scenariuszowych, które pokazują, dla jakiego przesunięciu d zadany kontakt jest zerwany po raz ostatni. Kontakt identyfikujemy poprzez odpowiadającą mu odległość sekwencyjną. Przyjmujemy, iż kontakt jest zerwany gdy odległość l_e zdefiniowana w sekcji 3.1.1 i wyznaczona osobno dla każdego modelu jest mniejsza niż r_{ij}. Zauważmy, iż maksimum siły zazwyczaj odpowiada równoczesnemu zrywaniu całej grupy kontaktów. Przykaładowy scenariusz jest pokazany na rysunku 4.1.

4.2 Wyniki — wybór modelu

4.2.1 Wybór temperatury dla symulacji rozciągania

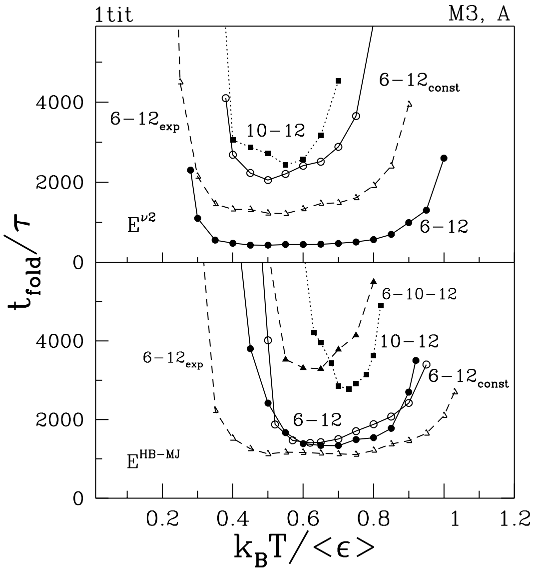
Istnieje kilka charakterystycznych temperatur, które mogą być powiązane ze stanem białka. Należą do nich między innymi $\tilde{T}_{min}$, $\tilde{T}_f$, i $\tilde{T}_{max}$. Pierwsza z nich charakteryzuje własności kinetyczne i określa temperaturę, w której proces zwijania jest najszybszy. Czas zwinięcia jest wyznaczony utworzeniem się po raz pierwszy wszystkich natywnych kontaktów. (Kinetykę procesu zwijania omówimy oddzielenie w paragrafie 5.) Druga odpowiada temperaturze zwijania, w której prawdopodobieństwo posiadania wszystkich natywnych kontaków wynosi $\frac{1}{2}$. (Własnośći termodynamiczne dla różnych modeli omówimy oddzielnie w paragrafie 4.3.) Trzecia temperatura określa położenie maksimum na wykresie ciepła właściwego w zależności od temperatury. W idealnym modelu te trzy temperatury powinny być zbliżone do siebie. Jednakże dla większości rozważanych w tej pracy modeli istnieje pewne przesunięcie pomiędzy tymi wartościami. Z tego powodu poniżej przeprowadzamy dyskusję nad wyborem temperatury dla rozciągania w danym modelu.

Zauważmy, iż $\tilde{T}_{min}$ i przede wszystkim $\tilde{T}_f$ zależą bezpośrednio od kryterium dynamicznego istnienia kontaktu $l_{e,ij}$. Natomiast $\tilde{T}_f$ i $\tilde{T}_{max}$ dodatkowo zależą także od czasu trwania symulacji i statystyki. W niniejszej pracy postulujemy, iż symulacje rozciągania powinny być przeprowadzane w pobliżu lub dokładnie w temperaturze T_{min}, gdyż ta temperatura odpowiada kinetycznie optymalnemu stanowi dla modelowanego białka, i w niej własności modelu są najbardziej sprzyjające białku. W wielu modelach, włącznie z modelem z jednorodną skalą energii i potencjałem Lennarda-Jonesa, wybór ten jest blisko efektywnej temperatury pokojowej [IV,VI], podczas gdy wartość $\tilde{T}_{max}$ często leży znacznie powyżej.

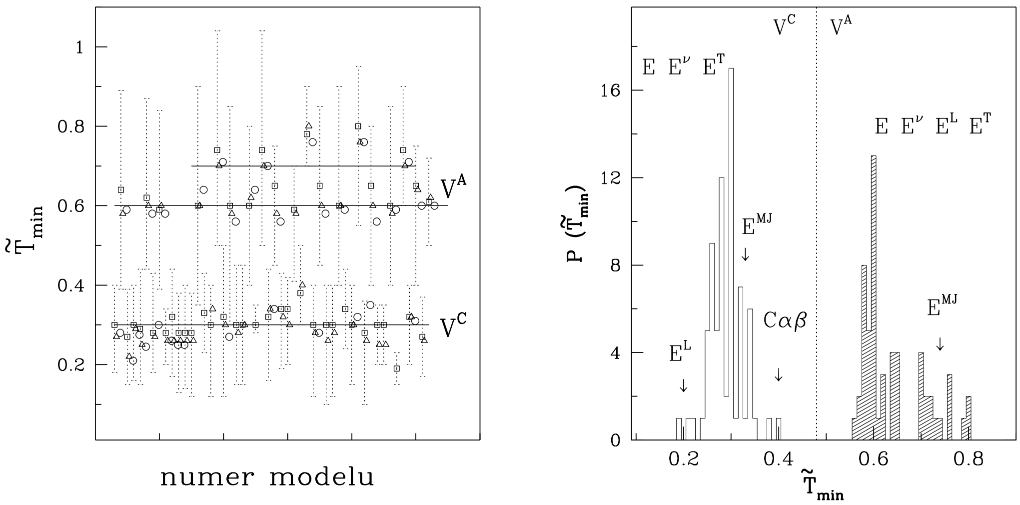


Rysunek 4.2: Zależność czasów zwijania od temperatury w różnych modelach Go dla krambiny. Lewe panele: model z jednorodną skalą energii E^o. Prawe panele: modele z niejednorodną skalą energii $E^{\nu 2}$ i $E^{HB,MJ}$. Panel górny odpowiada potencjałowi V^C, a panel dolny potencjałowi V^A. Linia ciągła z pełnymi kołami reprezentuje model z potencjałem V^{6-12}, linia ciągła z pustymi kołami V^{6-12}_{const}, linia przerywana z pełnymi kwadratami V^{10-12}, linia kropkowa z pełnymi trójkątami $V^{6-10-12}$, linia kropkowana z pustymi trójkątami V^{6-12}_{exp}.

Rysunki 4.2 i 4.3 pokazują czasy zwijania w funkcji temperatury dla białek krambiny i tytyny w zależności od parametrów modelu. W tej grupie modeli w każdym przypadku t_{fold} przyjmuje charakterystyczny kształt litery U, którego środek wyznacza położenie $\tilde{T}_{min}$. Wspólnie zestawiając te krzywe widzimy wyraźne rozróżnienie pomiędzy modelami, dla których do opisu lokalnej szytwności jest użyty potencjał V^C bądź V^A. W pierwszym przypadku $\tilde{T}_{min}$=0.3, podczas gdy dla drugiego potencjału $\tilde{T}_{min}$ jest znacznie wyżej i nie przyjmuje jednej charakterystycznej wartości. W przypdku krambiny, wartość $\tilde{T}_{min}$ jest dwukrotnie większa.



Rysunek 4.3: Zależność czasów zwijania od temperatury w różnych modelach dla tytyny. Oznaczenia takie same jak na rysunku 4.2.



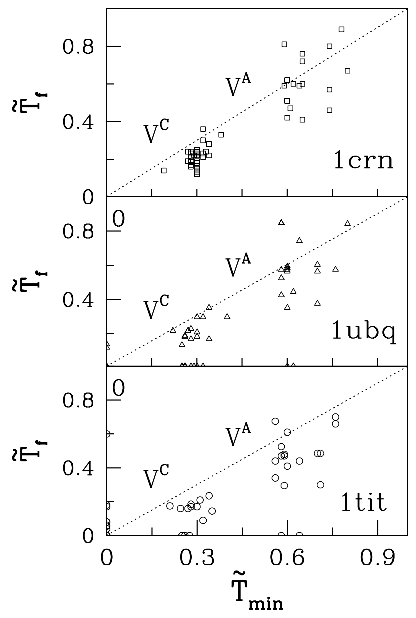
Rysunek 4.4: Lewy panel: wartość $\tilde{T}_{min}$ dla białek 1crn (kwadraty), 1ubq (trójkąty), 1tit (koła) na podstawie 61 modeli. Kropkowane linie wskazują przybliżony zakres temperatur, w których kinetyka zwijania jest optymalna. Linie ciągłe pokazują główne trędy w wartościach teperatur, które odpowiadają $\tilde{T}_{min}$=0.3, 0.6, 0.7. Prawy panel: rozkład wartości $\tilde{T}_{min}$ w rozważanych modelach, gdzie puste i cieniowane histogramy odpowiadają odpowiednio modelom z V^C i V^A.

Rysunek 4.4 podsumowuje wyniki $\tilde{T}_{min}$ i przybliżoną wartość zakresu temperatury, dla której proces zwijania jest optymalny, i które otrzymaliśmy z symulacji zwijania białek w 61 modelach dla trzech białek krambiny, tytyny i ubikwityny. Znajdujemy, iż $\tilde{T}_{min}$ dla różnych modeli ściśle się skupia wokół wartości 0.3, jeśli w modelu był użyty potencjał V^C. Wyjątek stanowi uogólniony model Go ($C^{\alpha-\beta}$) z $\tilde{T}_{min}$. Zauważmy, iż zastąpienie V^C poprzez V^A powoduje powstanie dodatkowego rozrzutu w wartościach $\tilde{T}_{min}$ wokół trzech wartości zredukowanej temperatury: 0.6, 0.7 (dla V^{6-12},E^{MJ}), i 0.8 (dla $V^{6-10-12}$ lub V_e^{6-12} z E^{MJ}).

Podsumowując, ilekroć w niniejszej pracy używamy modelu Go (C^{α}) wraz z V^C, rozciąganie białek jest wykonywane dla temperatury zredukowanej $\tilde{T} = 0.3$ (jeśli nie precyzujemy konkretnie innej temperatury). W modelach Go z grupą boczną $C^{\alpha-\beta}$ z V^C rozciąganie wykonujemy dla $\tilde{T}$= 0.4. W modelach zawierających potencjał V^A rozciąganie jest przeprowadzane dla $\tilde{T}$ 0.6, 0.7, 0.8, które są oznaczone odpowiednio symbolami *, ** i ***.

Otrzymane przesunięcie wartości $\tilde{T}_{min}$ w stronę wyższych temperatur po zastąpieniu V^C przez V^A w rozważanym modelu, wynika z istnienia dodatkowego usztywniającego członu w potencjale V^A. W podrozdziale 3.1.2 pokazaliśmy, iż jest to siła związana z więzami nałożonymi na kąty płaskie pomiędzy kolejnymi atomami C^i. Całkowita siła od potencjału V^A jest złożeniem V^C i potencjału od kątów płaskich, zgodnie ze wzorem (3.17). Rysunek 4.5 pokazuje zależność $\tilde{T}_{min}$ od $\tilde{T}_f$, z którego wyraźnie widzimy, iż dodatkowa sztywność w V^A prowadzi zarówno do większej termodynamicznej stabilności (większej wartość $\tilde{T}_f$) i wyższej wartości temperatury optymalnego zwijania $\tilde{T}_{min}$. Zerowe wartości $\tilde{T}_{min}$ na tym wykresie krzyżowym oznaczają modele, które nie prowadzą do zwinięcia co najmniej połowy rozważanych trajektorii. Zerowe wartości $\tilde{T}_f$ oznaczają modele, w których $\tilde{T}_f$ jest mniejsze niż 0.05.

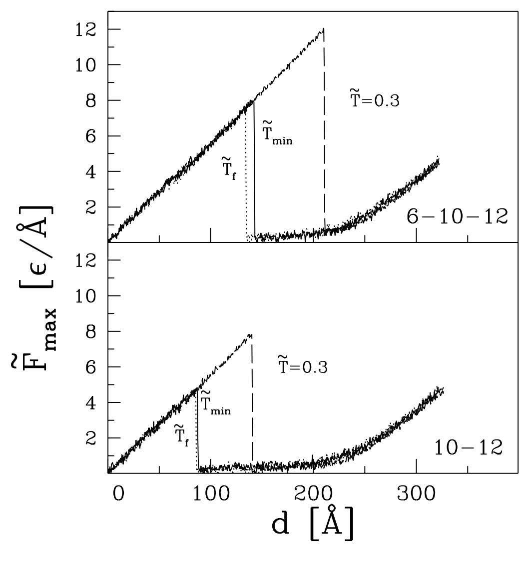
Powyżej scharakteryzowaliśmy 5 wartości temeperatury zredukowanej dla symulacji rozciągania. Zauważmy jednakże, że stopień w jakim wartość F_{max} i kształt $F(d)$ (przykładowo pokazany na rysunku 4.1) są czułe na wybór temperatury, zależy od parametrów modelu wymienionych w tablicy 3.1. W przypadku gdy lokalna sztywność jest opisywana przez V^A, zakres optymalnych temperatur jest dużo szerszy niż dla V^C. Ponadto znajdujemy, iż największą czułość na wybór temperatury wykazują modele z potencjałem $V^{6-10-12}$, a najmniejszą potencjał V^{10-12}; obydwa ze skalą energii $E^{HB,MJ}$, których wykrsy $F(d)$ są pokazane na rysunku 4.6. Dla pierwszego z tych potencjałów, aby otrzymać zgodność teoretycznej krzywej $F(d)$ (jej kształtu i położenia F_{max}) z doświadczalnymi wynikami, symulacja powinna być wykonana dla $\tilde{T}_f$ ∼0.75, która jest powyżej zaproponowanej przez nas $\tilde{T}_{min}$. W przypadku drugiego potencjału $\tilde{T}_f$ jest bardzo blisko $\tilde{T}_{min}$, tak więc nie obserwujemy tutaj żadnej zmiany. Zauważmy natomiast zasadniczą różnicę we własnościach białka dla rozciągania znacznie poniżej $\tilde{T}_{min}$, np. dla wartości 0.3. Warto też wspomnieć, iż czułość na wybór temperatury



Rysunek 4.5: Korelacja pomiędzy $\tilde{T}_{min}$ i $\tilde{T}_f$ dla 1crn (górny panel), 1ubq (środkowy panel) i 1tit (dolny panel).

także zależy od klasy białka. Wydaje się, że w przypadku białek helikalnych lepszą zgodność z doświadczeniem otrzymujemy, gdy rozciąganie jest przeprowadzane poniżej $\tilde{T}_{min}$.

Powyższa dyskusja miała na celu pokazanie, że wybór temperatury rozciągania może zmieniać znacząco korelację R^2 dla danego modelu. W kilku modelach rozciąganie przeprowadzamy dla różnych temperatur, a do wyboru optymalnego modelu używamy temperatury, w której R^2 jest największe. Rozciąganie tego samego modelu dla różnych temperatur jest oznaczone odpowiednią gwiazdką w tablicach z wynikami.



Rysunek 4.6: Zależność kształtu $F(d)$ od temperatury rozciągania w modelu z V^A. Kropkowane, ciągłe i przerywane linie odpowiadają rozciąganiu tytyny w temperaturze zwijania $\tilde{T}_f$, temperaturze optymalnego zwijania $\tilde{T}_{min}$ i $\tilde{T}$=0.3, która leży poza zakresem krzywych zwijania. Górny i dolny panel odpowiadają modelom z potencjałem 6−10−12 i 10 − 12.

Podsumowując, model skonstruowany z uwzględnienmiem potencjału V^C jest lepiej określonym modelem fizycznym. Dodatkowo jego atutem jest możliwość łatwego porównania różnych modeli, gdyż otrzymywany dla niego zakres temperatur optymalnego zwijania jest wąski i podobny dla różnych białek. Ponadto wartości $\tilde{T}_f$ i $\tilde{T}_{min}$ są zbliżone.

Tablica 4.2: Wyniki korelacji pomiędzy eksperymentalną a teoretycznie wyznaczoną wartością F_{max} na podstawie testowego zbioru 29 białek dla różnych wariantów modelu Go. Wszystkie modele zawierają potencjał $6-12$, pozostałe argumenty danego modelu są podane w nawisach $\{\}$. Współczynnik R^2 oznacza współczynnik korelacji, U — współczynnik Theil'a, a nachylenie prostej F^t_{max}=a $\cdot F^e_{max}$.

model	a	R^2	U
$\{C, M2, E^o\}$	0.0122	0.854	0.229
$\{C, M2c75, E^o\}$	0.0160	0.653	0.389
$\{C, M3c7, E^o\}$	0.0144	0.726	0.361
$\{C, M3c75, E^o\}$	0.0164	0.773	0.332
$\{C, M3, E^o\}$	0.0140	0.888	0.223
$\{C_{0.5}, M3, E^o\}$	0.0137	0.877	0.231
$\{A, M4, E^o\}$	0.0256	0.789	0.296
$\{A, M4, E^o\}$ **	0.0162	0.688	0.353
$\{C, CSU, E^o\}$	0.0150	0.858	0.212
$\{C, M3, E^{L1}\}$	0.0142	0.772	0.312
$\{C, M3, E^{L2}\}$	0.0146	0.821	0.274
$\{C, M3, E^{T2}\}$	0.0132	0.827	0.278
$\{A, M4, E^{T2}\}$	0.0250	0.767	0.309
$\{C, M3, E^{\nu 1}\}$	0.0142	0.792	0.230
$\{C, M3, E^{\nu 2}\}$	0.0147	0.671	0.361
$\{A, M4, E^{\nu 1}\}$ **	0.0156	0.667	0.365
$\{C, M3, E^{HB,MJ}\}$	0.0209	0.879	0.230
$\{A, M4, E^{HB,MJ}\}$	0.0332	0.833	0.271
$\{A, M4, E^{HB,MJ}\}$ **	0.0223	0.839	0.261

Tablica 4.3: Dane dla różnych modeli z potencjałem 6-12$_{const}$ — oznaczenia jak w tablicy 4.2.

model	a	R^2	U
$\{C, M2, E^o\}$	0.014	0.808	0.284
$\{C, Mc75, E^o\}$	0.016	0.777	0.331
$\{C, M3, E^o\}$	0.014	0.774	0.313
$\{A, M3, E^o\}$ *	0.020	0.764	0.309
$\{C, M3, E^{nu1}\}$	0.015	0.792	0.295
$\{C, M3, E^{HB,MJ}\}$	0.021	0.803	0.287

Tablica 4.4: Dane dla różnych modeli z potencjałem $6 - 10 - 12$ — oznaczenia jak w tablicy 4.2.

model	a	R^2	U
$\{ C, M2, E^o\}$	0.0186	0.828	0.274
$\{ C, M3c75, E^o\}$	0.0209	0.831	0.268
$\{ C, M3, E^o\}$	0.0154	0.807	0.284
$\{ A, M4, E^o\}$ **	0.0208	0.808	0.286
$\{ C, M3, E^{\nu 1}\}$	0.0190	0.668	0.367
$\{ C, M3, E^{HB,MJ}\}$	0.0268	0.761	0.317
$\{ A, M4, E^{HB,MJ}\}$	0.0494	0.866	0.240
$\{ A, M4, E^{HB,MJ}\}$ ***	0.0296	0.832	0.286

Tablica 4.5: Dane dla różnych modeli z potencjałem $6 - 12_{exp}$ — oznaczenia jak w tablicy 4.2.

model	a	R^2	U
$\{C, M3, E^o\}$	0.0153	0.854	0.250
$\{A, M3, E^o\}$ **	0.0179	0.791	0.297
$\{C, M3c75, E^o\}$	0.0185	0.768	0.328
$\{C, M3, E^{\nu 2}\}$	0.0164	0.816	0.277
$\{C, M3, E^{HB,MJ}\}$	0.0223	0.7739	0.306
$\{A, M3, E^{HB,MJ}\}$ ***	0.0241	0.831	0.269

Tablica 4.6: Dane dla różnych modeli z potencjałem 10-12 — oznaczenia jak w tablicy 4.2.

model	a	R^2	U
$\{ C, M2, E^o\}$	0.0167	0.751	0.323
$\{ C, M3, E^o\}$	0.0144	0.840	0.264
$\{ C, M3c75, E^o\}$	0.0169	0.791	0.326
$\{ A, M4, E^o\}$	0.0289	0.851	0.255
$\{ A, M4, E^o\}$ **	0.0174	0.731	0.329
$\{ C, M3, E^{\nu 1}\}$	0.0136	0.710	0.338
$\{ C, M3, E^{HB,MJ}\}$	0.0213	0.673	0.365
$\{ A, M4, E^{HB,MJ}\}$	0.0388	0.876	0.238
$\{ A, M4, E^{HB,MJ}\}$ ***	0.0242	0.816	0.280

4.2.2 Korelacja pomiędzy teoretyczną a eksprymentalną wartością F_{max}

Tablice 4.2, 4.3, 4.4, 4.5, 4.6, 4.7, 4.8 podsumowują statystyczną ocenę różnych wariantów modeli typu C^{α}, pogrupowanych w zależności od wyboru funkcjonalnej formy

Tablica 4.7: Dane dla różnych modeli z potencjałem Morse'a — oznaczenia jak w tablicy 4.2.

model	a	R^2	U
$\{ C, M3, E^o\}$	0.015	0.827	0.269
$\{ A, M3, E^o\}$	0.029	0.731	0.321
$\{ A, M3, E^o\}$ **	0.018	0.779	0.302

Tablica 4.8: Dane dla różnych modeli z mieszanym potencjałem $10-12$ z $6-12$ lub $6-12_{const}$ — oznaczenia jak w tablicy 4.2.

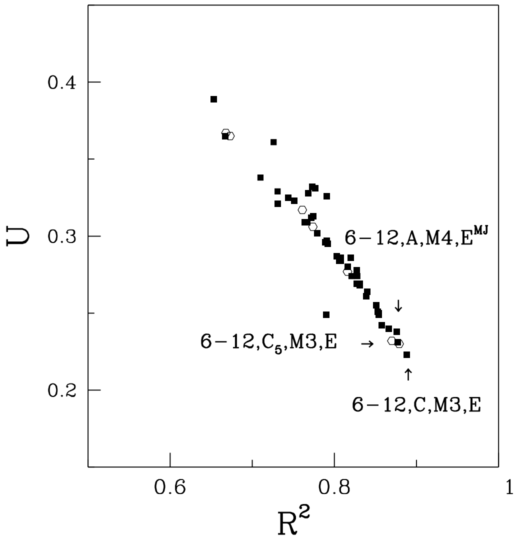
model	a	R^2	U
$\{6-12, C, M2, E^o\}$	0.014	0.870	0.232
$\{6-12, A, M3, E^o\}$ **	0.016	0.744	0.325
$\{6-12_{const}, C, M2, E^o\}$	0.015	0.796	0.319
$\{6-12_{const}, A, M3, E^o\}$ **	0.018	0.801	0.298

Tablica 4.9: Dane dla różnych wariantów modeli z grupami bocznymi $C^{\alpha-\beta}$ dla potencjałów $6-12$ i $10-12$ — oznaczenia jak w tablicy 4.2.

model	a	R^2	U
$\{ 6-12, C, M2, E^o\}$	0.0220	0.806	0.284
$\{ 6-12, C, M2, E^{\nu 2}\}$	0.0278	0.79	0.249
$\{ 10-12, C, M2, E^o\}$	0.0301	0.813	0.219

potencjału V^{NAT} dla natywnych kontaktów. Tablica 4.9 pokazuje podobne podsumowanie dla modeli typu $C^{\alpha-\beta}$ uwzględniające parametry: współczynnik korelacji Pearsona R^2 pomiędzy teoretycznie przewidywaną a doświadczalną wartością F_{max}, wartość a określającą nachylenie, i współczynnik U danym równaniem 4.4. Przed przystąpieniem do analizy wyników zauważmy, że dane eksperymentalne niekoniecznie zostały otrzymane dla tej samej prędkości rozciągania, podczas gdy wszystkie dane teoretyczne zostały uzyskane dla tej samej prędkości rozciągania. Dadatkowym utrudnieniem podczas oceny poprawności modeli jest fakt występowania więcej niż jednej drogi rozwijania, które także posiadają inne wartości F_{max}. Dla wszystkich białek, z wyjątkiem białka GFP, symulacja rozciągania jest wykonywana 10-krotnie i z nich wybieramy wartości F_{max} odpowiadające najbardziej typowej trajektorii rozwijania.

Reprezentacja graficzna wyników korelacji jest pokazana na rysunku 4.7, gdzie pozycję każdego modelu wyznaczają dwie wartości R^2 i U. Model jest tym lepszy, im R^2 jest większe i U mniejsze. Jednakże poprawny model powinien także posiadać rozsądne



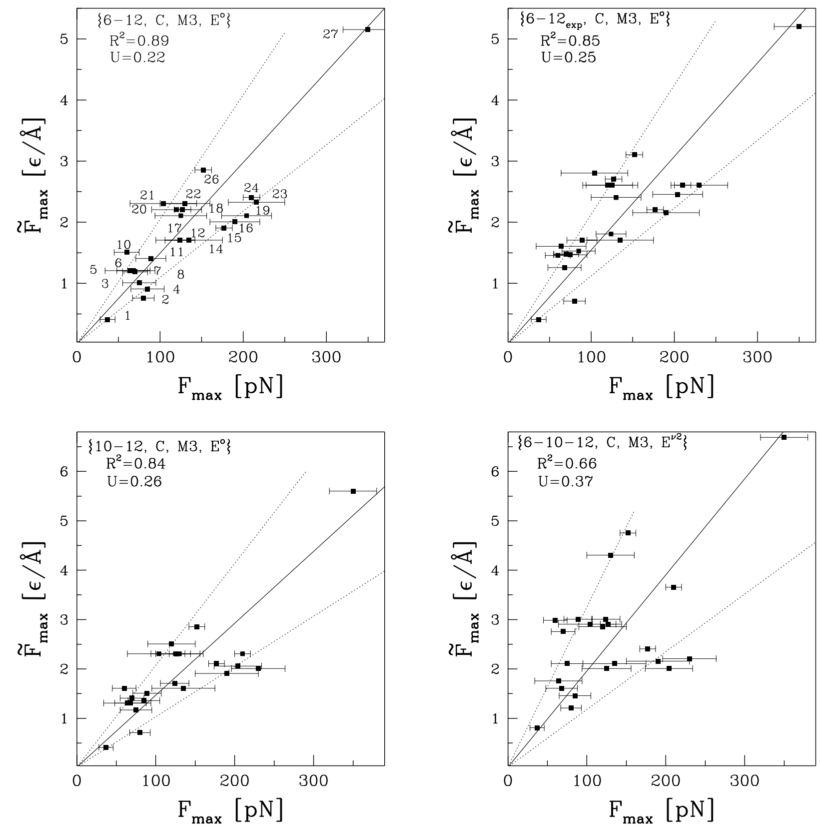
Rysunek 4.7: Zależność pomiędzy współczynnikiem korelacji Pearsona R^2 a współczynnikiem Theil'a U. Puste koła odpowiadają modelom, w których nie każde białko może być zwinięte.

własności zwijania (co najmniej połowa trajektorii powinna prowadzić do zwinięcia). Modele które prowadzą dla zwinięcia każdego z 3 białek (1crn, 1ubq,1tit) oznaczyliśmy czarnym pełnym kołem, podczas gdy modele, które nie pozwalają na zwinięcie chociaż jednego z tych białek, są oznaczone okręgami. Do tych odrzuconych modeli należą $\{6-12, C, M3, E^{HB,MJ}\}$ oraz $\{(6-12, 10-12), C, M3, E^{o,HB}\}$.

Otrzymaliśmy, iż cztery najlepsze modele zbudowane z potencjału chiralnościowego, które także umożliwiają zwijanie białek, odpowiadają następującym zbiorom parametrów:

1. { 6 − 12, C, M3, E^0 }
2. $\{\ 6-12_{exp}, C, M3, E^O\ \}$
3. $\{\ 10-12, C, M3, E^O\}$
4. $\{\ 6-10-12, C, M3, E^O\}$

dla których korelacja R^2 odpowiednio wynosi 0.89, 0.87, 0.85 i 0.84. Najlepszy model składa się z jednorodnej skali energii, potencjału Lennarda-Jones'a, sztywności łańcucha zadanej potencjałem chiralnościowym i mapy kontaktów nie uwzględniającej kontaktów dla $i, i+2$. Model o numerze 2 zawiera zamiast potencjału $6-12$ potencjał $6-12_{exp}$ z drugim minimum w odległości 3Å, pozostałe argumenty są takie same. Wykresy korelacji pomiędzy F^t_{max} a F^e_{max} dla najlepszych trzech modeli i jednego najgorszego są pokazne na rysunku 4.8. Podkreślmy, że w tej grupie nie znajdują się modele, które nie prowadzą do zwijania, jak np. model $\{\ (6-12, 10-12), C, M3, E^{o,HB}\}$, który mógłby być sklasyfikowany na drugim miejscu z korelacją R^2=0.87.



Rysunek 4.8: Korelacja pomiędzy eksperymentalną F_{max} a teoretyczną wartością $\tilde{F}_{max}$ dla trzech najlepszych i jednego najgorszego modelu, które zawierają potencjał V^C. Są to modele określone symbolami $\{6-12, C, M3, E^O\}$, $\{6\text{-}12_{exp}, C, M3, E^O\}$, $\{10\text{-}12, C, M3, E^O\}$, $\{6\text{-}10\text{-}12, C, M3, E^{\nu 2}\}$, o współczynniku korelacji R^2 odpowiednio 0.89, 0.85, 0.83 i 0.66 (najgorszy model). Symbole odpowiadają białkom 1(1n11), 2(1cfc), 3(1hci), 4($^{10}FNII$), 5(1u4q), 6(1aj3), 7(B), 8(1ubq$_{48-N}$), 9(1b6i), 10(1rsy), 11(^{13}FNII), 12(^{12}FNIII), 14(TNFNIII), 15(1qjo$_{N-41}$), 16(G), 17(^{1}FNII), 18(I1), 19(I27),20(1emb), 21(1emb$_{132-212}$), 22(1emb$_{3-212}$), 23(1ubq), 24(1nct), 26(L), 27(1emb$_{3-132}$, 28(1ksr), 29(1g1c)).

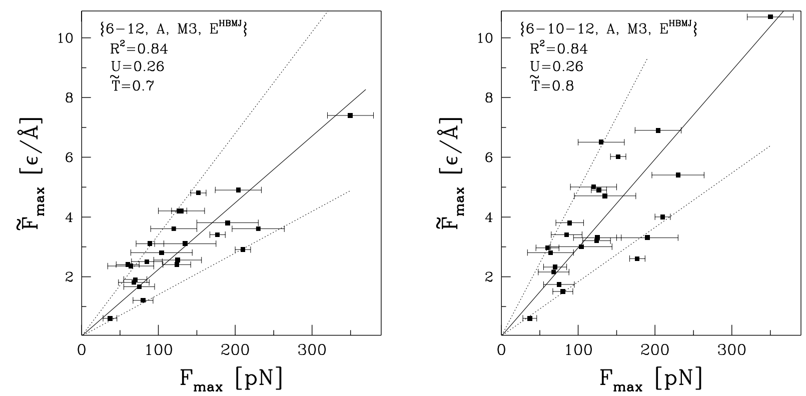
Dodatkowe informacje o modelach otrzymujemy po podzieleniu 29 białek na grupy odpowiadające czterem klasom strukturalnym. Znajdujemy, że jednorodna skala energii daje dość dobre wyniki (współczynnik korelacji R^2 ≃0.82) dla opisu białek z klasy $\alpha-\beta$, jednakże najlepszym wyjściem jest użycie skal energii E^{T1} lub T^{T2} gdzie współczynnik korelacji wynosi R^2 ≃89. Najgorsze korelacje dla białek klasy α daje skala energii E^{MJ} (R^2 ≃64)), podczas gdy użycie jej do opisu białek klasy β daje zadawalające wyniki (R^2 ≃83). Kolejną obserwacją jest znacząco lepsza korelacja wszystkich modeli gdy użyta jest mapa $M3$ zamiast $M2$, i dla obydwu map w porównaniu z którąkolwiek mapą typu Mc.

W gronie modeli z potencjałem $6-12$, dobrą korelację z wynikami eksperymentalnymi otrzymujemy także dla modeli ze skalami energii $E^{L1,\,L2}$, $E^{\,T1,\,T2,\,T3}$ jak i ich kombinacji. Skale te charakteryzują się siłą oddziaływania pomiędzy kontaktami pochodzącymi tylko z grup bocznych słabszą niż dla pozostałych natywnych kontatktów. Ponadto modele tego typu całkiem dobrze opisują proces rozwijania białek klasy α (1cfc,1aj3,1u4q,1n11).

Otrzymaliśmy, iż cztery najlepsze modele, w których lokalna sztywność łancucha białka jest opisana potencjałem kątowym V^A, i które także umożliwiają zwijanie białek, odpowiadają następującym zbiorom parametrów:

1. $\{6-12, A, M4, E^{HB,MJ}\}$
2. $\{6-10-12, A, M4, E^{HB,MJ}\}$ [25]
3. $\{10-12, A, M4, E^{O}\}$
4. $\{10-12, A, M4, E^{HB,MJ}\}$

Wśród tych modeli jednakże nie znajdujemy bardzo często używanego modelu o parametrach $\{6-12, A, M4, E^o\}$ [145], gdyż w jego przypadku współczynnik korelacji jest niższy, $R^2 \simeq 0.78$. Zestawienie otrzymanych z symulcji wartości F^t_{max} z ich odpowiednikiem doświadczalnym F^e_{max} dla najlepszych dwóch modeli jest pokazane na rysunku 4.9.



Rysunek 4.9: Dwa najlepsze modele z potencjałem V^A, $\{6-12, A, M4, E^{HB,MJ}\}$ i $\{6-10-12, A, M4, E^{HB,MJ}\}$ współczynnikiem korelacji R^2 opowiednio 0.839 i 0.832. Oznaczenia jak na rysunku 4.8.

Analiza otrzymanych współczynników R^2 w przypadku modeli z V^A pokazuje zgodnie z naszymi wcześniejszymi przypuszczeniami, iż jego wartość jest bardzo czuła na wybór zredukowanej temperatury dla symulacji rozciągania. Przypomnijmy, że optymalne zwijanie dla tego samego modelu z V^A może zachodzić nawet w zakresie temeperatur od 0.6 do 0.87. Tak więc, w przypadku modelu o parametrach $\{6-10-12, A, M4, E^0\}$, otrzymujemy korelację R^2 rzędu 0.82 z rozciągnięcia każdego białka ze zbioru S29 dla temperatury $\tilde{T}$=0.8, podczas gdy wykonanie tych symulacji dla temperatury $\tilde{T}$=0.5 daje korelacje aż R^2=0.86. Rozciąganie dla temperatury $\tilde{T}$=0.8 jest motywowane położeniem $\tilde{T}_{min}$ dla tytyny, a dla $\tilde{T}$=0.5 pochodzi z położenia $\tilde{T}_{min}$ dla białka 1aj3.

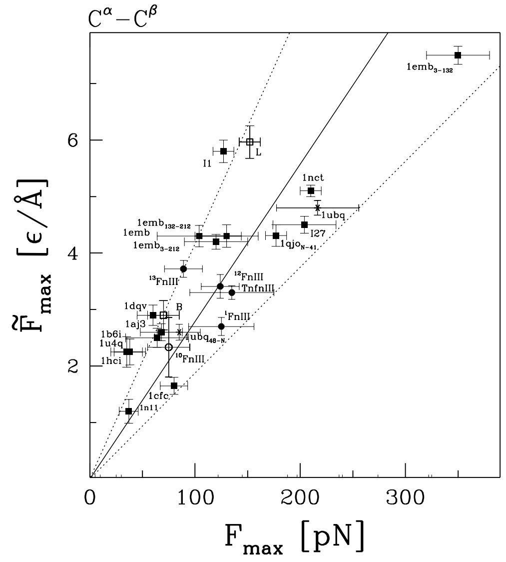
Najbardziej chaotyczne położenia F_{max} otrzymaliśmy dla modeli ze skalą energii $E^{\nu 2}$, i trochę lepsze dla $E^{\nu 1}$, w połączeniu z większością używanych V^{NAT} z V^C. Niską korelację daje także skala energii E^{MJ} z V^{10-12} i V^C. Modele te także nie prowadzą do zwinięcia każdego białka (np. 1tit).

Model z grupami bocznymi

Uogólniony model Go, w którym aminokwas (poza glicyną) reprezentujemy dwoma atomami C^α i C^β rozważamy na kilka sposobów, jednakże zawsze z potencjałem V^C. Dla tego modelu optymalne własności kinetyczne są obserwowane dla zredukowanej temperatury około 0.4 i ta temperatura jest użyta w symulacjach rozciągania. W przypadku tego typu modeli $\tilde{F}_{max}$ dla każdego białka jest średnią po trzech trajektoriach. Otrzymana wartość korelacji R^2 dla czterech modeli jest pokazana w tablicy 4.9. Najlepszą korelację z danymi doświadczalnymi w ramach modeli typu $C^{\alpha-\beta}$ daje skład argumentów $\{10-12, C, M3, E^o\}$, tym niemniej korelacja ta jest znacznie niższa niż dla czterech najlepszych modeli typu $C^{\alpha-\beta}$. Rysunek 4.10 pokazuje wykres krzyżowy F^{teo}_{max} z F^e_{max} w modelu $\{6-12, C, M3, E^o\}$. Podzielenie grupy 29 białek na grupy w zależności od klasy białka pokazuje, iż bardzo dobrą korelację (R^2=0.85) w klasie białek typu $\alpha-\beta$ daje model o parametrach $\{6-12, C, M3, E^{\nu 2}\}$, jednakże wydaje się, iż nie powinien on być zostosowany do opisu białek z klasy α.

4.2.3 Kształt krzywych $F(d)$

Zbiór wartości F_{max} charakteryzuje skalę siły, jednakże nie łącząc jej z całkowitym przebiegiem procesu rozwijania określonym między innymi przez kształt $F(d)$. Tak więc w proces identyfikacji optymalnego modelu włączamy także charakterystykę wzorów $F(d)$, gdyż w szczególności wprowadzenie różnych skal energii może prowadzić do



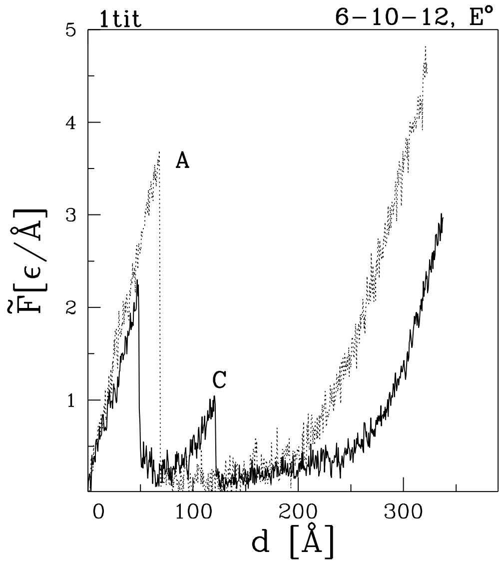
Rysunek 4.10: Wartość mechanicznego oporu $\tilde{F}_{max}$ otrzymana teoretycznie w modelu $C^{\alpha-\beta}$ o parametrach $\{6{-}12, C, M3, E^o\}$ w funkcji odpowiadającej jej wartości otrzymanej doświadczalnie F_{max}.

niepoprawnych trajektorii rozwijania.

Z naszych obliczeń wynika, iż użycie jednorodnej skali energii i V^C dla jakiejkolwiek formy potencjału dla oddziaływań natywnych prowadzi do podobnych kształtów wykresu $F(d)$. Rozciąganie np. tytyny w modelu $\{6-12, C, M3, E^o\}$ jest pokazane na rysunku 4.1, który zarazem jest charakterystyczny dla wszystkich modeli z parametrami $\{C, E^o\}$. Na wykresach tych położenie pierwszego największego maksimum zgadza się z wynikami doświadczalnymi, jednakże występowanie drugiego znacznie mniejszego maksimum jest defektem modeli. Ten defekt znika podczas rozciągania tandemu tytyn.

Uwzględnienie w modelach potencjału V^A zamiast V^C powoduje pojawienie się dużej różnorodności w kształtach wzorów $F(d)$, która zależy zarazem od doboru skali energii jak też formy potencjału V^{NAT}. Po pierwsze użycie jednorodnej skali energii, potencjału $6-10-12$ iA w przypadku tytyny prowadzi do nie pojawienia się niepożądanego drugiego maksimum na wykresie $F(d)$, zgodnie z wynikiem pokazanym na wykresie 4.11.

Po drugie, niejednorodne skale energii w zestawieniu z V^A prowadzą do dużej różnorodności w otrzymywanych kształtach $F(d)$, zwłaszcza dla białek klasy α. Największą rozbieżność w porównaniu do znanych wykresów $F(d)$ otrzymywanych doświadczalnie lub z symulacji pełnoatomowych otrzymujemy dla skal energii $E^{\nu 2}$. Najlepszą zgodność dają skale energii $E^{HB,MJ}$. Przykłady zmian kształtu wzoru $F(d)$ dla tytyn pod wpływem zmiany skali energii są pokazane na wykresach rysunku 4.12. Ponadto połączenie skali $E^{\nu 2}$ z V^{6-12} lub $V^{6-10-12}$ z V^C prowadzi do kształtu $F(d)$, który w żadnym wy-



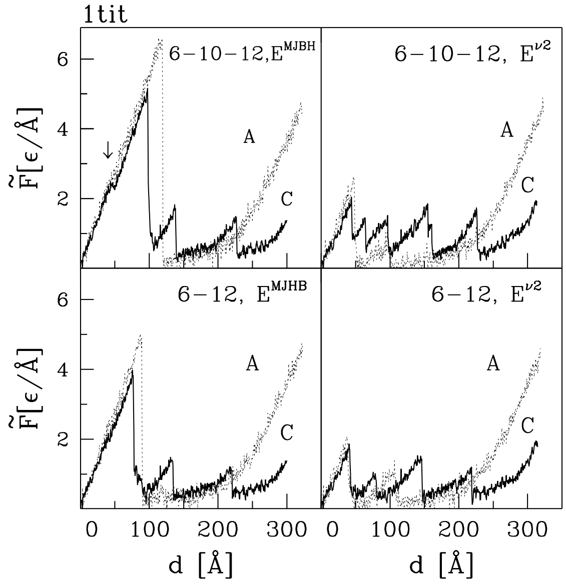
Rysunek 4.11: Zależność kształtu $F(d)$ od potencjału S dla jednorodnej skali energii z potencjałem $6-10-12$ Przerywana i ciągła linia odpowiadają opowiednio modelowi z potencjałem kątowym A i chiralnościowym C.

padku nie przypomina doświadczalnie otrzymywanych krzywych, w szczególności dla białek 1tit, 1ubq, i wszystkich domen fibronektyn typu III ze zbioru S29.

Skale energii $E^{HB,MJ}$ z V^C zazwyczaj wprowadzają do kształtu $F(d)$ ząbkowatość, obserwowaną poprzez pojawianie się dodatkowych małych maksimów. Szczegółowa analiza danych pokazuje, iż zestawienie tej skali energii z potencjałem $V^{6-10-12}$ lub V^{6-12}_{const} powoduje pojawienie się w kształcie $F(d)$ dla tytyny dodatkowego malutkiego maksimum na wzrastającej stronie największego maksimum, co doskonale zgadza się z wynikiem doświadczalnym [160]. Wynik ten pokazuje, iż obserwowana ząbkowatość może mieć znaczenie fizyczne.

Pokazaliśmy już, że jednorodna skala energii prowadzi do dodatkowego nadmiarowego małego maksimum na wykresie $F(d)$ dla tytyny, które znika przy rozciąganiu kilku domen naraz. Jednakże zastąpienie jednorodnej skali energii przez skale $E^{HB,MJ}$ powoduje także zniknięcie nadmiarowego maksimum dla wszystkich potencjałów V^{NAT} poza V^{6-12}_{exp},

W podrozdziale 4.2.2 pokazaliśmy, że korelacja R^2 pomiędzy F^t_{max} a F^e_{max} jest niska dla modeli ze skalami energii E^{T1}, E^{T2} i E^{L1}, E^{L2}. Tym niemniej okazuje się, iż skale te formują kształt $F(d)$ najbliższy doświadczalnemu dla białek klasy α (1aj3, 1cfc, 1u4q). Wynik ten jest zgodny z dużą wartością R^2 jeśli rozważy się oddzielnie tylko zbiór białek helikalnych, jednakże wartość R^2 jest niska dla całej grupy 29 testowych białek.

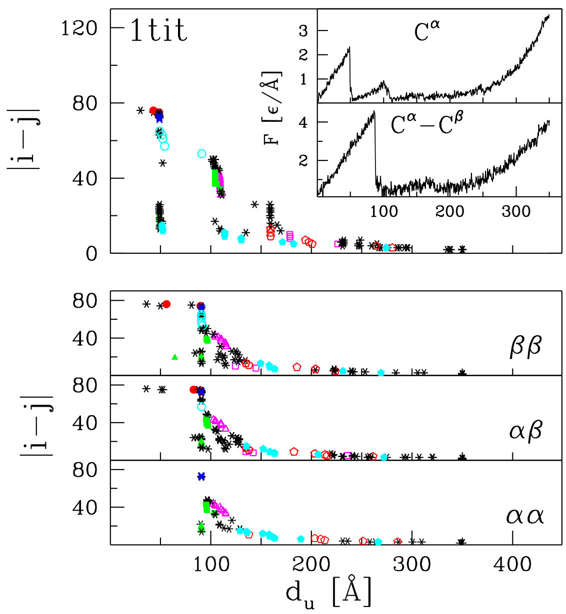


Rysunek 4.12: Zależność kształtu $F(d)$ od wyboru potencjału S dla modeli z niejednorodną skalą energii. Linie kropkowane i ciągłe odpowiadają odpowiadają odpowiednio potencjałom A i C. Lewy i prawy panel powstały odpowiednio dla skal energii $E^{HB,MJ}$ i $E^{\nu 2}$. Górne panele przedstawiają modele z potencjałem $6-10-12$, dolne panele z potencjałem $6-12$.

Kształt $F(d)$ w modelu z grupami bocznymi, $C^{\alpha-\beta}$

Okazuje się, że największą zmianę w kształcie $F(d)$ w całym zbiorze testowym powoduje uwzględnienie w modelu właśnie dodatkowego atomu C^{β}. W grupie 4 testowanych modeli typu $C^{\alpha-\beta}$ obserwujemy kilka typów zmian: zniknięcie jednego ze średniej wielkości maksimów (zazwyczaj białka klasy β np. 1tit, 1fnf, 1ttg), pojawienie się nowych maksimów (białka klasy α np. 1aj3, 1u4q), zmianę położenia głównego maksimum — inne drogi rozwijania niż w modelu C^{α} (1ten).

Rysunek 4.13 pokazuje najbardziej typową zmianę kształtu $F(d)$ w porównaniu do modelu typu C^{α}, o parametrach $\{6-12, C, M3, E^{o}\}$ i odpowiadający temu przejściu diagram scenariuszowy na przykładzie białka tytyny. Diagram scenariuszowy dla rozrywania kontaktów w modelu $\{C^{\alpha}, 6-12, C, M3, E^{o}\}$ jest pokazany na rysunku 4.1. Przypomnijmy, iż drugie maksimum w modelu C^{α} wynika z rozerwania kontaktów pomiędzy nićmi β, C+F i B+E. Zniknięcie tego maksimum w modelu z C^{β} wynika z jednoczesnego zerwania także i tych kontaktów w pierwszym maksimum. Jest to skutkiem po pierwsze zmiany siły oddziaływania między niektórym aminokwasami, po drugie większej kooperatywności pomiędzy elementami struktury.



Rysunek 4.13: Scenariusze pękania kontaktów podczas mechanicznego rozciągania dla tytyny otrzymane w modelu typu C^{α} (górny panel) i $C^{\alpha-\beta}$ (dolny panel), w ich temperaturach optymalnego zwijania. Rysunek dolny jest podzielony na trzy panele, które pokazują scenariusze rozwijania dla kontretnego typu kontaktów: $C^{\beta}-C^{\beta}$, $C^{\alpha}-C^{\beta}$ i $C^{\alpha}-C^{\alpha}$ odpowiednio od górnego do dolnego panelu. Kształty i kolory symboli są takie same jak na rysunku 4.1. Panel wewnętrzny w górnym panelu przedstawia krzywe $F(d)$ odpowiadające modelom C^{α} (górny panel) i $C^{\alpha-\beta}$(dolny panel).

Podsumowanie

W rozdziale tym rozważyliśmy 61 najbardziej obiecujących wariantów modeli z 504 możliwych. Otrzymaliśmy grono modeli o najwyższej korelacji gdzie wartość R^2 jest najwyższa dla modelu o parametrach $\{6-12, C, M3, E^{O}\}$. Dodatkowo analiza kształtów $F(d)$ pokazała, że model złożony z jednorodnej skali E^{o}, potencjału V^{6-12} i potencjału V^{C} najlepiej odzwierciedla znane z eksperymentu trajektorie rozwijania dla całego zbioru 29 białek. Ponadto, czego wcześniej nie dyskutowaliśmy a pokazaliśmy w pracy [VIII], zmiana siły od potencjału V^{C} jak też zamiana mapy kontaktów na mapę CSU daje taki sam lub gorszy współczynnik korelacji R^2. Wartość $\tilde{F}_{max}$ dla każdego białka z grupy 29 testowych otrzymana w optymalnym modelu jest zawarta w tablicy 4.1. Duży błąd wyznaczenia siły dla białek 1nct, 1pga, 1emb(132-212), 1emb(3-212) i wszystkich białek typu FNIII, reprezentuje zakres najniższej i najwyższej wartości $\tilde{F}_{max}$. Rozrzut w sile większy niż $0.2\epsilon/\text{Å}$ odpowiada różnym ścieżkom rozwijania.

Należy jednak pokreślić, że do badania poszczególnych klas jak i pojedynczych specyficznych białek, model złożny z innych argumentów może być bardziej adekwatny.

4.3 Wielkości termodynamiczne

W niniejszej części tego rozdziału dyskutujemy poprawność modeli poza procesem rozciągania. Konstrukcja modeli Go powoduje, iż nie są one najlepszym narzędziem do badania własności układu daleko od stanu natywnego, tak więc takie obserwable jak doświadczalnie wyznaczona temperatura zwijania czy czas zwijania nie mogą być użyte jako kryterium poprawności modelu. Jednakże ważnym elementem charakterystyki modeli jest także poznanie ich własności poza stanem natywnym, a tym bardziej modeli wybranych jako optymalne do symulacji rozciągania białek.

Modele typu C^{α}

Teoretycznie wyznaczone wartości $\tilde{T}_f$, $\tilde{T}_{min}$ i t_{fold} dla 1crn, 1ubq, 1tit, dla 61 wariantów modeli są pokazane w tablicach 4.10, 4.11, 4.12, 4.13, 4.14, 4.15, 4.16 i 4.17. Otrzymaliśmy, iż tylko małe białko klasy α/β może być zwinięte do swojego stanu natywnego poprzez każdy wariant modelu. Natomiast w przypadku białka tytyny tylko 35 wariantów modeli prowadzi tytynę do stanu natywnego.

Punktem wyjścia do dyskusji danych zawartych w wyżej wymienionych tablicach jest praca [23]. W pracy tej na podstawie 51 różnych białek pokazano, że wartość $\tilde{T}_{min}$ w modelu o parametrach $\{6-12, M2, C, E^o\}$ zależy od klasy i od długości białka N, ale jednocześnie większość białek charakteryzuje się optymalną temperaturą zwijania wokół $\tilde{T} = 0.3$. W przypadku modelu o parametrach $\{10-12, M2, C, E^o\}$, wartości $\tilde{T}_{min}$ były skoncentrowane wokół 0.225 i okazują się nie zależeć od parametru długości białka. W pracy [23] pokazano także, iż dla obu wariantów modeli $\tilde{T}_f$ znajduje się nieznacznie poniżej $\tilde{T}_{min}$, jednakże przerwa ta jest zauważalnie większa dla modelu z potencjałem $10-12$. Poniżej pokażemy wyniki podobnego porównania pomiędzy 61 wariantami modeli, jednakże tylko na podstawie 1crn, 1ubq i 1tit.

Rysunek 4.5 pokazuje korelację pomiędzy $\tilde{T}_{min}$ i $\tilde{T}_f$ dla każdego z białek oddzielnie. W każdym przypadku otrzymaliśmy wyraźny podział wyników na modele z potencjałem V^C i V^A. Modele, w których sztywność łańcucha opisywana jest potencjałem V^C, charakteryzują się wartością $\tilde{T}_f$ porównywalną z $\tilde{T}_{min}$, chociaż dla modeli o potencjale $10-12$, $6-10-12$ i skali energii E^{ν} rozrzut pomiędzy tymi temperaturami jest większy. Modele, w których zamiast potencjału V^C jest użyty V^A, charakteryzują się bardzo dużym rozrzutem w położeniu punktów odpowiadających relacji $\tilde{T}_f$ — $\tilde{T}_{min}$. Ponadto w przypadku kilku modeli wartość $\tilde{T}_f$ jest większa niż $\tilde{T}_{min}$; są to punkty powyżej diagonali na rysunku 4.5. Punkty poniżej diagonali odpowiadają przeciwnej realcji. Otrzymaliśmy także, że mimo iż w niektórych modelach $\tilde{T}_f$ jest mniejsze niż $\tilde{T}_{min}$, czasy zwijania dla temperatur $\tilde{T}_f$ są porównywalne do tych uzyskiwanych dla zwijania w $\tilde{T}_{min}$. Wynik ten pokazuje, że modele te są niesfrustrowane i prowadzą do zwinięcia w skończonym czasie. Powyższy wniosek dotyczy także modeli z niejednorodną skalą energii, które okazywały się być kiepskimi modelami do opisu procesu mechanicznego rozciągania białek. Tym niemniej zauważmy, że istnieją takie modele jak V^{6-12}_{const} z E^{MJ}

bez względu na dobór parametru potencjału V^C lub V^A, i V^{10-12} z E^{MJ} lub $E^{\nu 1,\nu 2}$, dla których czas zwijania t_{fold} jest zbyt długi do wyznaczenia w ramach naszych symulacji, zwłaszcza dla białek 1tit,1ubq. Niemożność wyznaczania czasu zwijania w modelach okazuje się być skorelowana dla większości z tych modeli z bardzo niską wartością $\tilde{T}_f$.

Sprawdziliśmy także proporcję pomiędzy położeniem $\tilde{T}_{max}$ (określającym położenie maksimum na wykresie ciepła właściwego w funkcji temperatury) a $\tilde{T}_f$ i $\tilde{T}_{min}$ (aczkolwiek wartość $\tilde{T}_{max}$ nie jest pokazana w tablicach). Największą rozbieżność pomiędzy $\tilde{T}_{max}$ a pozostałymi temperaturami (około 0.35) obserwujemy dla modeli z potencjałem V^C, chociaż w kilku przypadkach $\tilde{T}_{max}$ mieści się w granicy temperatur, w których kinetyka zwijania jest optymalna. W przypadku modeli z potencjałem V^A, zazwyczaj różnica pomiędzy $\tilde{T}_f$ i $\tilde{T}_{min}$ a $\tilde{T}_{max}$ jest wyraźnie mniejsza (około $\tilde{T}$=0.25), a w kilku przypadkach wynosi $\tilde{T}$=1.5. Jednakże wśród 61 wariantów modeli istnieją modele, dla których trzy temperatury $\tilde{T}_f$, $\tilde{T}_{min}$ i $\tilde{T}_{max}$ przyjmują bardzo wąski charakterystyczny zakres temperatur. Są to modele złożone z parametrów $6-10-12, A, M4$ ze skalami energii E^o, E^{MJ} lub $E^{\nu 2}$. Zbieżnosc trzech temperatur jest także połączona z wydajną kinetyką tych modeli; $\tilde{T}_f$ i $\tilde{T}_{max}$ znajdują się w górnym zakresie temperatur optymalnego zwijania. Należy pokreślić, iż zbieżność tych temperatur zależy także od własności rozważanego białka.

Modele typu $C^{\alpha-\beta}$

Poprzednio pokazaliśmy, iż żaden wariant modelu Go typu $C^{\alpha-\beta}$ nie jest w ścisłej czołówce najlepszych modeli do symulacji rozciągania białek. Jednakże wykazaliśmy jego pozytywny wpływ na poprawność kształtu $F(d)$ i kolejność rozrywania kontaktów. Ponadto otrzymujemy, że wzajemne położenie temperatur $\tilde{T}_f$, $\tilde{T}_{min}$ i $\tilde{T}_{max}$ w tym modelu jest bardziej zbliżone niż w modelu typu C^{α} o tych samych parametrach, mimo że zazwyczaj zakres temperatur optymalnego zwijania jest węższy. Wyniki te podnoszą rangę tego modelu.

Tablica 4.10: Temperatura zwijania $\tilde{T}_f$ oraz temperatura optymalnego zwijania $\tilde{T}_{min}$ wraz z czasem zwijania t_{fold}/τ, otrzymane w różnych modelach z potencjałem $6-12$ dla 1crn, 1ubq and 1tit. Symbole $-$ oznaczają niemożność wyznaczenia w danym modelu czasu zwinięcia białka.

model	1crn			1ubq		1tit	
	$\tilde{T}_f$	$\tilde{T}_{min}$	t_{fold}/τ	$\tilde{T}_{min}$	t_{fold}/τ	$\tilde{T}_{min}$	t_{fold}/τ
$\{ C, M3,\ E^o\}$	0.24	0.3	256	0.27	460	0.25	3926
$\{ A, M4, E^o\}$	0.59	0.64	243	0.58	459	0.59	1055
$\{ C, M3, E^{L1}\}$	0.24	0.28	328	0.22	773	0.21	4100
$\{ C, M3, E^{L2}\}$	0.25	0.30	337	0.29	824	0.27	2817
$\{ C, M3, E^{T1}\}$	0.22	0.29	318	0.25	747	0.24	6300
$\{ A, M4, E^{T1}\}$	0.60	0.62	216	0.60	419	0.58	1283
$\{ C, M3, E^{T2}\}$	0.23	0.28	318	0.27	696	0.30	5038
$\{ A, M4, E^{T2}\}$	0.59	0.59	225	0.60	460	0.58	1039
$\{ C, M3, E^{T3}\}$	0.21	0.28	351	0.26	766	0.26	4500
$\{ C, M3, E^{T1,L1}\}$	0.23	0.32	317	0.26	755	0.25	4947
$\{ C, M3, E^{T2,L1}\}$	0.24	0.28	320	0.26	778	0.25	3800
$\{ C, M3, E^{\nu1}\}$	0.17	0.28	336	0.26	771	-	-
$\{ C, M3, E^{\nu2}\}$	0.16	0.28	322	0.26	751	-	-
$\{ A, M4, E^{\nu2}\}$	0.62	0.60	209	0.60	443	0.64	1561
$\{ C, M3, E^{HB,MJ}\}$	0.22	0.30	300	0.34	1060	-	-
$\{ A, M4, E^{HB,MJ}\}$	0.80	0.74	229	0.70	348	0.72	1337

Tablica 4.11: Dane dla różnych modeli z potencjałem $6-12_{const}$ — oznaczenia jak w tablicy 4.10.

	1crn			1ubq		1tit	
model	$\tilde{T}_f$	$\tilde{T}_{min}$	t_{fold}/τ	$\tilde{T}_{min}$	t_{fold}/τ	$\tilde{T}_{min}$	t_{fold}/τ
$\{C, M3, E^o\}$	0.21	0.32	265	0.3	563	0.27	6647
$\{A, M3, E^o\}$	0.51	0.60	247	0.58	544	0.56	1296
$\{C, M3, E^{\nu1}\}$	0.14	0.3	388	0.28	908	-	-
$\{C, M3, E^{\nu2}\}$	0.15	0.3	375	0.3	915	-	-
$\{A, M4, E^{\nu2}\}$	0.51	0.6	214	0.62	563	0.64	2680
$\{C, M3, E^{HB,MJ}\}$	0.19	0.3	664	-	-	-	-
$\{A, M4, E^{HB,MJ}\}$	0.57	0.74	338	0.74	905	0.70	1500

Tablica 4.12: Dane dla różnych modeli z potencjałem $6-10-12$ — oznaczenia jak w tablicy 4.10.

	1crn			1ubq		1tit	
model	$\tilde{T}_f$	$\tilde{T}_{min}$	t_{fold}/τ	$\tilde{T}_{min}$	t_{fold}/τ	$\tilde{T}_{min}$	t_{fold}/τ
$\{C, M3, E^o\}$	0.36	0.32	603	0.34	1053	0.34	7500
$\{A, M4, E^o\}$	0.76	0.65	519	0.58	851	0.56	2500
$\{C, M3, E^{\nu 1}\}$	0.28	0.34	839	0.32	1880	-	-
$\{C, M3, E^{\nu 2}\}$	0.28	0.34	802	0.30	1832	-	-
$\{A, M4, E^{\nu 2}\}$	0.81	0.59	485	0.58	910	-	-
$\{C, M3, E^{HB,MJ}\}$	0.33	0.38	875	0.4	8034	-	-
$\{A, M4, E^{HB,MJ}\}$	0.89	0.78	601	0.80	1183	0.76	3259

Tablica 4.13: Dane dla różnych modeli z potencjałem $6-12_{exp}$ — oznaczenia jak w tablicy 4.10.

	1crn			1ubq		1tit	
model	$\tilde{T}_f$	$\tilde{T}_{min}$	t_{fold}/τ	$\tilde{T}_{min}$	t_{fold}/τ	$\tilde{T}_{min}$	t_{fold}/τ
$\{C, M3, E^o\}$	0.24	0.3	285	0.28	702	0.28	2171
$\{A, M3, E^o\}$	0.6	0.65	242	0.6	482	0.58	997
$\{C, M3, E^{\nu 1}\}$	0.18	0.3	239	0.26	532	-	-
$\{C, M3, E^{\nu 2}\}$	0.18	0.3	332	0.28	715	-	-
$\{A, M4, E^{\nu 2}\}$	0.62	0.60	205	0.60	450	0.59	1201
$\{C, M3, E^{HB,MJ}\}$	0.23	0.3	261	0.30	670	0.32	9980
$\{A, M4, E^{HB,MJ}\}$	0.67	0.78	244	0.76	411	0.75	1163

Tablica 4.14: Dane dla różnych modeli z potencjałem $10-12$ — oznaczenia jak w tablicy 4.10.

	1crn			1ubq		1tit	
model	$\tilde{T}_f$	$\tilde{T}_{min}$	t_{fold}/τ	$\tilde{T}_{min}$	t_{fold}/τ	$\tilde{T}_{min}$	t_{fold}/τ
$\{C, M3, E^o\}$	0.19	0.28	440	0.26	638	0.35	4300
$\{A, M3, E^o\}$	0.41	0.65	339	0.6	529	0.56	1500
$\{C, M3, E^{\nu 1}\}$	0.13	0.3	972	0.25	1100	0.29	5130
$\{C, M3, E^{\nu 2}\}$	0.12	0.3	928	0.25	1083	0.28	5280
$\{A, M4, E^{\nu 2}\}$	0.42	0.6	285	0.58	537	0.59	2320
$\{C, M3, E^{HB,MJ}\}$	0.14	0.19	2469	-	-	-	-
$\{A, M4, E^{HB,MJ}\}$	0.46	0.74	423	0.70	1914	0.71	2909

Tablica 4.15: Dane dla różnych modeli z potencjałem Morse'a — oznaczenia jak w tablicy 4.10.

model		1crn		1ubq		1tit	
	$\tilde{T}_f$	$\tilde{T}_{min}$	t_{fold}/τ	$\tilde{T}_{min}$	t_{fold}/τ	$\tilde{T}_{min}$	t_{fold}/τ
$\{C, M3, E^o\}$	0.3	0.32	358	0.32	677	0.31	3200
$\{A, M3, E^o\}$	0.72	0.65	344	0.64	614	0.60	1688

Tablica 4.16: Dane dla różnych modeli z mieszanym potencjałem $(10-12, 6-12)$ lub $(10-12, 6-12_{const})$ — oznaczenia jak w tablicy 4.10.

model		1crn		1ubq		1tit	
	$\tilde{T}_f$	$\tilde{T}_{min}$	t_{fold}/τ	$\tilde{T}_{min}$	t_{fold}/τ	$\tilde{T}_{min}$	t_{fold}/τ
$\{(10-12, 6-12), C, M3, E^{o,HB}\}$	0.19	0.27	253	0.26	485	-	-
$\{(10-12, 6-12), A, M3, E^{o,HB}\}$	0.47	0.61	217	0.62	428	0.60	1213
$\{(10-12, 6-12_{const}), C, M3, E^{o,HB}\}$	0.23	0.30	223	0.30	451	0.63	2940
$\{(10-12, 6-12_{const}), A, M3, E^{o,HB}\}$	0.58	0.61	209	0.60	397	0.60	1107

Tablica 4.17: Dane dla różnych modeli typu $C^{\alpha-\beta}$ — oznaczenia jak w tablicy 4.10.

model		1crn		1ubq		1tit	
	$\tilde{T}_f$	$\tilde{T}_{min}$	t_{fold}/τ	$\tilde{T}_{min}$	t_{fold}/τ	$\tilde{T}_{min}$	t_{fold}/τ
$\{\ 6-12, C, M2, E^o\}$	0.29	0.39	543	0.42	1280	0.40	3289
$\{\ 6-12, C, M3, E^o\}$	0.25	0.39	490	0.40	1010	0.40	2520
$\{\ 6-12, C, M3, E^{\nu2}\}$	0.31	0.38	591	0.40	1183	0.42	3629
$\{\ 10-12, C, M3, E^o\}$	0.22	0.36	641	0.38	1132	-	-

Rozdział 5

Modele Go do zwijania białek

W rozdziale poprzednim wyznaczyliśmy optymalny model Go do rozciągania białek na podstawie zgodności z danymi doświadczalnymi. Tym niemniej optymalny model Go powinien także umożliwiać zwijanie białek. W niniejszym rozdziale scharakteryzujemy własności kinetyczne modeli sprawdzając ogólną ich zdolność do zwijania. W szczególności zbadamy wpływ potencjałów do opisu lokalnej sztywności na poprawność struktury natywnej.

Wyniki dyskutowane w tym rozdziale są szerzej omówione w pracach [I] i [VIII].

5.1 Protokół symulacji

Symulację zwijania białek przeprowadzamy startując z przypadkowej konformacji rozwiniętej molekuły i następnie wyznaczamy czas zwinięcia t_{fold}, jako czas potrzebny na dotarcie białka do konformacji natywnej. Jeśli nie precyzujemy inaczej osiągnięcie stanu natywnego definiujemy poprzez utworzenie po raz pierwszy naraz wszystkich natywnych kontaktów. Dynamiczne istnienie kontaktu zdefiniowaliśmy w rozdziale 3.1.1 poprzez parametr l_e, który przyjmuje konkretną wartość w zależności od rozważanego potencjału V^{NAT}. Wartość czasu t_{fold} określmy jako wartość medialną po kilkunastu trajektoriach.

Zależność czasów zwijania od temperatury zazwyczaj przyjmuje kształt litery U, gdzie centrum U określa położenie optymalnej temperatury zwijania, $\tilde{T}_{min}$. t_c oznacza średni czas ustalenia się pojedynczego kontaktu. Proces zwijania w temperaturze $\tilde{T}$ zazwyczaj charakteryzujemy korzystając z diagramów scenariuszowych, zaproponowanych i opisanych w pracy [129], które pokazują czasy t_c w funkcji odległości wzdłuż sekwencji $|j-i|$ dla danego kontaktu.

Używając 61 modeli przeprowadziliśmy symulacje zwijania dla białek 1crn, 1ubq i 1tit, tak aby za każdym razem wyznaczyć charakterystyczny kształt U. W przypadku białek 1crn i 1ubq t_{fold} jest uzyskane z co najmniej 500 trajektorii podczas gdy dla 1tit rozważyliśmy po 200 trajektorii. Otrzymane $\tilde{T}_{min}$, t_{fold} są pokazane w tablicach 4.10, 4.11, 4.12, 4.13, 4.14, 4.15, 4.16 i 4.17, a zakres optymalnej temperatury zwijania jako błąd na wykresie 4.4.

5.2 Kinetyka zwijania w modelach typu C^{α}

W rozdziale 4 pokazaliśmy, iż optymalne temperatury zwijania $\tilde{T}_{min}$ otrzymane z symulacji zwijania białek w rozważanych modelach przyjmują trzy charakterystyczne wartości 0.3, 0.6, 07, na podstawie których wyznaczyliśmy temperatury dla symulacji rozciągania. W niniejszym rozdziale skupimy się na porównaniu własności kinetycznych dla konkretnych wariantów modeli. Zgodnie z wcześniejszą uwagą otrzymaliśmy, że tylko dla małego i prostego białka typu 1crn wszystkie modele prowadzą do zwinięcia. Rysuneku 4.2 przedstawia zależności czasów zwijania od temperatury dla 1crn w kilku przykładowych modelach o jednorodnej skali energii (lewy panel) i niejednorodnych skalach (prawy panel), różnych potencjałach dla oddziaływań natywnych V^{NAT} i dwóch sposobach opisu lokalnej sztywności łańcucha V^{C} i V^{A}. Natomiast rysunek 4.3 pokazuje te same zależności dla tytyny w modelach, w których zwinięcie było możliwe. W większości modeli czasy zwijania przyjmują charakterystyczne symetryczne kształty U, z rozbieżnością w skończonych temperaturach po obu stronach zakresu zwijania. Zakres temperatur optymalnego zwijania jest szerszy wraz z wyżej położonym $\tilde{T}_{min}$ dla modeli z potencjałem V^{A} niż V^{C}. Fakt, iż każdy model prowadzi do zwinięcia 1crn, pozwala na porównanie potencjałów. Najwęższy wykres U obserwujemy dla potencjału 10-12, szerszy dla 6-10-12, średniej wielkości dla 6-12, szeroki dla 6-12_{exp} i najszerszy dla 6-12_{const}. W wyniku zamiany białka 1crn na białko klasy β, np. 1tit, otrzymujemy podobne skalowanie U, z tym wyjątkiem, że potencjał 10-12 będzie najwęższy.

W celu znalezienia które z modeli wykazują rozsądne własności kinetyczne rozważania ograniczamy do dwóch białek, które nazywa się potocznie trudnymi zwijaczami 1ubq i 1tit. Tylko połowa spośród 61 modeli prowadzi do zwinięcia obu białek i dobrze określonej krzywej U. Do tego zbioru należą wszystkie modele z jednorodną skalą energii i mapą $M3$, z jednym wyjątkiem modelu o parametrach {10-12,C}. W tablicach z parametrami zwijania nie zawarliśmy informacji o proporcji zwiniętych do nie zwiniętych trajektorii. Dla białka 1tit ilości te są bardzo podobne. Trochę lepsza sytuacja ma miejsce po zamianie V^{C} na V^{A}. Jednakże istnieje wyjątek — model {6-12_{exp},C,M3,E^{0}} okazuje się mieć doskonałe własności kinetyczne, w tym przypadku dla tytyny zwija się aż 90% trajektorii (z małą wartością $RMSD$) i krótkimi czasami t_{fold}.

Modele zawierające skalę energii $E^{\ L1,L2}$ są interesującymi obiektami do badania własności kinetycznych, gdyż ich energia w stanie natywnym jest bardzo zbliżona do modelu z jednorodną skalą energii. Jednakże własności dynamiczne skal są różne. Skale energii $E^{L1,L2}$ powodują tłumienie sił oddziaływania pomiędzy natywnyni kontaktami wraz ze wzrostem odległości od stanu natywnego podczas dynamiki. Znajdujemy, iż modele ze skalą E^{L1} lub E^{L2} ułatwiają procesy zwijania, kontakty krótko i średnio zasięgowe tworzą się szybciej i najdłużej pozostają razem. Najlepsze rezultaty otrzymaliśmy dla E^{L2} wraz z V^{6-12}. Wynik ten pokazuje, że w przypadku prostych gruboziarnistych modeli w których nie uwzględnia się w żaden sposób ekranowania kontaktów, stopniowe tłumienie oddziaływań dalekozasięgowych jest bardzo ważne. Rozważenie tego rodzaju modeli dla dodatkowej grupy białek pokazało, że najlepiej opisują one własności kinetyczne białek klasy α.

Zwinięcie tytyny w modelach z niejednorodną skalą energii i potencjałem C jest bardzo czasochłonne i wymaga ogromnej statystyki, a otrzymywane zależności czasu zwijania od temperatury są często pojedynczymi punktami. Sytuacja ta zmienia się po zamianie V^C na V^A, co przedyskutujemy w następnym paragrafie. Tak więc modele z niejednorodna skalą energii i V^C omówimy na przykładzie ubikwityny i krambiny. Znajdujemy wyraźną korelację pomiędzy zakresem optymalnych własności kinetycznych i rodzajem skali energii. Najszerszą formę krzywej U otrzymujemy dla skali energii $E^{\nu1,\nu2}$, trochę węższą dla E^{L1},E^{L2}, średnią $(6-12, 10-12)$, $E^{o,HB}$ i E^{o}, i najwęższą dla $E^{HB,MJ}$. Nieznacznie inne proporcje daje tylko potencjał 10-12. Powyższe zależności otrzymaliśmy dla dwóch białek, tak więc nie jest to wynik determinujący ogólną zależność czasów zwijania od temperatury dla tego typu modeli, chociaż w dobrze określanym modelu położenie $\tilde{T}_{min}$ powinno być podobne.

Przegląd własności kinetycznych tytyny w modelach o niejednorodnych skalach energii z V^A pokazuje dla większości skal energii charakterystyczną dla procesu zwijania krzywą U. Modele w których zwijanie jest najszybsze lub innymi słowy najłatwiejsze, powstają z połączenia parametrów wymienionych w ramce zgodnie danymi zawartymi w tablicach 4.10 i 4.12.

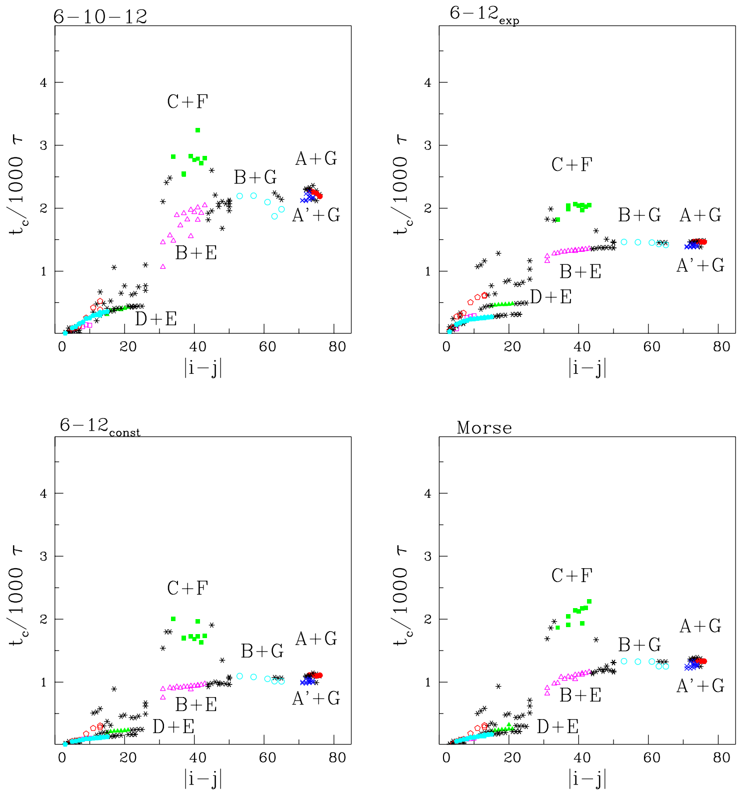
$$\{\ 6-12, A, M4, E^{T2}\ \}$$
$$\{\ 6-12_{exp}, A, M4, E^{HB,MJ}\ \}$$
$$\{\ 6-12_{exp}, A, M4, E^{\nu2}\ \}$$

Drugim dość dobrym zestawem modeli, które przede wszystkim dają optymalne warunki do zwijania białek helikalnych jest potencjał V^{6-12}_{const} niemalże dla każdej skali energii. Dodatkowe testy pokazały, że nie dla każdego białka helikalnego osłabienie energii oddziaływań dalekozasięgowych wygładza proces zwijania, jak np. 1rpo.

Otrzymaliśmy, iż najtrudniej jest znaleźć drogę do stanu natywnego dla modeli ze skalami energii $E^{HB,MJ}$ i $E^{\nu 1}$ wraz z V^C dla prawie wszystkich rodzajów potencjałów V^{NAT}.

5.2.1 Scenariusze zwijania

Powyżej pokazaliśmy ogólne własności kinetyczne nie badając poprawności procesu zwijania z danymi doświadczalnymi, który jest najważniejszym kryterium oceny wiarygodności modelu. Jako białko testowe ponownie wybieramy tytynę, której procesy zwijania były obszernie badane doświadczalnie [161, 162].



Rysunek 5.1: Diagram scenariuszowy ustalenia się kontaktów w białku 1tit podczas zwijania w modelu $C, M3, E^o$ dla potencjałów 6-10-12, 6-12$_{exp}$, 6-12$_{const}$, Morse'a.

Rysunek 5.1 przedstawia scenariusze zwijanie dla czterech potencjałów 6-10-12, 6-12$_{exp}$, 6-12$_{const}$ i Morse'a z jednorodną skalą energii. Natomiast rysunek 5.2 pokazuje scenariusz zwijania w optymalnym modelu Go do rozwijania. Początkowy proces zwijania w każdym z tych modeli jest podobny podczas gdy największa różnica występuje pomiędzy czasami ustalenia się kontaktów C+F w stosunku do A+G. Żaden z procesów nie jest idealnie zgodny z doświadczeniem, jednakże sytuacja ta zmienia się gdy zauważymy, że wypadkowe czasy ustalania się kontaktów są złożeniem kilku różnych trajektorii zwijania. W przypadku najlepszego optymalnego modelu Go do rozwijania otrzymujemy, że dla 95% wszystkich trajektorii początkowa kolejność ustalania się kontaktów jest taka sama: F+G → A'+B → D+E → A+B. Zasadnicze różnice w scenariuszu występują przy ustalaniu się kolejnych elementów struktury, tak że:

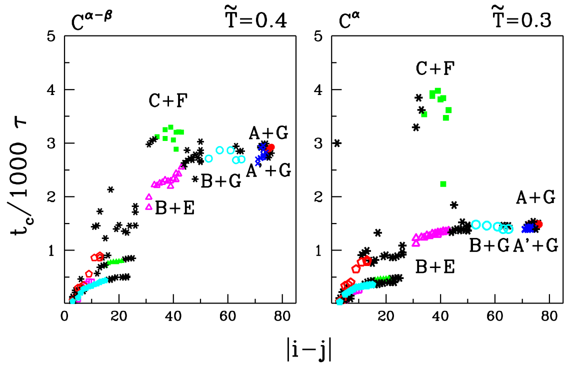
1. 50 trajektorii posiada scenariusz:
   B+E → B+G → $A' + G$ → $A + G$ → C+F

2. 36 trajektorii posiada scenariusz:
   B+E → B+G ≈ $A' + G$ ≈ $A + G$ ≈ C+F

3. 6 trajektorii posiada scenariusz:
   B+E → C+F → B+G ≈ $A' + G$ ≈ $A + G$

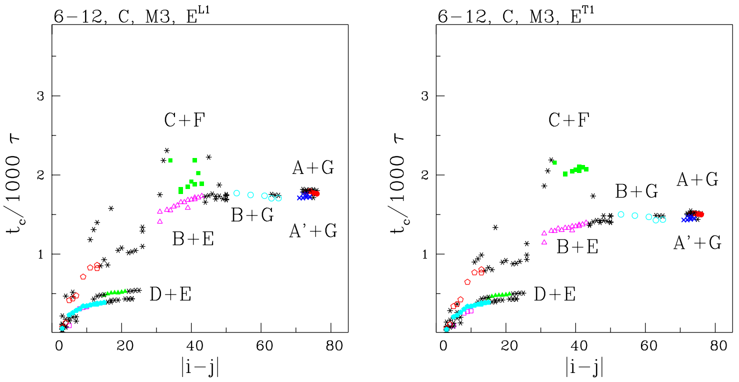
4. 4 trajektorie posiadają scenariusz:
   B+E → → B+G → $A' + G$ → $A + G$ ≈ C+F

5. 4 trajektorie posiadają scenariusz:
   → B+G → $A' + G$ → $A + G$ ≈ C+F → B+E



Rysunek 5.2: Diagram scenariuszowy ustalenia się kontaktów w białku 1tit podczas zwijania w modelu o parametrach $\{6-12, C, M3, E^o\}$ typu $C^{\alpha-\beta}$(lewy panel) i C^{α} (prawy panel).

Scenariusz rozwijania z rysunku 5.2 jest więc charakterystyczny dla 50% trajektorii, należy także podkreślić, iż dla tych przypadków różnica w czasie ustalenia się kontaktów z grupy C+F a A+G jest około 1500τ. Jednakże dla 36% trajektorii obserwujemy bardzo dobrą zgodność z doświadczeniem. Należy się spodziewać, iż także dla pozostałych modeli rozkład średnich czasów zwijania wyłoniłby drogi, które byłyby bardziej zgodne z doświadczeniem. To zagadnienie jest nadal tematem otwartym do dalszych badań.



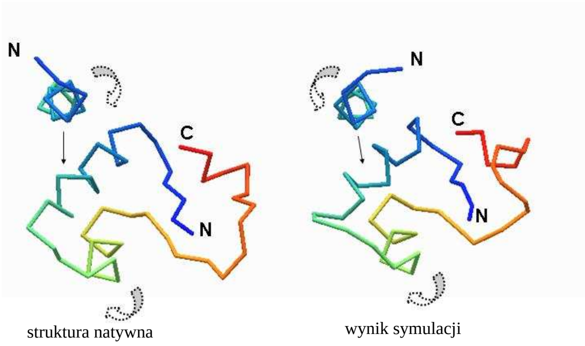
Rysunek 5.3: Diagram scenariuszowy ustalenia się kontaktów w białku 1tit podczas zwijania w modelu $6-12, C, M3$ i skali energii E^{L1}(lewy panel) i E^{T1}(prawy panel).

Rysunek 5.3 przedstawia scenariusze zwijania dla potencjału 6-12 i dwóch skal energii E^{T1}, E^{L1}, gdzie wyraźnie widzimy, iż największy wpływ na proces zwijania ma siła oddziaływania pomiędzy kontaktami natywnymi.

Chiralność — geometria łańcucha głównego

Przedyskutowaliśmy już wpływ wyboru skali energii i formy potencjału dla oddziaływań natywnych. Jednakże w rozważanych modelach istnieje jeszcze jeden ważny element dla procesu zwijania — potencjał określający lokalną sztywność łańcucha. Pokazaliśmy, iż potencjał V^A zazwyczaj daje szerszy zakres optymalnych temperatur i przesuwa $\tilde{T}_{min}$ w stronę wyższych temperatur w porównaniu do V^C. Obecnie odpowiemy na pytanie jak potencjały wpływają na przebieg zwijania i konformację stanu natywnego.

Rysunek 5.4 pokazuje otrzymaną z dynamiki molekularnej modelem Go natywną konformację krambiny w modelu bez więzów na lokalną sztywność łańcucha np. V^C, V^A, gdy kryterium zwinięcia jest powstanie wszystkich natywnych kontaktów. Otrzymana struktura jest błędna gdyż zawiera lewoskrętną helisę, mimo że kryterium ustalenia kontaktów zostało zrealizowane a $RMSD$ jest mniejsze niż 2.5 Å. W pracy [I] pokazaliśmy, na podstawie białek klasy α, $\alpha-\beta$ i β, iż taka sytuacja jest częsta a wyjściem z niej jest przyjęcie bardziej krytycznej definicji stanu natywnego lub właśnie zastosowanie jednej z form potencjałów określających lokalną sztywność. Aby

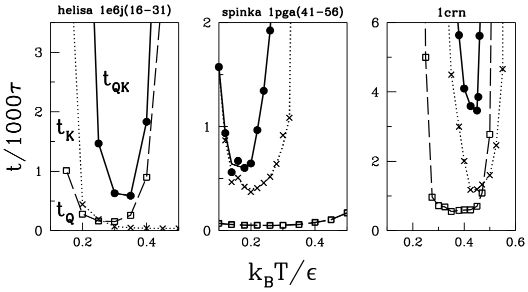


Rysunek 5.4: Reprezentacja patyczkowa konformacji krambiny. Lewy rysunek: konformacja krambiny w stanie natywnym danym na podstawie pliku PDB. Prawy rysunek: konformacja krambiny otrzymana z symulacji w modelu Go o parametrach {6-12,M2,E^o}. Mimo, że kryterium kontaktów zostało ustalone krambina posiada jedną lewoskrętną helisę.

rozstrzygnąć które podejście jest bardziej wydajne, zaproponowaliśmy dwa dodatkowe kryteria definicji stanu natywnego i parametr mierzący poprawność lokalnej chiralności łańcucha. Używane do tej pory kryterium utworzenia się wszystkich natywnych kontaktów oznaczamy symbolem Q. Ponadto, niech R oznacza kryterium, które określa wejście konformacji białka w ustalony zakres $RMSD$, względem oryginalnego stanu natywnego. Dla krambiny przyjęliśmy R$\leq$2 Å. Drugim kryterium oznaczonym symbolem A jest ustalenie się wielkości kątów płaskich θ_i i dwuściennych ϕ_i w pewnym zakresie $\pm\Delta\theta$, $\pm\Delta\phi$ od ich natywnej wartości. Przyjmujemy, $\pm\Delta\theta$, $\pm\Delta\phi$ w granicy 60^O, nie jest to bardzo mocne ograniczenie jednakże i tak czas jego realizacji jest najdłuższy w porównaniu z kryterium Q i R. Dodatkowo wprowadzamy parametr K, który umożliwia dynamiczne określenie poprawności geometrii łańcucha głównego poprzez pomiar odchylenia od stanu natywnego, określonego wartością chiralności

$$K = \sum_{I=2}^{N-2} H\Big(\frac{C_i}{C_i^{NAT}} - 0.5\Big), \tag{5.1}$$

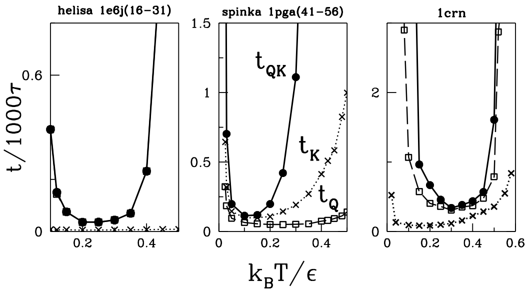
gdzie $H(X)$ jest funkcją Heaviside'a, taką że $\Theta(x)$=1 gdy x$\geq$0, a w przeciwnym wypadku 0. W pracy prześledziliśmy zachowania się parametru K w zależności od czasu i otrzymaliśmy, że zmienia on swoją wartość bardzo dynamicznie, tak więc zażądanie aby był on spełniony naraz dla każdego węgla C_i^α nie jest realnym kryterium zakończenia symulacji. W niniejszej pracy przyjmujemy $K_c \leq 0.75 \cdot K_{NAT}$, gdzie K_{NAT} określa liczbę natywnych kontaktów w białku. Na podstawie powyższych parametrów określiliśmy kryterium stanu natywnego, które jest także czułe na topologię konformacji białka, jest ono złożeniem jednego z kryteriów Q, R, A i spełnia warunek poprawnej chiralności K_c.



Rysunek 5.5: Średnie czasy zwijania dla helisy, spinki i krambiny. Linia przerywana odpowiada czasowi potrzebnemu na spełnienie kryterium kontaktowego Q, linia ciągła — kryterium kontaktowego wraz z kryterium chiralnościowym QK, linia kropkowana — ustalenie się dla 75% aminokwasów chiralności blisko stanu natywnego.

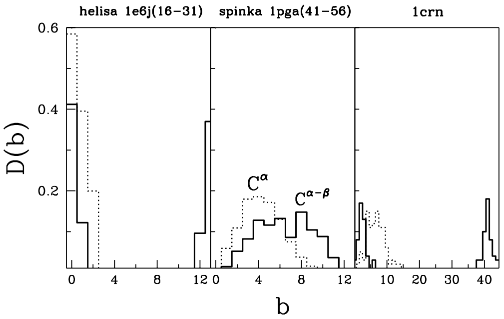
Rysunki 5.5 i 5.6 przedstawiją porównanie czasów zwijania i U-kształtnych zależności od temperatury dla kryteriów Q, QK i K, odpowiednio bez i z potencjałem chiralnościowym z κ=1 danym wzorem (3.12). W pierwszym przypadku różnica w czasach i w kształcie U jest bardzo duża pomiędzy ustaleniem się poprawnej topologicznie struktury (K) wraz z kryterium Q a tylko kryterium Q. Ponadto otrzymaliśmy, iż np. użycie kryterium R wraz RK, lub A wraz AK daje czasy zwijania o rząd wielkości większe. Spełnienie kryterium AK jest zwłaszcza trudne w obszarach nieustrukturyzowanych, ponadto zastosowanie kryterium R ($RMSD$) wymaga skalowania wraz z rozmiarem badanego białka. Diametralna zmiana sposobów zawijania następuje po uwzględnieniu potencjału chiralnościwego. Warunek poprawnej chiralności K_c i ustalenia się kontaktów natywnych Q jest niemal równoważny w $\tilde{T}_{min}$. Otrzymaliśmy także, iż bardzo podobnie do potencjału V^C na kinetykę zwijania działa potencjał V^A. Jednakże jego faktyczny czas obliczeniowy jest dużo dłuższy.

Alternatywnym sposobem zapewnienia poprawnej chiralności może być użycie zaproponowanego w rozdziale 3.2 uogólnionego modelu Go, w którym grupa boczna jest reprezentowana przez węgiel C^{β}. Porównaliśmy dwa modele typu C^{α} z potencjałem chiralnościowym i model uogólniony $C^{\alpha-\beta}$ bez uwzględnienia sił od lokalnej sztywności.



Rysunek 5.6: Czas zwijania dla helisy, spinki i krambiny w modelu z V^C, oznaczenia takie same jak na rysunku 5.5.

Rysunek 5.7 pokazuje, że przedstawienie grupy bocznej tylko jednym atomem typu C^β nie jest wystarczające aby zapewnić poprawną chiralność i usprawnić proces zwijania białek.



Rysunek 5.7: Prawdopodobieństwo otrzymania złego znaku chiralności $D(b)$ dla danego aminokwasu b wzdłuż sekwencji w helisie, spince i krambinie dla modelu typu C^α z potencjałem chiralnościowym (oznaczonego kropkowaną linią) i modelu typu $C^{\alpha-\beta}$ bez V^C (oznaczonego ciągłą linią).

Otrzymane wyniki pokazują, iż potencjały do opisu lokalnej sztywności są niezbędnym elementem modeli Go podczas symulacji zwijania. Usprawniają one proces zwijania zapobiegając frustracji łańcucha, a przede wszystkim powstaniu niepoprawnej geometrii stanu natywnego. Należy pokreślić, że potencjały V^C i V^A podobnie kształtują lokalną geometrię, chociaż V^A daje nieznacznie lepsze wyniki przy jednoczesnym prawie czterokrotnie dłuższym czasie obliczeniowym.

5.3 Kinetyka zwijania w modelach typu $C^{\alpha-\beta}$

Rysunek przedstawia zależność czasów zwijania t_{fold} od temperatury, otrzymane w 4 formach uogólnionych modeli typu $C^{\alpha-\beta}$. Wprowadzenie dodatkowego stopnia swobody powoduje ograniczenie zakresu optymalnej temperatury zwijania i przesunięcie $\tilde{T}_{min}$ w stronę wyższych temperatur. Ważnym wynikiem jest otrzymanie zbliżonych wartości $\tilde{T}_{min}$ do $\tilde{T}_f$ dla różnych białek w ramach tych samych parametrów modelu.

Scenariusze zdarzeń

Przy omawianiu diagramów scenariuszowych dla zwijania ograniczymy się do porównania najlepszego modelu do rozwijania z modelem o takich samych parametrach ale typu $C^{\alpha-\beta}$. Kolejność ustalania się kontaktów dla obu modeli w ich $\tilde{T}_{min}$ jest pokazana na rysunku 5.2.

W przypadku modelu $C^{\alpha-\beta}, C, M3, E^o$ około 95 procent trajektorii charakteryzuje się takim samym początkiem scenariusza zdarzeń jaki został podany dla modelu $\{C^\alpha, C, M3, E^o\}$. Ustalanie się kontaktów średnio i długo zasięgowych posiada natomiast 5 różnych ścieżek:

1. 52 trajektorie ze scenariuszem:
   B+E $\rightarrow$ B+G $\approx$ C+F $\approx A' + G \approx A + G$

2. 32 trajektorie ze scenariuszem:
   B+E $\rightarrow$ B+G $\approx A' + G \approx A + G \rightarrow$ C+F

3. 10 trajektorie ze scenariuszem:
   C+F $\rightarrow$ B+E $\rightarrow$ B+G $\approx A' + G \approx A + G$

4. 4 trajektorie z scenariuszem:
   B+E $\approx$ B+G $\approx$ C+F $\approx A' + G \approx A + G$

5. 2 trajektorie ze scenariuszem:
   B+G $\rightarrow A' + G \rightarrow A + G \approx$ C+F $\rightarrow$ B+E

W przypadku tego modelu tylko 32 trajektorie są typowe dla średniego scenariusza, ponadto tylko 5 procent z nich charakteryzuje się przerwą większą niż 1500 τ pomiędzy kontaktami $A + G$ a $C + F$.

Podsumowanie

Podsumowując, otrzymaliśmy iż optymalny model Go do rozwijania także może być użyty do opisu zwijania białek. Rozkład scenariusza tworzenia się kontaktów na poszczególne ścieżki pokazuje, że istnieje wśród nich zespół trajektorii bardzo podobny do procesu zwijania obserwowanego doświadczalnie, przynajmniej dla badanej przez nas tytyny.

Wydaje się, że istnieje możliwość doboru parametrów tak, aby scenariusze zwijania zgadzały się z doświadczeniem, jednakże nie było to naszym celem w tym rozdziale.

Rozdział 6

Własności mechaniczne białek — przegląd PDB

W rozdziale tym przedstawimy wyniki przeglądu własności mechanicznych białek dokonanego na podstawie bazy danych struktur białkowych *PDB*. Przegląd wykonamy przy użyciu optymalnej formy modelu Go do rozciągania białek $\{6-12, C, M3, E^o\}$. Odpowiemy na pytanie które białka są silne i dlaczego. Scharakteryzowanie własności mechanicznych dostępnych struktur białkowych jest bardzo ważne dla ukierunkowania dalszych badań doświadczalnych, a w konsekwencji zrozumienia i poznania procesów mechanicznych zachodzących w komórkach.

W pierwszej części niniejszego rozdziału określimy parametry białek i technikę przeglądu bazy danych *PDB*. W wyniku tego otrzymamy zbiór 7510 białek, które są nie dłuższe niż 150 aminokwasów. Dodatkowo tworzymy zbiór 239 niehomologicznych białek większych niż 150 aminokwasów. Dla zbiorów tych przeprowadzamy symulacje rozciągania, z których wyznaczamy maksymalny opór F_{max} biomolekuł. Następnie dokonujemy charakterystyki własności mechanicznych 3813 białek w zależności od klasyfiakcji $CATH$. Znajdujemy zbiór 134 najsilniejszych białek (1.8% wszystkich zbadanych białek) i na jego podstawie podajemy definicję sześciu motywów w białkach odpowiedzialnych za duży mechaniczny opór. Wyniki nasze uzupełniamy o własności mechaniczne białek zawierających mostki siarczkowe. Wyznaczony ranking 134 najsilniejszych białek testujemy ponadto uogólnionym modelem Go, $C^{\alpha-\beta}$.

Wyniki zaprezentowane w tym rozdziale jak i ich szczegółowa analiza zawarte są w pracach [IV] i [VI].

6.1 Protokół przeglądu *PDB*

Wybór białek

W dniu 26 lipca 2005 roku w bazie danych *PDB* znajdowało się 29 385 struktur, które stanowią punkt wyjścia do przeglądu własności mechanicznych białek. Ze zbioru tego odrzuciliśmy struktury niebiałkowe, oraz odpowiadające białkom które wchodzą w skład kompleksów (np. z DNA lub RNA), są integralną częścią innych struktur (np. wirusów), nie posiadają jednolitych łańcuchów (np. nie została wyznaczona doświadczalnie pełna struktura lub zawierają inne jony bądź kofaktory). Z pozostałych ok. 15000 białek wyznaczyliśmy białka o długości łańcucha w zakresie od 40 do 150 aminokwasów. Dolna granica dla długości białka wynika z możliwości eksperymentalnych (pomiar mniejszych białek jest obarczony bardzo dużym błędem), a w górnej granicy zawierają się białka jednodomenowe. Należy zauważyć, iż przedział ten zawiera najbardziej typową długość domeny białkowej N=120. (Rozkład wartości N jest pokazany na rysunku 2 w pracy [IV].) Po zastosowaniu powyższych kryteriów otrzymaliśmy zbiór 7510 białek, który będziemy oznaczać symbolem S7510.

Ze zbioru S7510 tylko 3810 białek ma automatycznie określoną strukturę według klasyfikacji *CATH*. Zbiór tych białek będziemy oznaczać S3810 i do niego ograniczymy się badając korelację siły ze strukturą.

Powyższy zbiór nie uwzględnia białek o długości większej niż 150 aminokwasów, jak też struktur wielodomenowych. Aby uwzględnić te cechy białek przeanalizowaliśmy dodatkowy testowy zbiór o symbolu S239. Składa się on z 239 białek wyselekcjonowanych z 3 podzbiorów. Tak więc zbiór S239 zawiera białka niehomologiczne, kompletne (m. in. nie posiadające dziur) wybrane z dwóch podzbiorów z pracy [163] gdzie największe białko to 1no3 z N=851, natomiast ilość domen w jednym białku może wynosić od jednej do ośmiu. Ponadto S239 zawiera 11 kompletnych struktur białek ze zbioru wyselekcjonowanego przez autorów *PDB*, składającego się 19 najbardziej typowych, znaczących „foldów" białek. Dodatkowo do tego zbioru zostało dołączone białko 1cpy, złożone z 421 aminokwasów.

Protokół symulacji

Technika rozciągania białek, a w szczególności wybór parametrów modelu takich jak prędkość ciągnięcia, stałe sprężystości oraz temperatura, zostały opisane w podrozdziale 4.1.2. W modelu tym rozciągnięcie 7510 białek przy średnim czasie około 3 godziny (serwer komputerowy z procesorem 2.8 GHz) na białko wynosiłoby 22530 godzin czyli 2.6 roku. Z tego powodu zastosowaliśmy dodatkowe rozciąganie z prędkością dziesięciokrotnie szybszą v_p=0.05 σ/τ od tej optymalnej dla modelu (dyskutowanej w podrozdziale 4.1.2). Zgodnie z wynikami doświadczalnymi [96] i teoretycznymi [129] wielkość F_{max} zależy zaledwie logarytmicznie od prędkości i jest mniejsza dla mniejszej prędkości. Natomiast kształt krzywej $F(d)$ jest bardziej czuły na wybór prędkości, ale zachowuje takie samo największe maksimum dla największego oporu. Te własności pozwalają zastosować technikę szybkiego rozciągania do wyznaczenia położenia najwyższego maksimum. Tak więc białka ze zbioru S7510 zostały najpierw rozciągnięte z dużą prędkością, z której wyznaczyliśmy położenie najwyższego maksimum. Następnie symulację powtórzyliśmy, jednakże tylko do momentu wyznaczenia F_{max}, z prędkością v_p=0.005 σ/τ, która jest bardziej zbliżona do wartości doświadczalnej.

Interpretacja danych

Własności mechaniczne białek uzyskanych z symulacji scharakteryzujemy na podstawie maksymalnego oporu F_{max}, klasyfikacji $CATH$ opisanej w rozdziale 1, wzoru dla krzywych $F(d)$ i względnego położenia λ opisanego poniżej. W przypadku najsilniejszych białek S134 do identyfikacji pochodzenia bardzo dużego oporu zostaną użyte diagramy scenariuszowe zdarzeń.

Wykresy krzywych $F(d)$

Wykresy siły F w funkcji długości d rozciągniętego białka różnią się nie tylko położeniem F_{max} ale także ilością mniejszych maksimów przed i po osiągnięciu najwyższego maksimum. W przypadku białek 1tit i 1ubq najwyższe maksimum pojawia się na wykresie $F(d)$ jako pierwsze, a po nim – w zależności od białka – następuje jedno dodatkowe maksimum lub seria kilku mniejszych maksimów. Odwrotna sytuacja występuje w białku integrynie, gdzie największe maksimum jest poprzedzone serią małych maksimów. Należy tu jednak pokreślić, iż liczba maksimów po największym maksimum może być różna i zależna od trajektorii (możliwych dróg rozwijania), natomiast liczba maksimów przed jest dobrze określona. Na podstawie ilości i położenia małych maksimów względem głównego maksimum, kształt $F(d)$ może być zaklasyfikowany jako M, BnM, BnMA, lub MA, gdzie M (największe maksimum) odpowiada przebiegowi $F(d)$ tylko z jednym największym maksimum, oznaczenie BnM określa przebieg z liczbą n małych maksimów przed głównym maksimum, sytuacja BnMA jest podobna do BnM jednak dodatkowo po głównym maksimum występują jedno lub seria kilku małych maksimów, i wreszcie symbol MA oznacza brak maksimów przed głównym maksimum a serię maksimów po nim. Zgodnie z tą klasyfikacją białka 1tit, 1ubq i 1ido są typu odpowiednio MA, MA i B3MA.

Względne położenie λ głównego maksimum na krzywej $F(d)$ w relacji do odległości między końcami białka jest także dobrze określoną własnością i może być użyte do określenia charakteru mechanicznego oporu. Zamiast d wyliczamy jednak chwilową odległość między końcami białka, L, i definujemy paramter λ jako

$$\lambda = \frac{L_m - L_n}{L_f - L_n}, \tag{6.1}$$

gdzie L_m jest wartością L odpowiadającą położeniu najwyższego maksimum, L_n jest odległością między końcami białka w stanie natywnym, a L_f odpowiada całkowitemu rozciągnięciu $(N-1)3.8$ Å. Tak więc λ określa względne położenie F_{max} w funkcji $L_f - L_n$ wyrażone w procentach.

Diagram scenariuszowy

Mechanizmy rozwijania jak i identyfikacja mechanicznych imadeł będzie wyjaśniona poprzez użycie diagramów scenariuszowych, które pokazują przy jakim przesunięciu d_u dany kontakt jest zerwany po raz ostatni. Przykład interpretacji danych na podstawie diagramu jest podany w podrozdziale 4.1.3.

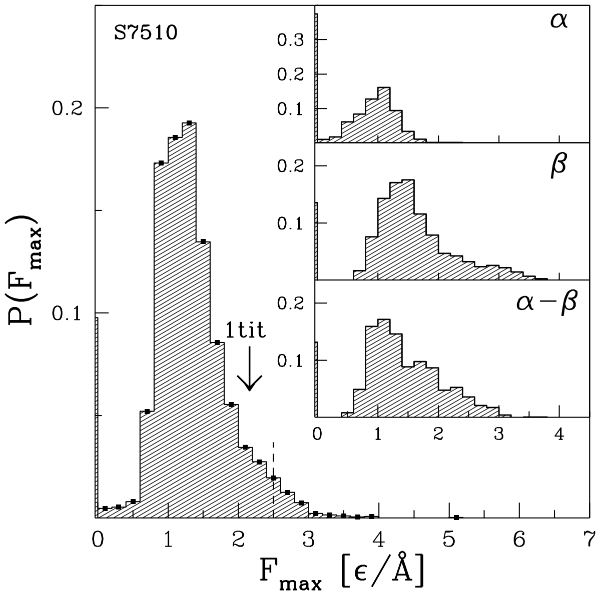
W optymalnym modelu do rozwijania $\{6-12, C, M3, E^o\}$ dany kontakt uważamy za zerwany gdy r_{ij}>$1.5\sigma_{ij}$, gdzie r_{ij} oznacza odległość między atomami C_i^{α}. Zgrupowanie zerwanych kontaktów dla tej samej wielkości d_u odpowiada maksimom siły. Mechaniczne imadło można zidentyfikować poprzez wyznaczenie odległości dla której występuje największy opór, a następnie sukcesywną eliminację występujących tam kontaktów odpowiedzialnych za powstanie tego oporu. Eliminacja najmniejszej grupy kontaktów dającej największą redukcję siły oznacza, że elementy struktury pomiędzy którymi występują owe kontakty tworzą mechaniczne imadło.

6.2 Wyniki

6.2.1 Wyniki dla białek z $40 \leq N \leq 150$: zbiór S7510

Rysunek 6.1 przedstawia rozkład wartości F_{max} dla zbioru S7510 białek otrzymany podczas symulacji rozciągania modelem Go {6-12,C,M3,E^o}. Wartości siły są w zakresie od 0ϵ/Å do 5.1ϵ/Å i najbardziej prawdopodobne są rzędu 1.5 ϵ/Å. Typowy błąd wynikający ze zmienności pomiędzy poszczególnymi trajektoriami jest rzędu 0.1-0.15 ϵ/Å. Jedno z najsilniejszych znanych do tej pory białek 1tit o sile $\tilde{F}_{max}$=2.1 klasyfikuje się obecnie także w gronie białek silnych, jednakże jest ono o ponad połowę słabsze niż najsilniejsze znalezione białko. Długi płaski ogon w histogramie sił od wartości $\tilde{F}_{max}$=2.5 określa zbiór 134 najsilniejszych białek, które stanowią 1.8% wszystkich rozważanych białek i są hierarchicznie wymienione w tabeli A.1. Wraz z wartością siły

F_{max} tabela podaje także kod strukturalny PDB, wielkość N, względne położenie F_{max} wyrażone poprzez λ oraz symbol strukturalnej klasyfikacji $CATH$, jeśli jest dostępny. Szczegółowa analiza najsilniejszych białek jest zawarta w sekcji 6.2.5 i pracy [IV].



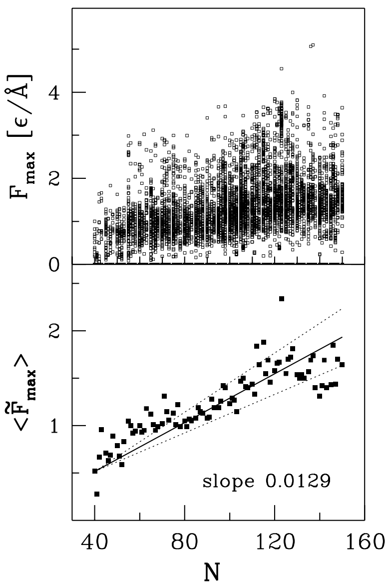
Rysunek 6.1: Rozkład prawdopodobieństwa F_{max} dla białek ze zbioru S7510. Kwadraty dla dużych wartości F_{max} odpowiadają pojedynczym danym. Strzałka pokazuje wartość F_{max} dla 27 domeny tytyny. Linia przerywana zaznacza próg, powyżej którego zaczyna się zbiór najsilniejszych białek. Panel wewnętrzny przedstawia rozkład F_{max} dla klas strukturalnych α, β, α/β. Czarne wąskie kolumny w zerach odpowiadają liczbie białek, w których naprężenie od mostków siarczkowych pojawiało się wcześniej niż F_{max}.

Górny panel rysunku 6.2 pokazuje wartość F_{max} w zależności od długości białka N, gdzie wydawałoby się, iż dla danego N siła może przyjmować dowolną wartość. Jednakże przedstawienie wartości siły jako średniej dla danej długości białka pokazuje rosnący trend dla F_{max} w funkcji N, przedstawiony na dolnym panelu rysunku 6.2. Tak więc czym większe N, tym większe prawdopodobieństwo wygenerowania większej siły oporu. Obserwacja ta jest zgodna z wynikami symulacji dla 239 białek dużych białek (N pomiędzy 153-851) omówionych poniżej.

6.2.2 Wyniki dla białek z $150 \leq N \leq 851$: zbiór S239

Wyznaczenie wartości F_{max} dla wolnej prędkości w przypadku białek z N>500 jest bardzo kosztowne numerycznie nawet gdy wiadomo, w którym miejscu wystąpi maksimum. Najprostszym sposobem ominięcia tego ograniczenia numerycznego jest wyznaczenie korelacji między wartościami sił otrzymanymi z szybkiego i wolnego ciągnięcia. Panel dolny na rysunku 6.3 przedstawia zależność między średnimi wartościami sił dla wolnego i szybkiego ciągnięcia $F_{max,slow} = 0.86 \cdot F_{max,fast} - 0.13\epsilon/\text{Å}$ na podstawie S7510 i 148 białek z S239. Tak więc w przypadku 91 białek z N>500 na podstawie powyższej zależności wyliczamy F_{max} dla wolnego ciągnięcia.

Histogram rozkładu wartości F_{max} jest pokazany na rysunku 6.3, z którego widać iż

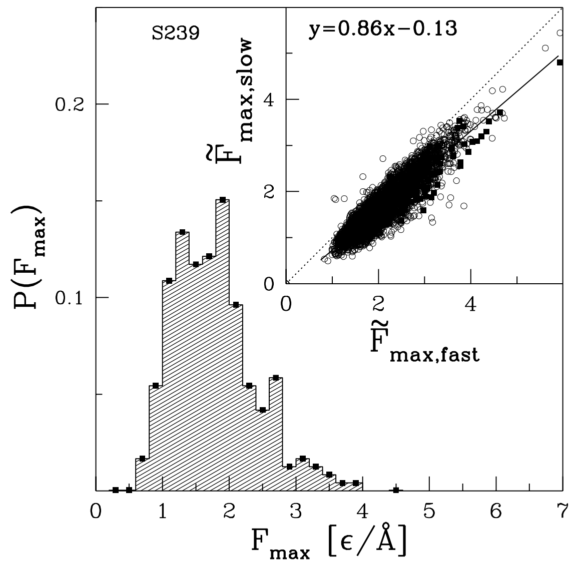


Rysunek 6.2: Górny panel: wartość F_{max} w zależności od długości N białka. Dolny panel: średnia wartość F_{max} w zależności od N.

jest on wyraźnie szerszy niż rozkład pokazany na rysunku 6.1 dla białek z $40 \leq N \leq 150$. Obserwacja ta potwierdza wcześniejszy wniosek, iż wartość mechanicznego oporu rośnie wraz ze wzrostem N. Powyższe wyniki symulacji zostały otrzymane dla modelu $\{6-12, C, M2, E^o\}$. Użycie bardziej optymalnej formy modelu z mapą kontaków typu $M3$ po pierwsze nie zmienia zależności między siłą dla wolnej i szybkiej prędkości, po drugie w mniejszym stopniu obniża siłę dla białek długich. W konsekwencji rozkład z rysunku 6.3 w modelu z mapą typu $M3$ byłby przesunięty o ok. 0.2 ϵ/Å w kierunku niższych sił, co nie wpływa na przeprowadzoną charakterystykę własności mechanicznych długich białek i wybór najsilniejszego (największa siła należy do białka krótkiego).

Tabela A.3 przedstawia 40 białek (co stanowi 17% całego zbioru S239), które według naszych obliczeń powinny być najsilniejsze, tzn. o sile F_{max}>2.9 ϵ/Å. Białka w tabeli są ułożone od największej do najmniejszej wartości siły F_{max}, wraz z którą podana jest liczba domen w białku n_D i symbol klasyfikacji $CATH$ dla dwóch pierwszych domen n_1,n_2. Stosunek białek najsilniejszych do pozostałych (wyrażony w procentach) jest dużo większy w gronie białek długich (9.5%) niż w krótkich (2%), co wskazuje ponownie na korelację między siłą a wielkością białka N. Kolejną charakterystyczną cechą tego zbioru jest niemalże takie same prawdopodobieństwo pojawienia się F_{max} na początku jak i na końcu procesu rozwijania, określone wartością parametru λ, rysunek 6.4. W przypadku najsilniejszych białek z N≤150 zbiór S134, rozkład λ posiada wyraźne maksimum dla małych wartości.

Największy mechaniczny opór wykazuje białko jednodomenowe o średniej wielkości N oznaczone symbolem 1bg2, którego struktura i wykres $F(d)$ są pokazane na rysun-



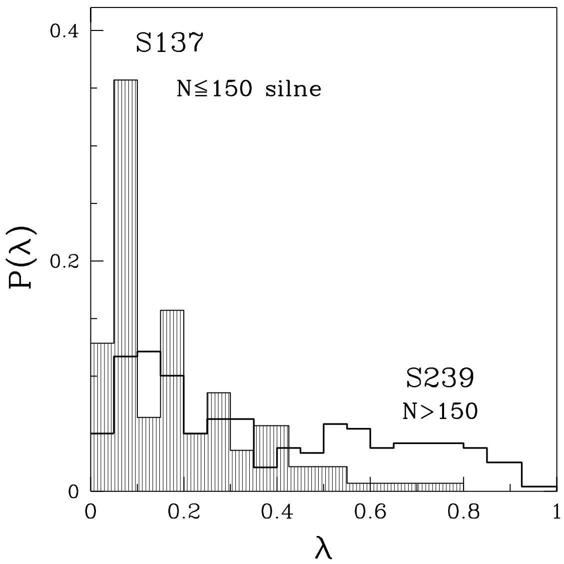
Rysunek 6.3: Rozkład prawdopodobieństwa F_{max} dla S239. Wewnętrzny panel pokazuje zależność między siłami otrzymanymi z szybkiego i wolnego ciągnięcia.

kach 6.5 i 6.6. Kolejne drugie (2sli), trzecie (1tmo) i czwarte (1bfd) najsilniejsze białka są już wielodomenowe. Składają się one z trzech, czterech i trzech domen tworzących czwartorzędową strukturę pokazaną na rysunku 6.5. Cechą łączącą te cztery najsilniejsze białka jest struktura typu α/β, oprócz jednej domeny klasy α dla białka 1tom. Klasyfikacja CATH dla pierwszych dwóch domen dla każdego białka jest zawarta w tabeli A.3. Ważną obserwacją jest fakt, iż wśród najsilniejszych długich białek są jeszcze tylko cztery białka z domeną klasy α. Duży mechaniczny opór dla żadnego z tych białek jednakże nie pochodzi od domeny klasy α.

Bardzo silny opór w białku 1bg2 wynika ze ścinania się dwóch nici β dodatkowo stabilizowanych przez otaczające nici β. Jest to rodzaj krótkiego mechanicznego imadła, ale dobrze umiejscowionego w białku. Analiza pochodzenia mechanicznego oporu dla pozostałych trzech najsilniejszych białek jest mocno utrudniona ze względu na złożoność gęsto upakowanych domen lub nawet przeplatanie się ich, tak jak w przypadku białka 1tmo. Wykonanie ponownej symulacji rozciągania ale z wyeliminowanymi kontaktami pomiędzy domenami pokazało znaczną redukcję wartości maksymalnej siły. Tak więc i w tych białkach występuje ścinanie się nici β, ale pochodzących z przylegających różnych domen.

6.2.3 Zależność siły od klasy i architektury: S3813

Obecnie powracamy do analizy białek z N<150, zawężając zbiór S7510 do zbioru S3813 złożonego z 3813 białek o znanej klasyfikacji $CATH$. Ideę hierarchicznego podziału białek na klasę (C), architekturę (A), topologię (T) i homologię (H) opisaliśmy

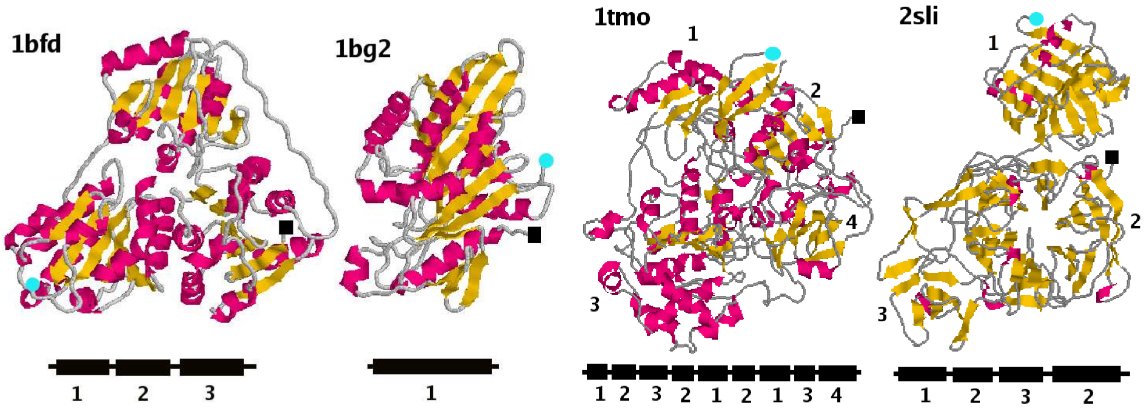


Rysunek 6.4: Rozkład prawdopodobieństwa wystąpienia parametru λ określającego stosunek położenia F_{max} podczas rozciągania względem całkowitego rozwinięcia ze stanu natywnego. Zaciemniony i pusty histogram reprezentują odpowiednio najsilniejsze białka z $N \leq 150$ i wszystkie długie białka S239.

w podrozdziale 1.1. Przykład podziału białek ze zbioru S3813 według *CATH* jest podany w dodatku A w tabeli A.1. Rozkład F_{max} dla S8313 jest bardzo podobny do S7510 pokazanego na rysunku 6.1. Istnieją jednak znaczące różnice w rozkładach odpowiadających poszczególnym klasom, widoczne np. na wewnętrznych panelach rysunku 6.1. Histogramy dla białek z klasy β i α/β charakteryzują się rozkładem asymetrycznym z długim ogonem w obszarze dużych sił. Tak nie jest w przypadku białek klasy α, których siły są dane rozkładem Gaussa z maksimum w 1.1 ϵ/Å, co wskazuje iż białka tej klasy charakteryzują się słabym mechanicznym oporem. Wynik ten jest zgodny z doświadczalnymi pomiarami dla pojedyńczych białek klasy α: kalmoduliny, spektryny czy pojedynczej domeny ankiryny, wymienionych w tabeli 1.1.

Interesującym wynikiem jest fakt, iż przewidywania wskazują na kilka białek z domenami klasy α (1ile, 2ng1, 1ciy i 1cii) ale o dużych wartościach N (pomiędzy 293-821), które mogą wykazywać duży opór na rozciąganie z uwagi na ruchy ścinające pomiędzy domenami. Białka nieustrukturyzowane (dane nie pokazane) zazwyczaj posiadają najmniejszą siłę, co także jest zgodne z danymi doświadczalnymi np. dla PEVK (domena łącznikowa w tytynie) zawartymi w tabeli 1.1. Szczegółowa analiza zależności siły na poziomie klasy białek jest przedstawiona w pracy IV.

Obecnie przechodzimy do dokładniejszej strukuralnej analizy (na poziomie architektury) otrzymanego zbioru 3810 białek. 80% białek klasy α w S3810 ma architekturę typu wiązki ortogonalnej. W klasie β 36% to beczułki, 31% kanapki, 13% wstążki, i 13% rolki. W klasie α/β 40% to rolki, a 39% to kanapka dwuwarstwowa. Podział ten ulega radykalnej zmianie w zbiorze najsilniejszych białek.



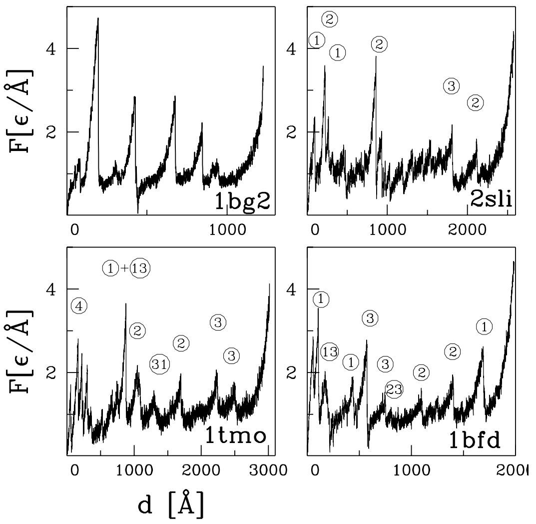
Rysunek 6.5: Reprezentacja wstążkowa czterech najsilniejszych długich białek. Czarne prostokąty pokazują liczbę i kolejność ułożenia domen.

Najbardziej interesującą architekturą do analizy korelacji siły ze strukturą białka jest kanpka β i kanapka dwuwarstwowa α/β. Histogram F_{max} dla grupy takich białek jest pokazany na rysunku 6.7. W obu architekturach rozkład jest szeroko rozciągnięty i w pierwszym przypadku posiada 3 szerokie lokalne maksima, w drugim dwa bardziej zwarte maksima. Kolejny podział na topologię i homologię umożliwia identyfikację maksimów w rozkładach sił. I tak dla kanapki β maksima dla większych sił pochodzą od topologii immunoglobuliny, na które składają się głównie na poziomie homologii, białka transportowe i immunoglobuliny. Maksimum dla małych sił jest reprezentowane przez rolki Jelly (z ang. Jelly roll). W wypadku architektury typu kanapki dwuwarstwowej α/β maksimum przy wyższych siłach tworzą białka o topologii P-30. Jednakże pokrywają one także pozostały zakres sił wraz z białkami ubikwityno-podobnymi. Maksimum dla słabszych sił jest tworzone przez pozostałe topologie wymienione na rysunku 6.7.

Powyższe zależności pokazują, iż klasyfikacja CATH może być użyta do charakterystyki własności mechanicznych białek.

Białka o tej samej klasyfikacji $CATH$ i różnym mechanicznym oporze

Choć mechaniczny opór jest skorelowany ze strukturą białka według klasyfikacji $CATH$, to jednak zadany kod $CATH$ struktury może odpowiadać różnym zachowaniom dynamicznym. W zbiorze S3810 istnieją białka o tym samym kodzie CATH, a o skrajnie innym zachowaniu dynamicznym. Przykładem takich białek są silne białka o kodzie 1p7e i 1pga (z pozycji 42 i 47 w tabeli A.1 z rozdziału A), które posiadają ten sam kod klasyfikacji $CATH$ co dużo słabsze (nawet czterokrotnie) białka 1mpe i



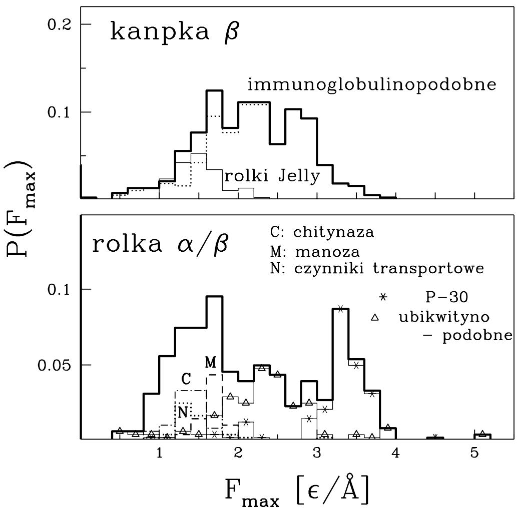
Rysunek 6.6: Wykres siły w funkcji przemieszczenia dla czterech najsilniejszych długich białek. Pojedyncze cyfry w kółkach oznaczają numer domeny, z których kontakty są zerwane w danym miejscu. Podwójne cyfry odpowiadają zrywaniu kontaktów między domenami. Oznaczenia domen są takie same jak na rysunku 6.5.

1q10. Ten kod to 3.10.20.10 i opisuje on białka klasy α/β, architektury rolki, topologii — ubikwityno-podobne i homologii — wiążące immunoglobulinę, a ich struktura wraz z wykresem siły jest pokazana na rysunku 6.8. Te cztery białka są tej samej długości N=56, przy czym słabe różnią się od silnych mutacją tylko w 3 miejscach. Mutacja ta eliminuje kontakty dalekozasięgowe, a w konsekwencji zmienia strukturę przestrzenną białka na taką która wręcz ułatwia rozciągnięcie białek.

6.2.4 Wpływ temperatury

Powołując się na dyskusję z podrozdziału 4.2 należy podkreślić, że temperatura pełni ważną rolę podczas rozciągania, gdyż termiczne fluktuacje pomagają rozwijać białko. W większości przypadków tak długo jak poszczególne maksima są wyraźne kształt krzywych $F(d)$ w różnych temperaturach jest taki sam, a jedynie wielkość maksimów skaluje się. Różnice mogą też być widoczne we fluktuacjach.

W zbiorze S3810 zostały znalezione białka dla których temperatura istotnie wpływa na kształt wzoru $F(d)$. Rysunek 6.9 pokazuje krzywe $F(d)$ dla dwóch białek 1nme i 1amx, dla temperatury pokojowej $\tilde{T}$=0.3 oraz gdy termiczne fluktuacje nie mają miejsca, $\tilde{T} = 0$. Naturalnym zachowaniem się maksimów na krzywej $F(d)$ przy obniżaniu temperatury jest wzrost ich amplitudy i pojawienie się jej później. Otrzymane wyniki przedstawiają niespotykany do tej pory wpływ temperatury na wykres $F(d)$. Mak-



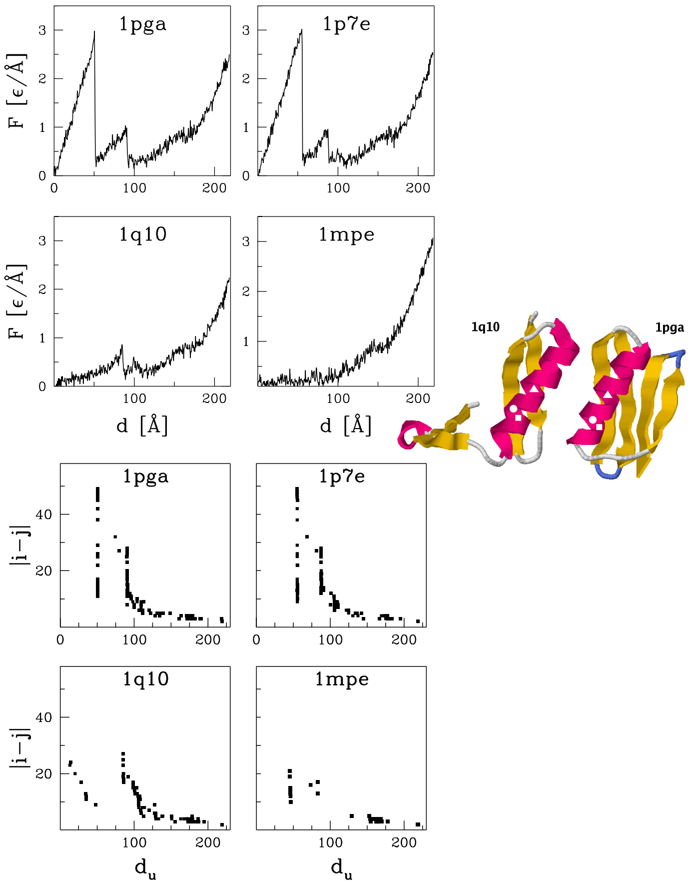
Rysunek 6.7: Górny panel: rozkład siły dla architektury typu kanapki β (ciągła gruba linia), i dalej wewnątrz niej podział jej na dwie topologie oznaczone cienką i przerywaną linią. Dolny panel: rozkład siły dla architektury typu α/β rolki (ciągła gruba linia) i wewnątrz niej podział na topologie oznaczone pozostałymi liniami.

simum siły pojawia się w innym miejscu oraz zmienia się liczba średniej wielkości maksimów.

6.2.5 Najsilniejsze białka — zbiór S134

Zbiór 134 najsilniejszych białek z $F_{max} >$ 2.5 ϵ/Å dla 40$\leq N \leq$150 jest zawarty w tabeli A.1. Korzystając z nachylenia wyliczonego dla głównego trendu między siłą wyznaczoną eksperymentalnie a teoretycznymi wynikami z optymalnego modelu pokazanymi na rysunku 4.8 otrzymujemy, że wartość progowa odpowiada $\simeq$ 180pN. Zauważmy, iż w zbiorze S134 nie ma tytyny z F_{max}=2.1ϵ/Å, chociaż odpowiadająca jej wartość doświadczalna to 210pN; jednakże leży ona poniżej głównego trendu i przelicznik dla niej jest zawyżony.

Histogram wartości λ w tym zbiorze ma ostre maksimum dla wartości około 10%, rysunek 6.4, podczas gdy w zbiorze S7510 rozkład jest raczej płaski (dane nie pokazane). Wynik ten wskazuje na istnienie dużych sił związanych ze zrywaniem kontaktów w pobliżu terminalnych odcinków białka, tak jak to miało miejsce dla tytyny omówionej w rozdziale 1.2.3. Na podstawie analizy kształtu $F(d)$ stwierdzilismy, iż dla 72% silnych białek najpierw występuje duże maksimum, po którym ukazuje się słabsze maksimum (wzór typu MA). Dla 7% słabsze maksimum występuje przed głównym maksimum(BnM), dla 19% jest kilka maksimów przed maksimum głównym i przynaj-

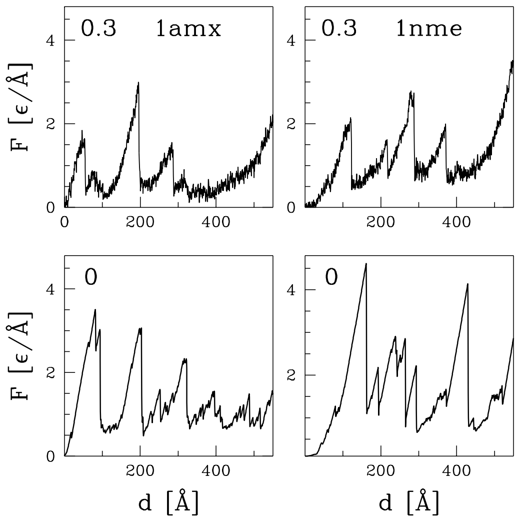


Rysunek 6.8: Prawy rysunek pokazuje wstążkową reprezentację białek 1q10 i 1pga, gdzie białe symbole oznaczają położenie trzech aminokwasów, którymi różnią się te białka. Górne 4 panele przedstawiają wykresy sił i pierwsze dwa odpowiadają silnym, następne dwa słabym mechanicznie białkom. Odpowiednio 4 dolne panele pokazują scenariusze rozwijania najpierw dla silnych, następnie dla słabych mechanicznie białek. Wszystkie białka mają tą samą klasyfikację $CATH$.

mniej jedno po nim ($BnMA$), a tylko cztery białka, w tym dwa najsilniejsze, mają tylko jedno maksimum siły (M). Wykresy $F(d)$ o wzorach typu MA i M, reprezentowane przez cztery najsilniejsze białka, są przedstawione na rysunku 6.10.

Rozkład sił w zależności od klasyfikacji $CATH$ dla S134 różni się zasadniczo od ogólnego zbioru S7510, jak i też od rozkładu dla najsilniejszych długich białek S239. Zbiór S134 nie zawiera białek klasy α. Otrzymaliśmy, iż białka te należą tylko do 6 typów architektur (A) i dwie z nich są szczególnie dobrze reprezentowane: kanapki β (60%) i rolki α/β (30%). Najsilniejsze białka należą także do dziewięciu klas topologicznych. Najczęstsze topologie to immunoglobuliny (2.60.40 według klasyfikacji $CATH$) i ubikwitynopodobne rolki UB (3.10.20). Pozostałe występujące tu kody to 3.90.79, 3.40.50, 3.10.130, 3.10.13330, 3.10.50, 2.40.40 i 2.40.240. Kolejny rozkład ze względu na pełnioną funkcję wskazuje także tylko na kilka rodzin: białka układu immunologicznego, wiążące, sygnałowe, transportowe i immunoglobuliny.

Niektóre białka wymienione w dodatku A w tabeli A.1 mogą być reprezentowane przez kilka kodów PDB, które odpowiadają bliskim lub dalszym homologom tego sa-

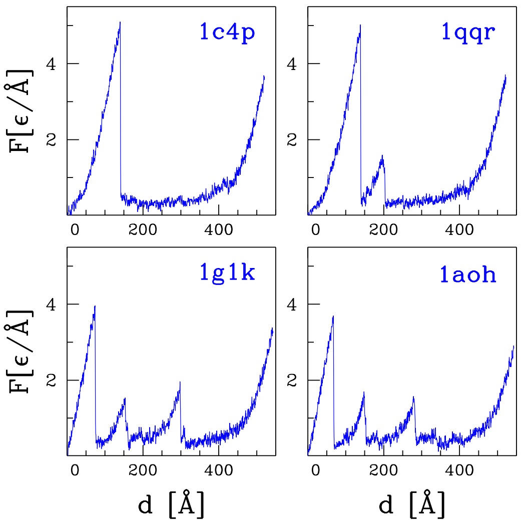


Rysunek 6.9: Wykres siły w funckji przemieszczenia dla białek 1amx (lewe panele) i 1nme (prawe panle) dla temperatury $\tilde{T}$=0.3 i $\tilde{T}$=0 (odpowiednio górne i dolne panele).

mego białka, jak i też temu samemu białku ale pochodzącemu z różnych warunków krystalizacyjnych. Dynamika rozciągania jest wrażliwa na szczegóły strukturalne, i nawet identyczne sekwencyjnie białka różniące się mapą kontaktów mogą wykazywać inne siły. Z tego powodu uwzględniliśmy różne kody dla danego białka, a następnie przy użyciu oprogramowania FATCAT [164] sklasyfikowaliśmy relację między białkami. Otrzymaliśmy, że 41 białek ze zbioru S134 nie jest homologicznie spokrewnionych, podczas gdy pozostałe 93 białka należą do 32 różnych grup, które składają się przynajmniej z 2 białek. Tabela A.4 przedstawia podział białek na grupy. W szczególności dwie najsilniejsze struktury 1c4p i 1qqr odpowiadają temu samemu białku — β-domenie streptokinazy (o topologii zwoju UB) która uczestniczy w mechanizmie krzepnięcia krwi, ale skrystalizowanych w różnych warunkach. Trzecie skrystalizowane białko jest w tej samej grupie co 20. 1anu — jest to domena kohezyny celulozy.

6.2.6 Mechanizm rozrywania silnych krótkich białek

Mechanizmy rozwijania wyjaśniamy przy użyciu diagramów scenariuszowych. Elementy struktury które tworzą mechaniczne imadła będziemy identyfikować poprzez grupę kontaktów, których eliminacja w największym stopniu redukuje opór białka. Okazuje się, iż w zbiorze 134 najsilniejszych białek 90% mechanicznych imadeł jest tworzonych przez dwie długie równoległe nici β połączone wiązaniami wodorowymi, które są zrywane poprzez ruchy ścinające. Ruchy ścinające są widoczne poprzez specyficzne ułożenie chwilowej konformacji białka podczas dynamiki molekularnej. Mo-

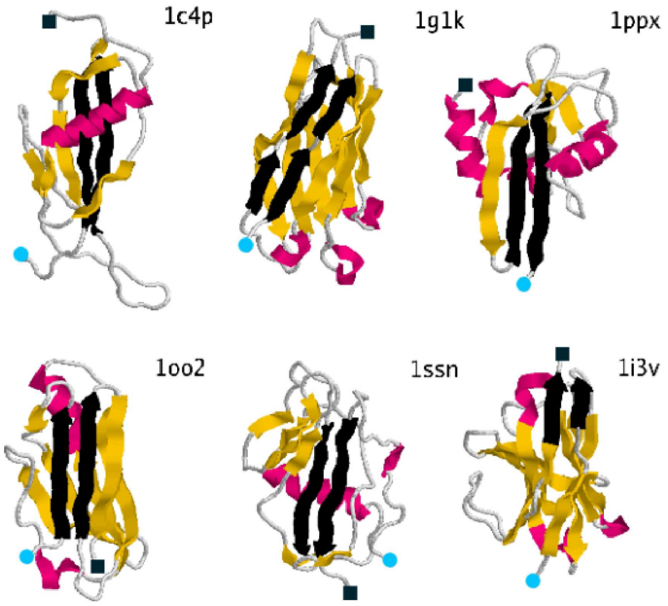


Rysunek 6.10: Wykres siły w funkcji przemieszczenia dla czterech najsilniejszych białek.

tyw ścinający został już wcześniej zidentyfikowany [8, 119, 99] w kilku białkach jako odpowiedzialny za duża siłę, jednakże obecne badania są przeprowadzane dla dużo większego zbioru białek i pozwalają na identyfikację dodatkowych 5 możliwych jego form, które są odpowiedzialne za jeszcze większy mechaniczny opór białek podczas rozciągania. Białka w których najwyraźniej jest widoczna różnica między sześcioma zidentyfikowanymi typami imadła są pokazane na rysunku 6.11.

Mechaniczne imadło, typ I — białko 1c4p

Diagram scenariuszowy najsilniejszego białka wraz z wykresem siły jest przedstawiony na rysunku 6.12. Białko to składa się z czterech łańcuchów o podobnych własnościach i rysunek 6.12 dotyczy pierwszego z nich. Łańcuch ten jest zbudowany z helisy α (aminokwasy między 196-210), oznaczonej symbolem I i ośmiu nici β oznaczonych literami od A do H, poczynając od końca N. Z rysunku 6.12 widać, iż największa siła pochodzi z jednoczesnego zerwania kontaktów pomiędzy nićmi β, terminalnymi A+G, A+I, G+I, jak i też D+H, A+B, E+G, D+G, C+G i A+C. Następnie zostają rozerwane $C + F$ i $E + F$, a jako ostatnia rozwija się helisa α. Panel wewnętrzny na rysunku 6.12 pokazuje, iż usunięcie kontaktów pomiędzy A i G redukuje maksimum siły o około 50%, podczas gdy usunięcie kontaktów między pozostałymi parami ma znacznie mniejszy wpływ. Grupy usuwanych kontaktów i otrzymane wartości sił są zestawione w tabeli 6.1. Na tej podstawie stwierdziliśmy, że rdzeniem mechanicznego imadła są kontakty wodorowe między nićmi A i G, nici te są zaznaczone czarnym kolorem na rysunku 6.11. Kontakty A+I, G+D i D+H dodatkowo tworzą sieć połączeń

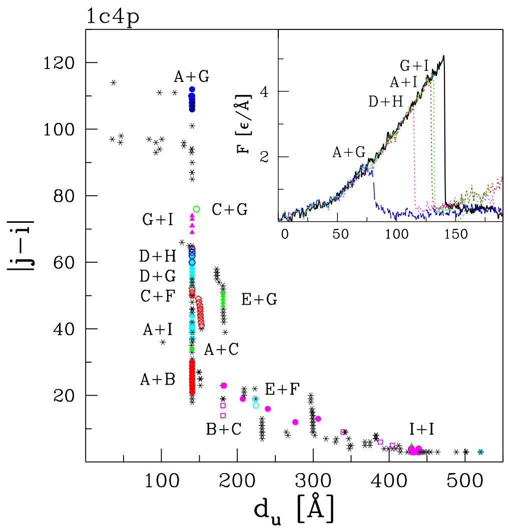


Rysunek 6.11: Wstążkowa reprezentacja struktury białek zawierających 6 typów mechanicznych imadeł. Elementy struktury zaznaczone na czarno reprezentują 6 typów najbardziej stabilnych mechanicznie imadeł. Dla białka 1g1k imadło składa się z dwóch segmentów.

Tablica 6.1: Wartość opru mechanicznego F_{max} po eliminacji danej grupy kontaktów.

PDB	K_{nat}	$F_{max}[\epsilon/\text{Å}]$	K_r	$F^r_{max}[\epsilon/\text{Å}]$	aa-aa
1c4p	366	5.1	27	2.1	(158-168, 266-278)
1g1k	441	3.9	16	2.7	(4-8, 131-135); (10-13, 131-141)
1aoh	456	3.8	21	2.2	(5-10, 137-140); (12-15,142-146)
1c76	327	3.8	23	1.4	(24-32, 124-135)
2sak	334	3.7	30	1.9	(24-32, 109-119)
1ssn	382	3.5	22	2.1	(24-32, 124-132)
1ppx	343	3.5	21	1.2	(2-12, 78-86)
1ie5	281	3.4	10	2.4	(10-16, 35-40)
1oo2	316	3.3	17	1.5	(12-18, 105-112)
1i9e	321	3.2	15	2.3	(9-13, 105-110)
1amx	456	3.1	24	1.6	(19-26, 105-115)
1sp0	371	3.1	19	1.8	(32-41, 138-144)
1sn0	323	3.1	21	1.4	(12-18, 105-112)
1i3v	373	2.9	7	2.6	(11-13, 123-128)
1yn4	295	2.9	29	1.7	(44-53, 132-140)
1qp1	315	2.7	13	2.1	(9-13, 102-116)

stabilizujących rdzeń. Konstrukcja ta daje motyw bardzo stabilnego mechanicznego imadła.



Rysunek 6.12: Scenariusz rozrywania kontaktów dla najsilniejszego białka 1c4p. Wewnętrzny panel: wykres siły w funkcji przemieszczenia z symulacji rozciągnia 1c4p dla pełnej mapy kontatków (czarna linia) oraz po usunięciu kontaktów tworzących mechaniczne imadło (granatowa przerywana linia).

Droga mechanicznego rozwijania pokazana na rysunku 6.12 jest typowym sposobem rozwijania białka 1c4p. Jednakże w przypadku tego białka istnieje też druga ścieżka rozwijania, która prowadzi do mniejszej maksymalnej siły F_{max}=4.8 ϵ/Å i towarzyszy jej tylko rozrywanie kontaktów między A+G, A+I, A+C i A+B. Zauważmy, iż na wykresie $F(d)$ pokazuje się drugie bardzo słabe maksimum, tworzone przez C+G, G+I, D+H, D+G i E+C (dane zawarte w pracy [IV]). Rozważyliśmy także inne sposoby ciągnięcia 1c4p, np. za aminokwasy: 1-108, 1-63, 1-89, 22-136, 63-136, które charakteryzują się największym oporem 2.5, 2.9, 1.7, 1.9 i 1.7 ϵ/Å. Wynik ten pokazuje, że rozciąganiu za terminale towarzyszy zerwanie najbardziej stabilnego mechanicznie imadła.

Mechaniczne imadło, typ II — białko 1g1k

Białko 1g1k jest zbudowane z jedenastu nici β, które tworzą cztery kartki β a, b, c i d. Bardzo podobną strukturę posiadają białka 1aoh i 1anu. Mechaniczne imadło tych białek składa się z dwóch szeregów połączonych segmentów, równoległych nici β stabilizowanych przez sieć kontaktów wodorowych z pozostałymi nićmi β z tych samych kartek β. Wykres siły w funkcji przemieszczenia na rysunku 6.13 pokazuje, że białko 1g1k rozwija się w trzech etapach. Z porównania diagramu scenariuszowego i chwilowych konformacji z dynamiki molekularnej otrzymujemy, iż największa siła pochodzi z ruchów ścinających między A+J i B+K stabilizowanych przez I+A, I+D, I+K, I+J i H+K. Kolejne maksimum na krzywej $F(d)$ jest tworzone poprzez zerwanie kontaktów wewnątrz kartek β C+H, C+A. Następne pochodzi z deformacji pomiędzy kartkami β i rozprostowania spinki β F+G.

Identyfikacja mechanicznego imadła została potwierdzona poprzez proces usuwania kontaktów, którego wyniki są przedstawione w tabeli 6.1. Usunięcie 6 kontaktów między A+J i 3 między B+K redukuje maksymalną siłę o ok. 40%. Wykres siły dla białka bez imadła jest zobrazowany granatową przerywaną krzywą na wewnętrznym panelu rysunku 6.13.

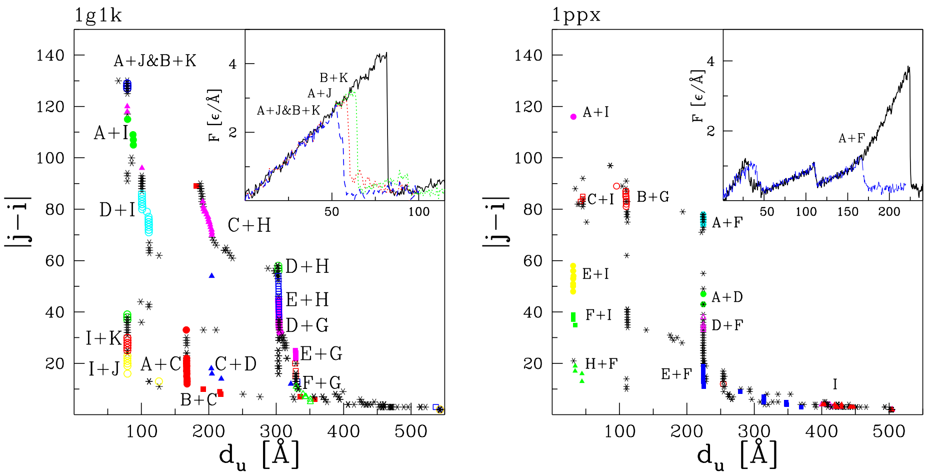
Białko 1g1k, podobnie jak 1c4p, posiada co najmniej dwie drogi rozwijania. Jednakże w przeciwieństwie do 1c4p, drogi mechanicznego rozwijania dla 1g1k różnią się dopiero po przejściu przez maksimum siły. Różnica jest widoczna w dwukrotnym wzroście drugiego maksimum, które jest konsekwencją rozerwania kartki β w jednym ruchu.

W przypadku tego białka znaleźliśmy miejsca przyłożenia siły, aminokwas 1 i 104, które prowadzą do jeszcze większego oporu mechanicznego niż ciągnięcie za aminokwasy terminalne 1 i 143. Ponadto otrzymana siła F_{max}=5.5 ϵ/Å jest jeszcze większa niż dla najsilniejszego białka 1c4p.

Mechaniczne imadło, typ III — białko 1ppx

Białko 1ppx składa się z pięciu nici β tworzących jedną kartkę β, dwóch helis α i trzech helis 3-10. Posiada kod *CATH* 3.90.79.10, który jest ten sam także dla białek 1jrk, 1k26, 1pun i 1pus.

Identyfikacja pochodzenia dużego oporu dla tego białka byłaby niemożliwa bez diagramu scenariuszowego i eliminacji kontaktów. Wewnętrzny panel z rysunku 6.13 pokazuje, iż największe maksimum jest poprzedzone dwoma mniejszymi. Pierwsze maksimum powstało z zerwania kontaktów z helisą I. Drugie maksimum wynika głównie z rozerwania wstęg B+G. W wyniku rozerwania tych kontaktów, konformacje chwilowe uzyskane z dynamiki molekularnej pokazują, iż pozostałe spinki β przyjmują pozycję równoległą do kierunku ciągnięcia, tworząc idealne mechaniczne imadło. Tak więc największy opór białka pochodzi z ruchów ścinających między A+F, w wyniku których następuje gładkie otworzenie (z ang. unzipping) spinki β E+F i rozwinięcie całego białka. Identyfikacja mechanicznego imadła została potwierdzona redukcją siły o ponad 60%, po eliminacji 21 kontaktów między A+F. Wyniki redukcji kontaktów są



Rysunek 6.13: Scenariusz rozwijania i wykresy siły po eliminacji kontaktów (wewnętrzny panel) dla białka 1g1k i 1ppx.

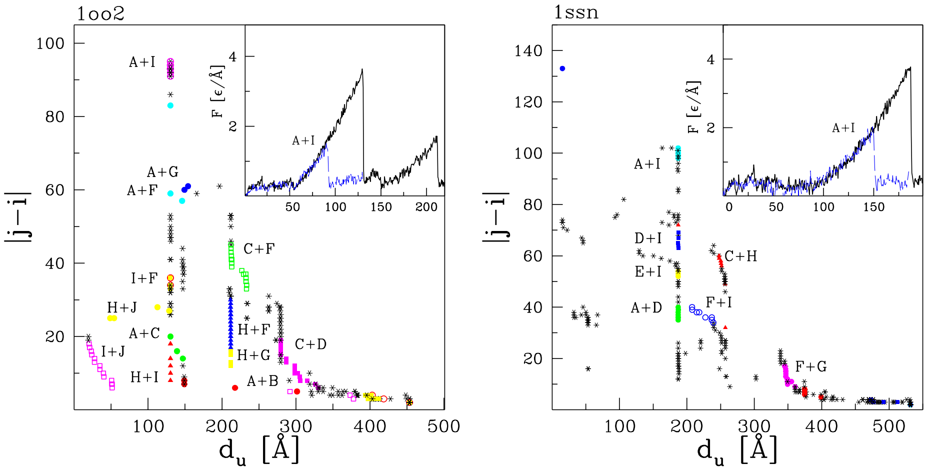
zawarte w tabeli 6.1.

Mechaniczne imadło, typ IV — białko 1oo2

Identyfikacja mechanicznego imadła w białku 1oo2 została przeprowadzona w taki sam sposób jak dla białka 1ppx, poprzez technikę eliminacji kontaktów 6.11. 1oo2 jest zbudowane z pojedynczej hesliy α, helisy 3-10 oraz dwóch kartek β: kartki a (nici A,B,I,J) i kartki b (nici C,D,F,H). W tym przypadku terminale znajdują się po tej samej stronie białka, tak więc kierunek wektora rozciągania nie jest równoległy do białka wzdłuż spinek β, lecz prostopadle do nich. Ponadto terminale nici A oraz I rozdzielone są jeszcze jedną nicią J. Z rysunku 6.14 wnioskujemy, iż pierwszym krokiem podczas rozwijania jest sukcesywne rozsuwanie spinki β I i J, które powoduje rotację białka tak, że nici A+I stają się równoległe do kierunku rozciągania. W następnym ruchu pękają kontakty pomiędzy pięcioma elementami strukturalnymi, jednakże eliminacja 30 kontaktów pomiędzy A+I redukuje wartość mechanicznego oporu o połowę (wewnętrzny panel na rysunku 6.14).

Mechaniczne imadło, typ V — białko 1ssn

Białko 1ssn składa się z 8 nici β tworzących 3 kartki β, oraz dwóch helis: α i 3-10. W tym przypadku wektor kierunku rozciągania łączący aminokwasy terminalne jest bardzo krótki i jest skierowany prostopadle w stosunku do najdłuższej kartki β. W jej skład wchodzi nić β z końca aminowego, która tworzy kontakty z nicią A. W tym przypadku obrót białka do pozycji, w której opisany powyżej motyw będzie równole-



Rysunek 6.14: Scenariusz rozwijania i wykresy siły po eliminacji kontaktów (wewnętrzny panel) dla białek 1oo2 i 1ssn.

gły do kierunku ciągnięcia, jest możliwy ze względu na istnienie wcześniej w strukturze łatwej do rozerwania pętli i kontaktów z terminalem N. Zerwanie tego odcinka umożliwia obrót o 90 stopni. Mechaniczne imadło dla białka 1ssn jest tworzone przez ruchy ścinające między A+I i oddziaływania z E+I. Eliminacja kontaktów A+I powoduje redukcję siły o $\simeq$ 50%, wewnętrzny panel na rysunku 6.14.

Mechaniczne imadło, typ VI — białko 1i3v

Białko 1i3v jest strukturalnie bardzo podobne do I27 domeny tytyny 1tit, tak więc posiada ten sam indeks $CATH$ 2.60.40.10. W tym przypadku wektor kierunku rozciągania jest równoległy do nici B i K. Mechaniczne imadło jest konsekwencją ruchów ścinających między B i K. Znacznie większy mechaniczny opór dla tego białka w porównaniu do 1tit wynika także z dodatkowej sieci kontaktów wodorowych rdzenia imadła z antyrównoległymi przylegającymi nićmi β. Ponadto w tym przypadku rdzeń jest tworzony przez osiem wiązań wodorowych, a nie sześć (białko tytna uważane do tej pory za najsilniejsze posiada w mechanicznym imadle sześć wiązań wodorowych). Redukcja tych 6 wiązań oraz 11 kontaktów stabilizujących redukuje siłę o połowę (dane przedstawione są w tabeli 6.1).

Reasumując, mechanicznie imadło może przyjąć jedną z możliwych form:

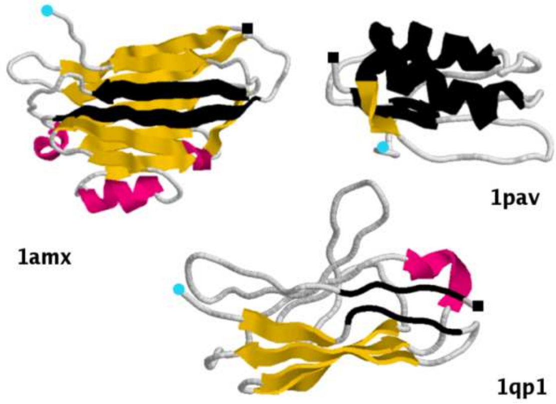
- równolegle ułożone do kierunku ciągnięcia terminalne nici β, połączone siecią wiązań (najlepiej wodorowych) przez otaczające naprzemianległe elementy struktury (najlepiej nici β),

- dwie równoległe nici β połączone wiązaniami wodorowymi, z których choć jedna jest terminalną; wyróżniamy tutaj dwa przypadki:
  - obie nici są ułożone równolegle do kierunku rozciągania,
  - nici są ułożone pod kątem do kierunku rozciągania (np. terminale są po tej samej stronie białka, więc wiązania wodorowe między wstęgami nie są prostopadłe do kierunku ciągnięcia); w tym przypadku siła będzie zależeć od możliwości obrotu białka do pozycji, w której nie naruszone nici β będą równoległe do kierunku ciągnięcia.

Siła tych mechanicznych imadeł po pierwsze zależy od liczby kontaktów wodorowych pomiędzy nićmi β, a po drugie od stabilizacji imadła przez otoczenie, co sugeruje, iż typ z terminalnymi nićmi β powinien wykazywać większy opór, gdyż białko nie traci od razu kontaktów z otoczeniem.

6.2.7 Nowe rodzaje mechanicznych imadeł

W zbiorze S134 znaleźliśmy także nowe rodzaje imadeł. Zaznaczone na czarno elementy struktury w białkach 1amx, 1pav i 1qp1, przedstawione na rysunku 6.15, pokazują przykłady tych nowych imadeł.



Rysunek 6.15: Reprezentacja wstążkowa białek 1amx, 1qp1, i 1pav o nietypowych mechanicznych imadłach zaznaczonych na czarno.

Mechaniczne imadło z antyrównoległymi nićmi β

Białko 1amx jak też wyżej notowane białko 1ai5 i niżej notowane 1lme składają się między innymi z dwóch kartek β ułożonych na przeciwko siebie. Analizę pochodzenia mechanicznego imadła zawęziliśmy do białka 1amx, gdyż było one już wcześniej przez nas dyskutowane i charakteryzuje się najbardziej ukształtowanym imadłem. Struktura 1amx powoduje przynależność terminalnych nici β do oddzielnych kartek β, a ponadto w całej strukturze nie ma ani jednej pary równoległych nici β. Eliminując grupy kontaktów otrzymaliśmy, iż największy wpływ na mechaniczny opór ma ścinanie dwóch antyrównoległych nici β z B+I. Wykluczenie tych oddziaływań redukuje siłę o $\simeq$50%. Zauważmy, iż dodatkowo kontakty pomiędzy końcem aminowym i najbliższą pętlą, pomiędzy B i C (170-174, 195-205), stabilizują rdzeń imadła w $\simeq$ 30%.

W przypadku białka 1ie5 usunięcie kontaktów pomiędzy antyrównoległymi nićmi β redukuje siłę o $\simeq$30%, co jest znacznie mniej niż dla białka 1amx. Jednakże eliminacja oddziaływań przyciągających pomiędzy otaczającymi elementami struktury zmniejsza opór tylko o $\simeq$ 5%. Tak więc rdzeniem imadła są ruchy ścinające antyrównoległych spinek β, a całe imadło jest tworzone przez sieć połączeń z jego otoczeniem.

Nieustrukturyzowane lub zdelokalizowane mechaniczne imadła

W kilku białkach, takich jak 1qp1 i 1tum, znalezione przez nas elementy struktury odpowiedzialne za największy opór są nieustrukturyzowane, jeśli przyjmie się ścisłe kryterium istnienia struktury drugorzędowej, w tym przypadku dla nici β. Np. w białku 1tum średnia odległość między odcinkami łańcucha głównego w przedziale aminokwasów zidentyfikowanych jako mechaniczne imadło 2-6, 78-82 wynosi 4.3 Å, jednakże eliminacja w tym rejonie kontaktów obniża siłę z 3.1 do 2.1 $\epsilon/$Å. W przypadku białka 1qp1 mechaniczne imadło zaznaczone czarnym kolorem na rysunku 6.15 zawiera się między aminokwasami 9-13 oraz 102-106. Eliminacja kontaktów w tym rejonie powoduje redukcję siły z 2.9 do 2.1 $\epsilon/$Å. W przypadku białka 1f86 eliminacja kontaktów między 11-19 oraz 104-122 redukuje siłę z 2.9 do 2.1 $\epsilon/$Å.

Rodzaj nieustrukturyzowanego i dodatkowo zdelokalizowanego imadła ma miejsce w białku 1b88. Po pierwsze eliminacja kontaktów w rejonie aminowego terminalu redukuje siłę z 3.2 do 2.1 $\epsilon/$Å. Po drugie eliminacja kontaktów między antyrównoległymi nićmi β zmniejsza siłę z 3.2 do 2.8, tworząc drugi słabszy rdzeń imadła.

Identyfikację elementów struktury drugorzędowej przeprowadziliśmy na podstawie informacji zawartej w pliku PDB.

Prostopadłościan jako motyw mechanicznego imadła

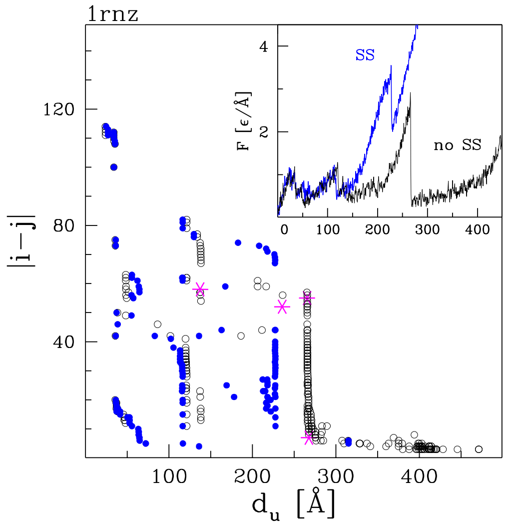
Ten motyw jest realizowany przez białko 1pav, które należy do klasy α/β. Mechaniczne imadło jest tworzone przez prostopadłościan zbudowany z antyrównoległych nici B (58-60) i E (73-76), połączonych wiązaniami wodorowymi i dwóch helis A (19-28) i C (44-57). Helisy te leżą równolegle i są rozciągane w przeciwnych kierunkach. Podczas rozciągania wszystkie ściany są ścinane w przeciwnym kierunku i trą o siebie. Usunięcie A+C, A+E, C+E i B+E redukuje siłę niemalże w równym stopniu od każdej pary o $\simeq$0.5 ϵ/Å. Mniejszy wpływ na ścinanie mają kontakty między helisami.

Reasumując, w pierwszym omówionym motywie mechanicznego imadła nici β były antyrównoległe. W drugim części imadła są tworzone z nieustrukturyzowanych równoległych odcinków białka, przypominających nici β, połączonych w większości wiązaniami wodorowymi. W trzecim typie imadło było kształtu prostopadłościanu, którego ściany ulegają ścinaniu i tarciu o siebie podczas rozciągania.

6.2.8 Rola mostków siarczkowych

Mostki siarczkowe (SS) pomiędzy cysteinami są natury kowalencyjnej i nie podlegają zrywaniu w zakresie sił mikroskopu atomowego. Zastosowanie takiej samej siły wiązania do opisu oddziaływań dla mostków SS i pozostałych kontaktów natywnych może prowadzić do ich zerwania i zmiany mechanizmu rozwijania w porównaniu do wyników doświadczalnych. Powstały problem jest zilustrowany na rysunku 6.16 na przykładzie białka 1rnz (ribonukleaza A). W przypadku, gdy mostki SS nie są w symulacji rozciągania białka 1rnz wzmocnione, obserwujemy największy opór F_{max}=2.7 ϵ/Å dla d_u=250Å. W białku tym są cztery mostki SS, z których w sytuacji nie wzmocnienia, dwa pękają przed F_{max}, a dwa pozostałe dają wkład do F_{max} (miejsce zerwania kontaktów SS jest oznaczone dużą gwiazdką na rysunku 6.16). Wprowadzenie 20-krotnie większej siły dla kontaktów typu SS prowadzi do wcześniejszego pojawienia się największego mechanicznego oporu o większej wartości F_{max}=3.2 ϵ/Å. Tym samym białko to awansuje na pozycję 14. w zbiorze S134. Scenariusz z rysunku 6.16 pokazuje, iż pierwsza połowa procesu rozwijania przebiega w obu modelach tak samo, jednakże potem przebieg jest zupełnie odmienny i imadło mechaniczne jest skonstruowane z innych wiązań. Podobna sytuacja występuje dla homologicznie pokrewnego białka 1rbj.

W większości rozważanych przypadków białek obecność wiązań SS nie wpływa na wartość F_{max}, zwłaszcza gdy pęknięcie tego rodzaju kontaktu odbywa się w prostym modelu po osiągnięciu F_{max}. Wyjątek stanowią białka, w których wiązania SS zamykają większą część struktury uniemożliwiając dalsze jej rozciągnięcie. Taka sytuacja występuje np. w białku 1lsl, które zawiera 6 mostków siarczkowych i jedno z nich ogranicza rozciągnięcie 113. aminokwasu białka do odcinka złożonego z 37 aminokwasów. Ten rozciągnięty segment nie wykazuje ponadto żądnego maksimum siły. Wszystkie tego typu białka obecne w zbiorze S7510 zostały uwzględnione w wykresie siły 6.1 po-



Rysunek 6.16: Diagramy scenariuszowe do procesu rozwijania oraz wykres siły $F(d)$ (panel wewnętrzny) dla białka 1rnz w ramach dwóch rodzajów modelu Go. W pierwszym modelu siła oddziaływania mostków SS jest taka sama jak innych natywnych kontaktów. Odpowiadająca mu krzywa $F(d)$ jest cieńsza, a symbole na scenariuszu są zaznaczone przez otwarte symbole i gwiazdki. Drugi model nie pozwala na zrywanie mostków siarczkowych. Odpowiada mu oznakowanie niebieskim kolorem krzywej $F(d)$ i pełne symbole.

przez wysokość kolumny w zerze. Porównanie rozciągania białek z wiązaniami SS i bez nich ma sens fizyczny, gdyż wiązania SS można zmienić na znacznie słabsze wiązania typu SH (grupa hydrosulfidowa) poprzez zastosowanie czynnika redukującego DTT (1,4 ditiotreitol). Tego typu badania mechanizmu rozciągania pod wpływem DDT były przeprowadzone np. dla molekuły adhezji o symbolu Mel-Cam (dane zawarte w tabeli 1.1). Tak więc zbiór testowy 29 białek, który służył do weryfikacji poprawności modelu, uwzględniał także ten rodzaj mechanicznego oporu.

6.2.9 Wpływ grup bocznych

W rozdziale 4 pokazaliśmy, iż model $\{6-12, C, M3, E^o\}$ jest optymalną formą modelu Go do pomiaru maksymalnej mechanicznej siły oporu dla każdej klasy białek. Tym niemniej model z grupami bocznymi $\{C^{\alpha-\beta}, 6-12, C, M3, E^o\}$ w przypadku białek klasy β lub α/β dawał porównywalną korelację F_{max} z doświadczeniem. Dodatkowym jego atutem było w przypadku białek 1tit, ^{10}FnIII, ^{12}FnIII, ^{13}FnIII i GFP otrzymanie kształtu krzywej $F(d)$ zgodnego z wynikami doświadczalnymi. Tak więc ten właśnie model wybraliśmy do przeprowadzenia weryfikacji otrzymanego rankingu najsilniejszych białek.

Wartości $F_{max}^{C\alpha\beta}$ dla zbioru S134 białek w modelu z grupami bocznymi są wymienione w tabeli A.2 na końcu pracy. W wyniku porównania otrzymanych wartości F_{max} dla obu typów modeli otrzymaliśmy liniową korelację między siłami ze współczynnikiem Pearsona 1.749 i wartością współczynnika U wynoszącą 0.34, co oznacza, iż ranking dla większości białek pozostaje taki sam, poza kilkoma wyjątkami. W szczególności stwierdziliśmy, iż cztery najsilniejsze białka pozostają na swoich pozycjach. Do wyjątków, które stają się znacząco silniejsze w nowym modelu, należą 1pun, 1kot i 1tum – zmieniają one pozycje z 38, 77 i 112 na 5, 8 i 6. Pozostałe wyjątki tworzą grupę białek, które stają się zdecydowanie słabsze; należą do nich: 1kiq, 1a2y, 1lve, 1amx, 1qd0, 1j05, 1rbj, 1gke, 1etb, 1em7 i 5lve.

Wpływ modelu z grupami bocznymi na kształt krzywych $F(d)$

Uwzględnienie grup bocznych w modelu spowodowało zmianę kształtu $F(d)$ w 49 przypadkach ze zbioru S134. Typowe zmiany polegają na

- zaniku średniej wielkości maksimów,
- w kilku przypadkach pojawieniu się jednego lub dwóch mniejszych maksimów,
- wystąpieniu znacznie większej wartości siły, przesuniętej w kierunku większych odległości.

Pierwszy przypadek jest najbardziej typowy i prowadzi do wchłonięcia średniej wielkości maksimum występującego po największym maksimum w jego obszar. Taka sytuacja miała miejsce w przypadku 1tit i była pokazana w podrozdziale 4.2.3 podczas wyboru optymalnego modelu Go. Poza tymi głównymi trendami w białkach o podobnej klasyfikacji $CATH$ do 1oo2 znikają zarówno maksima przed i po głównym maksimum, jak też przesunięcie jego położenia. Szczegółowa analiza wpływu grup bocznych na najsilniejsze białka jest zawarta w pracy [IV].

Rozdział 7

Własności mechaniczne białek z węzłami

W rozdziale niniejszym analizujemy proces rozciągania białek z węzłami, używając w tym celu opisanego w rozdziale 3 modelu Go o parametrach $\{6-12, C, M3\}$. Głównym naszym wynikiem jest wykazanie, iż podczas rozciągania białek węzły zaciskają się w sposób skokowy, w rezultacie przeskakiwania ich końców w ściśle określone położenia zdeterminowane geometrią głównego łańcucha, takie jak ostre zakręty lub początki helis. Jest to pierwszy opis takiego zachowania; w szczególności jest ono odmienne od dobrze zbadanego zaciskania się węzłów w polimerach, które następuje w wyniku dyfuzyjnego ruchu ich końców. Wykazujemy ponadto, iż skokowy ruch węzłów w białkach prowadzi do ich całkowitego zaciśnięcia na łańcuchu głównym, w przeciwieństwie do schodzenia węzłów z łańcucha polimeru w wyniku ruchu dyfuzyjnego. Badamy także czasy oczekiwania na skok końców węzła do kolejnych położeń, i pokazujemy iż rosną one stochastycznie wraz ze wzrostem temperatury. Ponadto pokazujemy, iż termodynamiczna stabilność białek z węzłami zależy między innymi od głębokości osadzenia węzła wzdłuż łańcucha. Opisujemy także proces powrotu białka do konfiguracji natywnej po jego rozciągnięciu, a następnie swobodnym puszczeniu.

Wyniki tych badań określające główne własności węzłów w białkach są opublikowane w pracy [VII].

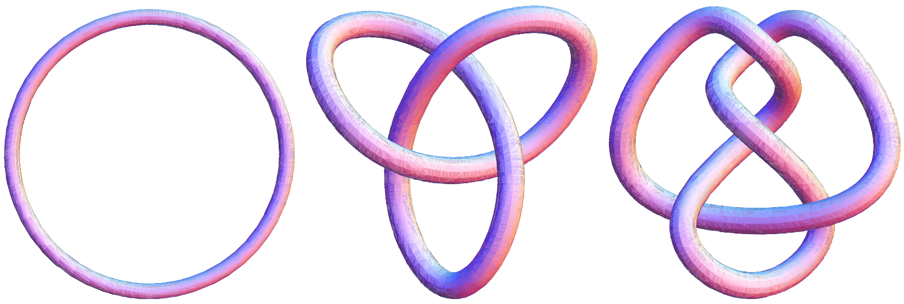
7.1 Węzły

Klasyfikacja węzłów

Według definicji matematycznej za węzeł uważa się zamkniętą nie przecinająca się krzywą zanurzoną w trzech wymiarach. Jednym z głównych celów matematycznej teorii węzłów jest dokonanie ich pełnej klasyfikacji. W tym celu konstruuje się tzw.

niezmienniki węzłów, czyli pewne obiekty które zależą tylko od topologii węzła, a nie jego kształtu. Jeśli dla danych dwóch węzłów ich niezmienniki są różne, oznacza to, iż nie można ich nałożyć na siebie bez dokonywania żadnych przecięć. Najprostszym takim niezmiennikiem, aczkolwiek wystarczającym do naszych celów, jest minimalna liczba skrzyżowań otrzymana w wyniku zrzutowania węzła na dowolną dwuwymiarową płaszczyznę. Niezmiennik ten jest też podstawą powszechnie stosowanego nazewnictwa węzłów, przy czym kolejne węzły o tej samej liczbie skrzyżowań rozróżnia się dodatkowym numerem, w oznaczeniach umieszczanym w dolnym indeksie liczby skrzyżowań.

Trzy najprostsze typy węzłów są pokazane na rysunku 7.1. Pierwszy z nich to tzw. węzeł trywialny z zerową liczbą skrzyżowań, oznaczony symbolem 0_1. Drugi to tzw. węzeł trójlistny 3_1 z trzema skrzyżowaniami. Kolejny węzeł to tzw. ósemka 4_1 z czterema przepleceniami. Są to jedyne węzły z jednym, trzema i czterema skrzyżowaniami. Istnieją natomiast dwa różne węzły z pięcioma skrzyżowaniami, oznaczane 5_1 oraz 5_2. Dla większej liczby skrzyżowań, ilość różnych węzłów je posiadających rośnie coraz szybciej.

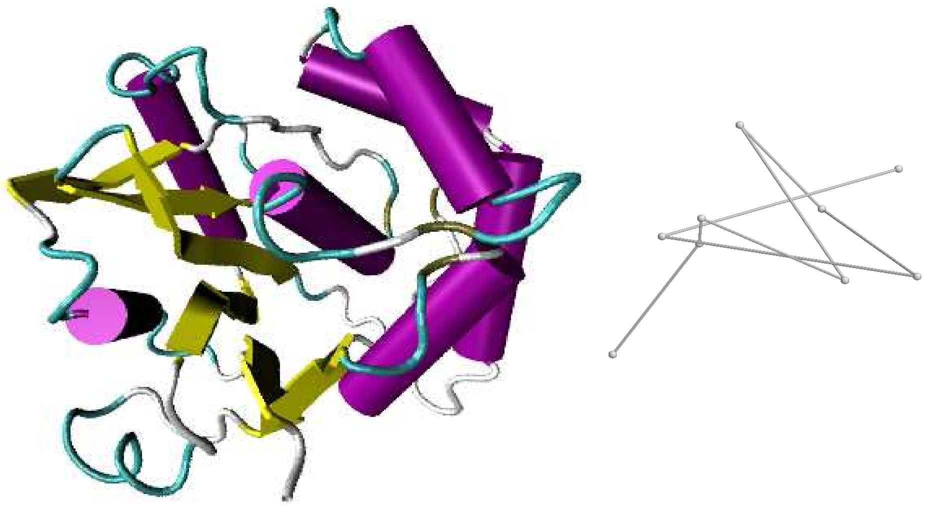


Rysunek 7.1: Od lewej: węzły typu 0_1, 3_1 i 4_1.

Łańcuch aminokwasów z których jest zbudowane białko nie tworzy krzywej zamkniętej. Tym niemniej często się zdarza, iż końce białka w konformacji natywnej są wyraźnie oddalone od jego mocno pozwijanego rdzenia, np. nie mniej niż o 10 atomów C^α. W takim przypadkach, na których też skupiamy się w niniejszej pracy, końce te można w jednoznaczny sposób połączyć dodatkowym zewnętrznym segmentem, tak aby nie tworzył on żadnej dodatkowej pętli, nie przecinał wnętrza białka i tworzył wraz z łańcuchem aminokwasów zamkniętą pętlę. Okazuje się, iż rzeczywiście w przyrodzie występują białka, dla których łańcuch główny tworzy nietrywialny węzeł po domknięciu go takim dodatkowym segmentem.

Algorytm KMT

Aby stwierdzić czy dane białko reprezentuje nietrywialny węzeł używamy tzw. algorytmu KMT [45, 36]. Algorytm ten polega na redukcji głównego łańcucha poprzez usuwanie kolejnych węgli C^α, pod warunkiem, że trójkąt utworzony przez rozważany atom i dwóch jego najbliższych sąsiadów nie jest przecinany przez żaden inny odcinek łańcucha głównego. Jeśli w wyniku wielokrotnego zastosowania tej procedury łańcuch zostanie zredukowany jedynie do dwóch węgli C^α – początkowego i końcowego – oznacza to, iż białko nie tworzyło nietrywialnego węzła. W przeciwnym wypadku, gdy nie jest możliwe zredukowanie łańcucha głównego jedynie do skrajnych atomów, oznacza to, iż białko reprezentuje pewien nietrywialny węzeł. Przykład działania algorytmu KMT dla białka 2etl posiadającego węzeł 5_2 jest pokazany na rysunku 7.2.



Rysunek 7.2: Po lewej: białko 2etl (posiadające węzeł 5_2) w stanie natywnym. Po prawej: algorytm KMT redukuje to białko do 9 atomów C^α połączonych 8 odcinkami z których żadnego nie można już usunąć, co świadczy o obecności węzła w wyjściowej konfiguracji.

Ponadto algorytm KMT pozwala na zdefiniowanie końców węzła oraz jego rozmiaru (długości) wzdłuż łańcucha głównego. W tym celu przeprowadzamy następująca procedurę. Dla danego białka złożonego z N aminokwasów (węgli C^α) posiadającego nietrywialny węzeł, usuwamy po kolei po jednym atomie C^α z ustalonego końca łańcucha głównego. Po każdym takim usunięciu atomu sprawdzamy, czy w dalszym ciągu algorytm KMT wykrywa obecność węzła. Odległość pomiędzy danym końcem łańcucha głównego (dla ustalenia uwagi — końcem aminowym) a najbliższym mu końcem węzła definiujemy jako minimalną liczbę atomów C^α które należy usunąć, by węzeł przestał być wykrywany. Liczbę tę oznaczamy jako n_1. Podobnie możemy wyznaczyć odległość n'_1 pomiędzy drugim (karboksylowym) końcem białka a drugim końcem węzła. Wprowadzamy ponadto odległość drugiego końca węzła od końca aminowego białka $n_2 = N - n'_1 + 1$. Wreszcie, rozmiar węzła wzdłuż łańcucha głównego definiujemy jako $n = N - n_1 - n'_1 = n_2 - n_1 - 1$.

Należy podkreślić, iż algorytm KMT pozwala jedynie na stwierdzenie czy białko tworzy nietrywialny węzeł, natomiast nie pozwala on na identyfikację rodzaju węzła. Tym niemniej dla naszych potrzeb jest on wystarczający, jako że w białkach występują jedynie najprostsze węzły, których rodzaj bez trudu można określić w razie potrzeby z rysunku białka. Wspomnimy tu jedynie, iż do rozróżniania rodzaju węzłów stosować można bardziej skomplikowane niezmienniki, np. różnorakie wielomiany węzłowe; stosowane są one np. w badaniu topologii polimerów.

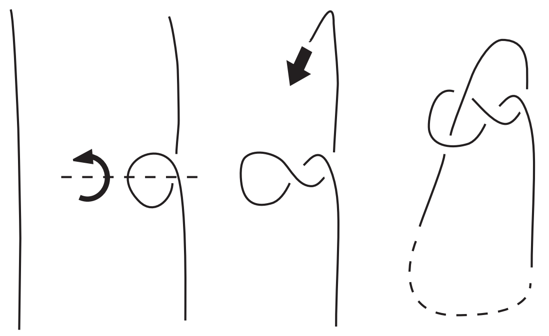
Występowanie w białkach węzłów 5_1 i 5_2

Jak wyżej wspomnieniliśmy, istnieją dwa węzły z pięcioma skrzyżowaniami. Tym niemniej, przynajmniej dotychczas w białkach znaleziono jedynie węzeł 5_2. Między innymi białka z tym właśnie węzłem analizujemy w dalszej części tego rozdziału. Zanim tego dokonamy, wyjaśnijmy dlaczego węzeł 5_1 w białkach nie występuje (lub też dotychczas nie został znaleziony). Związane jest to z faktem, iż w białkach pojawiają się jedynie najprostsze rodzaje węzłów. Konstrukcja węzła typu 5_2 jest pokazana na rysunku 7.3 i polega ona na złożeniu łańcucha w motyw spinki, a następnie podwójnym jego skręceniu i przełożeniu przez skręcenie pozostałego końca. Aby ten węzeł rozwiązać, wystarczy jeden z końców wysunąć z powstałej pętli. Natomiast węzeł 5_1 jest bardziej skomplikowany w tym sensie, że wysunięcie jednego jego końca z analogicznej pętli nie prowadzi do zupełnego rozwiązania węzła, lecz do powstania struktury typu 3_1, do której rozwiązania trzeba jeszcze raz wysunąć jeden z końców węzła z kolejnej pętli. [1]

Fizyczne własności węzłów

Węzły powszechnie występują w innych obiektach biologicznych, takich jak DNA i RNA [41, 42, 43]. Węzły w DNA powstają między innymi podczas replikacji i są usuwane przez enzymy o nazwane topoizomerazami [32]. Własności węzłów na DNA i RNA są badane między innymi poprzez techniki mikroskopu sił atomowych i szczypców optycznych, oraz porównywane do własności węzłów na polimerach. Najlepiej od

[1] Obserwacja ta została poczyniona niezależnie także w pracy [38].



Rysunek 7.3: Schemat powstawania węzła typu 5_2.

strony teoretycznej i symulacji numerycznych poznane zostały własności węzłów na polimerach [45, 46, 47, 165, 166] oraz na linach idealnych [44, 167].

W kontekście uzyskanych przez nas wyników dotyczących białek warto najpierw scharakteryzować zachowanie się węzłów na polimerach, które w symulacjach modeluje się jako zbiór kulek połączonych sprężynkami. Modele polimerów przypominają zatem stosowane przez nas modele gruboziarniste dla białek, aczkolwiek zasadniczą różnicę stanowi brak oddziaływań natywnych, które są obecne w białkach. W przypadku stosunkowo zaciśniętego węzła, jego położenie na polimerze można w przybliżony sposób określić jednym parametrem odpowiadającym odległości węzła od jednego z końców łańcucha. Bardziej ogólnie, nawet w przypadku „luźnego" (nie zaciśniętego) węzła, jego zachowanie można scharakteryzować położeniem porzez podanie położenia jego końców n_1 oraz n_2, które dla polimerów można zdefiniować tak samo jak w przypadku białek (jak opisaliśmy powyżej w części dotyczącej algorytmu KMT). W szczególności z przeprowadzonych symulacji numerycznych wynika, że w warunkach otoczenia odpowiadających naszym symulacjom białek, węzły w polimerach poruszają się wzdłuż łańcucha głównego ruchem dyfuzyjnym, a w większości przypadków po pewnym czasie ześlizgują się z niego [165, 166]. Wyniki te są także zgodne z obserwacjami doświadczalnymi dla polimerów [41]. Jak przedstawimy w dalszej części tego rozdziału, zachowanie się węzłów w białkach jest zupełnie inne i w pewnym sensie dużo ciekawsze, jako że daje ono wgląd w strukturę geometryczną białek.

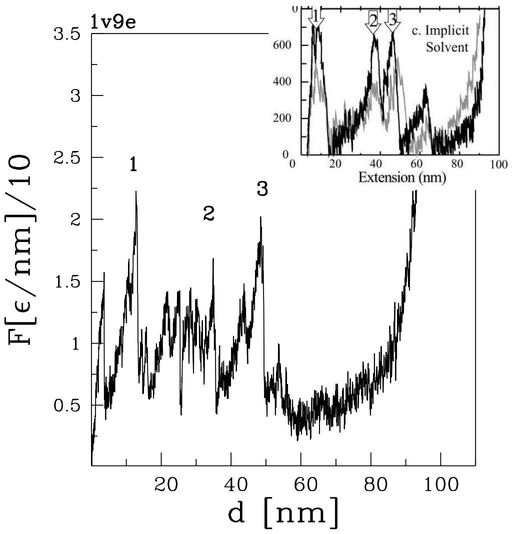
Natomiast wyniki prac dotyczące lin idealnych [44, 167] nie stosują się bezpośrednio do badanych przez nas układów. Po pierwsze, lina idealna modelowana jest poprzez cylinder o stałej szerokości, w którym istotną rolę odgrywa ślizganie się po sobie różnych jego fragmentów — taki efekt nie występuje w rozpatrywanych przez nas modelach węzłów złożonych z kulek połączonych odcinkami (sprężynkami). Po drugie, przy pomocy modelu liny idealnej dotychczas badano w szczególności maksymalnie zaciśnięte konfiguracje węzłów; w procesie ich zaciskania uwzględniane były stosunkowo niewielkie fluktuacje termiczne. Natomiast ruch dyfuzyjny węzłów w polimerach — z którym porównujemy nasze wyniki dla białek — wynika z rozpatrywanych przez nas znacznie większych fluktuacji. Tym niemniej, przeprowadzając nasze symulacje w temperaturze bliskiej zera, istotnie zaobserwowaliśmy zanik ruchu dyfuzyjnego końców węzła na rzecz jego jednostajnego zaciskania, co jakościowo odpowiada charakterowi zaciskania węzłów na linie idealnej.

7.2 Rozciąganie białek z węzłami

Węzły w białkach do tej pory pozostawały słabo zbadanymi obiektami. Obecnie znanych jest około 300 białek, w których zostały zidentyfikowane opisane wyżej trzy typy węzłów 3_1, 4_1, 5_2. Pierwsze prace doświadczalne nad stabilnością termodynamiczną i mechanizmem zwijania tego typu białek były przeprowadzone zaledwie dwa lata temu przez Mallam i in. [168, 169]. Rok później dla tego samego białka zostały przeprowadzone pierwsze symulacje teoretyczne procesu zwijania [170]. Doprowadziły one do zapętlenia białka, jednakże bez wyjaśnienia mechanizmu, który determinowałby ten proces. Jeden z możliwych mechanizmów powstawania węzłów w białkach został zaproponowany w pracy [38]. Jednakże do tej pory nie zostały wyjaśnione własności dynamiczne węzłów w białkach.

Własności dynamiczne węzłów w białkach mogą być badane (tak jak zostało to zrobione w przypadku węzłów na DNA i RNA) poprzez ich rozciąganie np. za pomocą mikroskopu sił atomowych lub szczypiec optycznych. Mimo że wydawałoby się, iż ten rodzaj eksperymentu został dobrze opanowany, do tej pory nie zostało rozciągnięte żadne białko z węzłem w sposób który umożliwiałby poznanie dynamiki węzłów. Z przeglądu literatury wynika, iż białko z węzłem zostało poddane rozciąganiu, jednakże miejsce zaczepienia było tak wybrane aby rozsupłać węzeł [171]. Eksperyment ten został następnie powtórzony przy pomocy pełnoatomowej symulacji teoretycznej [172].

Główne własności węzłów w białkach zbadamy na podstawie symulacji rozciągania ze stałą prędkością w ramach modelu Go o parametrach $\{6-12, C, M3\}$. Dynamikę oraz mechaniczne własności białek z węzłami badamy na przykładzie 18 białek z węzłem trójlistnym 3_1 (1j85, 1od6, 1dmx, 1jd0, 1j86, 1ipa, 1js1, 1k3r, 1kop, 1nxz, 1v9e, 1x7p, 1v2x, 1fug, 1vh0, 1zrj, 1hcb, 1keq) oraz dwóch białek z węzłem 5_2 (2etl i 1xd3) [39]. Charakterystyka modelu i protokół symulacji jest zgodny z opisem z rozdziału 4.1.2. Dodatkowo wykonaliśmy tutaj rozciąganie w kilku różnych prędkościach v_p 0.0005, 0.001, 0.01 i 0.05 Å$/\tau$, które są średniowane odpowiednio co 1Å, 0.5Å, 0.25Å i 0.25Å. Symulacje rozciągania dla białek 1j85, 1od6, 1vho, 1kop i 2etl były przepro-

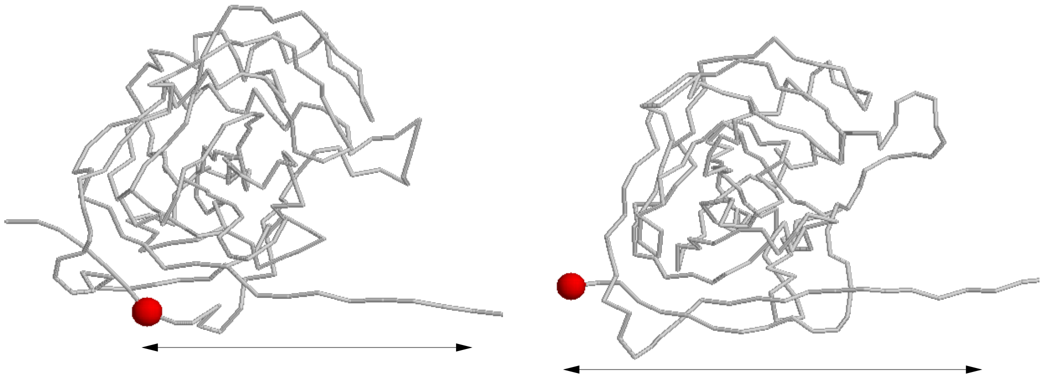


Rysunek 7.4: Panel główny: wykres siły w funkcji rozciągania dla białka z węzłem (1v9e) otrzymany z symulacji modelem Go. Panel wewnętrzny: ten sam proces zarejestrowany w symulacji pełnoatomowej z pracy [172]. Ciemna i jasna krzywa odpowiadają różnym prędkościom ciągnięcia, odpowiednio szybszej i wolniejszej.

wadzone 700 razy, a dla pozostałych po 10 razy. Zgodność wyników teoretycznych z doświadczalnymi dla procesu rozciągania białek w tym modelu pokazaliśmy już w rozdziale 4.1.2 i pracy [V]. Tym niemniej, rysunek 7.4 pokazuje wykres siły w funkcji przemieszczenia dla rozciągania białka z węzłem w modelu $\{6-12, C, M3\}$ i pełnoatomowej symulacji. W obu przypadkach białko anhydraza (1v9e) było rozciągane za aminokwas N, gdy 253 aminokwas był zaczepiony na stałe. Ten kierunek ciągnięcia umożliwia rozsupłanie węzła. Rysunek 7.5 przedstawia konformacje białek w tej samej chwili czasu dla kierunku rozciągania powodującego rozsupłanie (253-N) i zasupłanie (259-N). Otrzymaliśmy trzy maksima dla rozciągnięć w okolicach 130, 370 i 490 Å, które są zgodne z wynikami eksperymentalnymi [171] i teoretycznymi [172]. Ponadto analiza diagramu scenariuszowgo (dane nie pokazane) pokazała, iż rozrywanie kontaktów dla kluczowych elementów białka jest zgodne z wynikami pełnoatomowymi [172].

7.2.1 Mechanizm zaciskania się węzłów

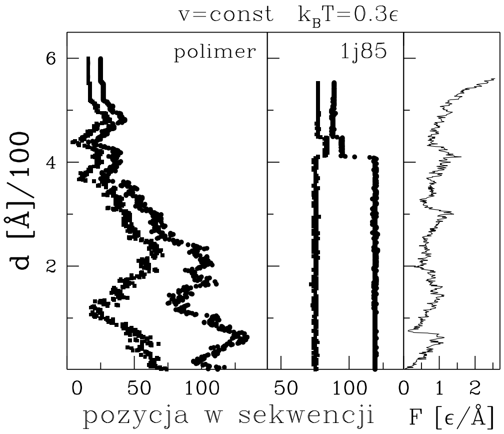
Aby zanalizować ruch końców węzła wprowadzamy diagramy takie jak ten pokazany na panelu środkowym rysunku 7.6. Odpowiada on białku YIBK (1j85) o 156 aminokwasach, które tworzą prosty trójlistny węzeł o końcach usytuowanych w aminokwasach o numerach n_1=75 i n_2=119 w stanie natywnym. Tak więc węzeł zajmuje w przybliżeniu 1/3 całej sekwencji w stanie natywnym. Diagram środkowy pokazuje



Rysunek 7.5: Chwilowe konformacje białka 1v9e podczas rozciągania w modelu Go dla d 105 Å. Na lewo białko jest rozciąganie za aminokwasy 253 (oznaczony czerwona kulką) i 1, co prowadzi to do rozsupłania węzła. Na prawo białko jest rozciąganie za aminokwasy 259 (oznaczony czerwoną kulką) i 1, co prowadzi do zaciśnięcia węzła.

ewolucję końców węzła n_i podczas gdy białko jest rozciągane, a odpowiadający temu wykres siły oporu F w funkcji przemieszczenia d jest pokazany na prawym panelu tegoż rysunku. Otrzymany wykres siły ma nietypowo największe maksimum pod sam koniec rozciągania, poprzedzone serią średniej wielkości maksimów wzdłuż całej drogi rozwijania. Tym niemniej widzimy, iż końce węzła pozostają w pozycji natywnej aż do momentu wystąpienia największego maksimum. Największy opór pojawia się w odległości około 400 Å i wtedy to oba końce węzła skaczą w swoim kierunku, aby pozostać tam do kolejnego skoku po około 50 Å, także związanego z maksimum siły. Drugi skok powoduje powstanie minimalnej objętości węzła stworzonej z około 12 atomów C^{α} i zaciśnięcie się na łańcuchu aminokwasów.

Obserwowana własność ruchów skokowych węzła nie występuje ani podczas podgrzewania białek (dane nie pokazane) ani też podczas rozciągania homopolimerów z kontaktami czysto odpychającymi (lewy panel na rysunku 7.6). Konformacją początkową dla symulacji homopolimeru było białko z węzłem w którym natywne kontakty zostały zastąpione odpychającymi oddziaływaniami. Natomiast zachowanie się białka podobne do homopolimerów można zaobserwować gdy temperatura otoczenia zostanie podgrzana powyżej maksimum ciepła właściwego. W homopolimerach końce węzła poruszają się ruchem dyfuzyjnym, a odległość między końcami polimeru gwałtownie wzrasta gdy jeden z końców węzła ześlizgnie się z łańcucha. W obserwowanych przez nas 5 różnych homopolimerach (różnią się one długością łańcucha i wielkością węzła), dla ponad 90% trajektorii po jakimś czasie końce węzła pozostają cały czas ruchliwe aż do momentu ześlizgnięcia się z łańcucha. Należy jednak zauważyć, iż w pozostałych $\simeq$10% przypadków węzeł może się zacisnąć w przypadkowym miejscu, tworząc mocno upakowaną formę. Węzeł pozostaje wtedy nieruchomy, chyba że nastąpi bardzo duża fluktuacja termiczna [165, 44, 173]. Otrzymane przez nas wyniki zgadzają się z wcześniejszymi badaniami nad dynamiką węzłów w polimerach i DNA, gdzie dyfuzyjny charakter ruchów węzłów był pokazany zarówno doświadczalnie [41] jak i teoretycznie [165, 166]. Głównym powodem tak odmiennej dynamiki węzłów w białkach od tej występującej w polimerach jest obecność złożonej sieci połączeń pomiędzy aminokwasami w białku.



Rysunek 7.6: Ruchy dyfuzyjne końców węzła w polimerze (lewy panel) oraz ruchy skokowe końców węzła w białku 1j85 (środkowy panel) podczas rozciągania z stałą prędkością. Kwadraty (lewy panel) i koła (środkowy panel) wskazują pozycję końców węzła wzdłuż sekwencji. Należy podkreślić, iż węzły z polimerów zazwyczaj schodzą podczas rozciągania. Tutaj jednakże pokazujemy przypadek kiedy węzeł zaciska się blisko jednego z końców. Przeciwnie zachowują się węzły w białkach, które po serii skoków zawsze się zaciskają w szczególnych położeniach wewnątrz konfiguracji początkowej węzła. Skok węzła do nowego położenia jest widoczny jako maksimum dla około 400 Å na wykresie siły w funkcji przemieszczenia otrzymanej z symulacji rozciągania (prawy panel).

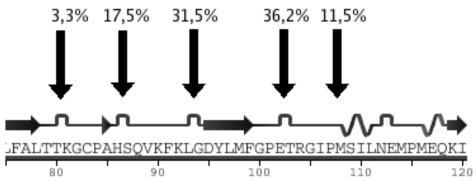
W obu rozważanych przypadkach, polimeru i białka, ruchy końców węzła zależą od konkretnej trajektorii, mimo że wykres siły w funkcji przemieszczenia $F(d)$ może pozostawać bardzo podobny. W szczególności końce węzła mogą chwilowo przeskoczyć do nowego położenia wraz z pojawieniem się maksimum na wykresie $F(d)$. Jednakże, jak to jest pokazane na środkowym panelu 7.6, podczas rozciągania część białka, a zwłaszcza elementy struktury drugorzędowej zawarte wewnątrz węzła, pozostają nienaruszone, podczas gdy pozostałe elementy są sukcesywnie rozwijane. Dobrą ilustracją tego zjawiska jest zachowanie się białka 1od6. W tym przypadku dla ponad połowy trajektorii znajdujemy „częściowo zaciśnięty” węzeł zawierający nienaruszoną helisę α.

Należy zauważyć, iż w zastosowanym opisie dynamiki węzłów śledzimy tylko ruch końców węzła n_i wzdłuż sekwencji aminokwasów. Tym niemniej znaleźliśmy, iż rzeczywista odległość w przestrzeni między reziduami w rdzeniu węzła pozostaje w większości nie zmieniona pomiędzy kolejnymi skokami i ulega bardzo dużym zmianom jedynie podczas samych skoków węzła. Własność ta wskazuje na istnienie zależności między dynamiką węzła w rzeczywistej przestrzeni a ruchami węzła w przestrzeni sekwencji aminokwasów (wzdłuż sekwencji).

7.2.2 Miejsca zaciskania się węzłów

W paragrafie powyższym pokazaliśmy, iż ruch węzłów w polimerach odbywa się skokowo, aż do momentu ostatniego skoku, który prowadzi do zaciśnięcia się węzła w zakresie jego natywnego obszaru. Obecnie przeanalizujemy ostateczne miejsca zatrzymania końców węzła.

Analiza ponad 700 trajektorii pokazuje, iż końce węzła zatrzymują się tylko w kilku szczególnych miejscach wzdłuż sekwencji. Chwilowe zatrzymanie jednego z końców będziemy nazywać położeniem (stanem) metastabilnym. W szczególności stwierdzilismy, iż końcowe i metastabilne położenia końców węzła są zlokalizowane w ostrych zakrętach w łańcuchu białka tworzonych przez aminokwasy prolinę i glicynę, jak też w początkach helis. Ta odpowiedniość dla białek 1j85 i 2etl jest pokazana na rysunkach 7.7 i 7.8.

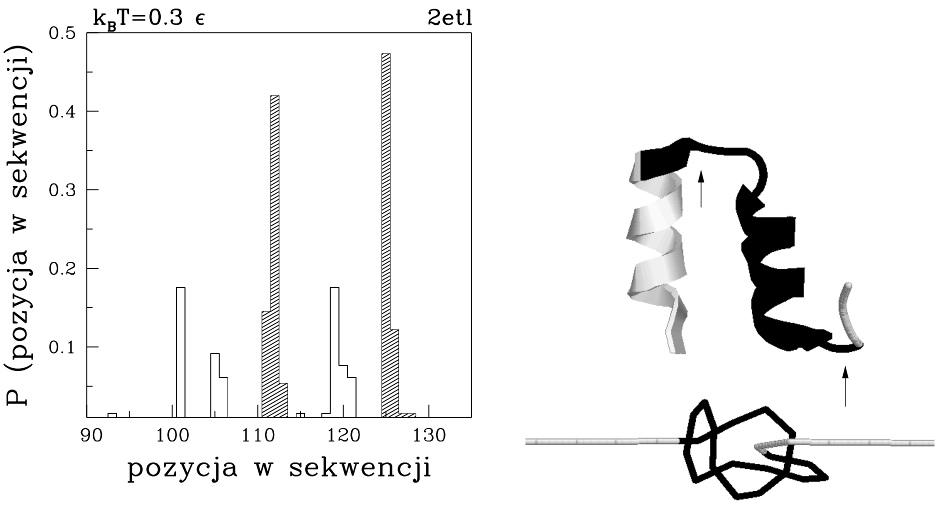


Rysunek 7.7: Sekwencja aminokwasów pomiędzy położeniem końców węzła n_1=75 i n_2=119 w stanie natywnym w białku 1j85. W konfiguracji zaciśniętej końce węzła przyjmują położenie między n_1 a n_2, z jednym z końców zlokalizowanym na ostrym zakręcie lub na początku helisy. Strzałki wskazują te charakterystyczne miejsca zatrzymania końców węzła podczas symulacji rozciągania. Liczby pokazują w procentach częstość zatrzymania w danym miejscu.

Charakterystykę miejsc zatrzymania przeprowadziliśmy badając dodatkowo wartości kątów wzdłuż łańcucha, korzystając z kryterium detekcji ostrych zakrętów (które także obejmują charakterystyczne wygięcie łańcucha na początku i końcu helisy) zaproponowanego przez Kuntza [174]. Znaleźliśmy, iż punkty zatrzymania węzłów odpowiadają głębokim minimom na wykresie wartości kątów liczonych pomiędzy co drugim wektorem wzdłuż łańcucha (np. $C_i^\alpha, C_{i+1}^\alpha$ i $C_{i+2}^\alpha, C_{i+3}^\alpha$). Zakręty te często są także stabilizowane przez otaczające je wiązania wodorowe, co powoduje, iż jeszcze trudniej jest je rozerwać. Dla wysokich temperatur ($k_B T > 0.5\epsilon$) ruchy węzłów stają się stopniowo coraz mniej przewidywalne i w konsekwencji ich końcowe położenie niekoniecznie musi się pokrywać z zakrętami ze stanu natywnego. W tych warunkach węzeł jest już w stanie wędrować poza obszar jego natywnego rdzenia. Z kolei dla $k_B T \gg \epsilon$ obserwujemy gładkie przejście do zachowania typowego dla homopolimerów, ze swobodną dyfuzją węzła wzdłuż łańcucha białka.

Białko w stanie natywnym zazwyczaj zawiera kilka ostrych zakrętów. Tak więc końce węzła skaczą do kilku możliwych położeń, nazwanych przez nas wcześniej położeniami metastabilnymi. Położenie tych miejsc wzdłuż sekwencji i prawdopodobieństwo zatrzymania się w nich końców dla białka 1j85 jest zilustrowane na rysunku 7.7. Potwierdzeniem tych obserwacji jest także zbadana przez nas dynamika znacznie bardziej skomplikowanego i większego węzła typu 5_2 w białka 2etl (węzeł ten obejmuje 174 spośród 223 atomów), jak pokazano na rysunku 7.8. W tym przypadku znajdujemy tylko dwa charakterystyczne miejsca końcowego zaciśnięcia się węzła, między reziduami 110 – 126 lub 101 – 119, w szerokim zakresie temperatur. Należy podkreślić, że wybór zakrętu dla początku lub końca węzła nie odnosi się do sytuacji natywnych, tzn. w stanach natywnych położenia końców węzłów w białkach nie wykazują żadnej szczególnej regularności.

Analizując procesy zaciskania dla całego zbioru 20 białek stwierdziliśmy, iż węzły dla niewysokich temperatur zaciskają się skokowo, zmniejszając początkowy zajmowany obszar odpowiadający węzłowi. Proces ten polega na skoku przynajmniej jednego z końców węzła do ostrego zakrętu. Mechanizm ten powtarza się aż do momentu zaciśnięcia się węzła na jednym z charakterystycznych zagięć łańcucha. W większości przypadków ostre zakręty zawierają aminokwas prolinę, która wymusza charakterystyczne zagięcie łańcucha o około 75° poprzez swoją strukturę pierścienia. W pozostałych sytuacjach zakręt w którym zatrzymuje się jeden z końców jest tworzony poprzez aminokwas glicynę (taka sytuacja ma miejsce w białkach 1od6, 1fug, 1vho, 1zrj, 1keq, 1v9e, ponadto 1v9e zawiera także zakręt z proliną). Glicyna jako jedyny aminokwas nie posiada grupy bocznej, co powoduje, iż przy odpowiednim ułożeniu otaczających ją aminokwasów w bardzo dużym stopniu łańcuch białka może się wygiąć. Pewnym

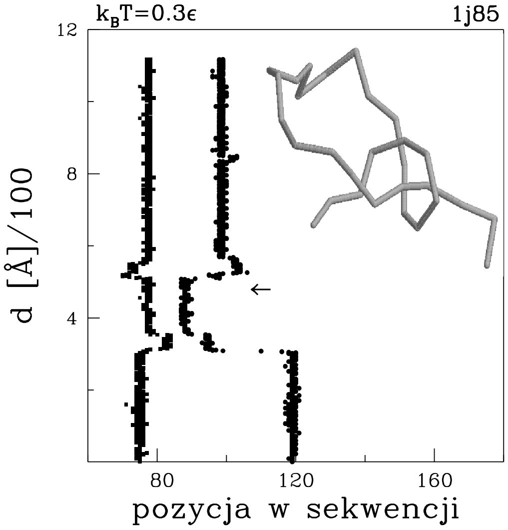


Rysunek 7.8: Rozkład prawdopodobieństwa zatrzymania się końców węzła w danym miejscu wzdłuż sekwencji dla białka 2etl. Zaciemnione słupki określają najbardziej prawdopodobne położenie końców zaciśniętego węzła: $n_1 \simeq$ 111 (koniec helisy) i $n_2 \simeq 126$ (zakręt). Otrzymana konfiguracja zaciśniętego węzła jest pokazana na prawo na dole, a odpowiadający jej natywny wycinek białka z zaznaczonymi strzałkami miejscami n_i nad nią. Mniej prawdopodobne miejsca zaciśnięcia oznaczone są białymi słupkami, i odpowiadają one $n_1 \simeq 101$ lub 106 i $n_2 \simeq 119$.

odstępstwem od opisanych powyżej reguł jest białko 1hcb, w którym zakręt zatrzymujący węzeł zawiera alaninę. W przypadku, kiedy węzeł podczas swoich skoków nie napotka na ostry zakręt, to zatrzymuje się on na najbliższym początku helisy, które to miejsce także tworzy charakterystyczne wygięcie.

7.2.3 Kinetyczne własności węzłów

W celu dokładniejszego poznania dynamiki węzłów przeprowadziliśmy zmodyfikowaną symulację rozciągania białek, w której po przejściu węzła do jednego z możliwych stanów metastabilnych końce białka zostały oswobodzone. Jeśli podczas symulacji rozciągania widocznych jest kilka maksimów siły i węzeł przejdzie przez co najmniej jeden stan metastabilny, wtedy, mimo oswobodzenia końców, ani białko, ani węzeł nie jest w stanie powrócić do konformacji natywnej. Węzeł zatrzyma się w jednym z ostrych zakrętów, a białko pozostanie w stanie zwiniętym częściowo. Nieodwracalność procesu zwijania zaobserwowaliśmy dla białek 2etl, 1vho i 1v2x, oraz dla 80% trajektorii dla białka 1od6. Zupełnie inną własnością charakteryzuje się białko 1j85, w którym poza kilkoma stanami pół-zwiniętymi – jak ten pokazany na rysunku 7.9 – białko znajduje



Rysunek 7.9: Położenie końców węzła podczas symulacji, w której dla rozciągania d=500Å (oznaczonego strzałką) końce białka zostały oswobodzone. Jest to rzadka trajektorii, w której końce węzła zatrzymują się w stanie metastabilnym uniemożliwiając ponowne zwinięcie się białka (odpowiadająca tej sytuacji konformacja węzła jest pokazana na prawo). Natomiast w większości trajektorii białko 1j85 powraca do stanu natywnego.

drogę do stanu natywnego, wraz z odtworzeniem konfiguracji natywnej węzła.

W celu zrozumienia odmiennej własności 1j85 w porównaniu do pozostałych 4 białek, sprawdziliśmy ich termiczną stabilność. Okazuje się, iż białko 1j85 stosunkowo łatwo znajduje drogę do rozwiązania węzła, podczas gdy proces ten jest znacznie dłuższy dla białka 1od6, a wręcz niemożliwy dla białka 1v2x. Białko 2etl bardzo szybko traci węzeł, jednakże wynika to z faktu, iż końce tylko nieznacznie wystają poza obszar węzła, tak więc lekkie drgania termiczne są wystarczające do rozwinięcia węzła, tak że przestaje być on wykrywany przez algorytm KMT.

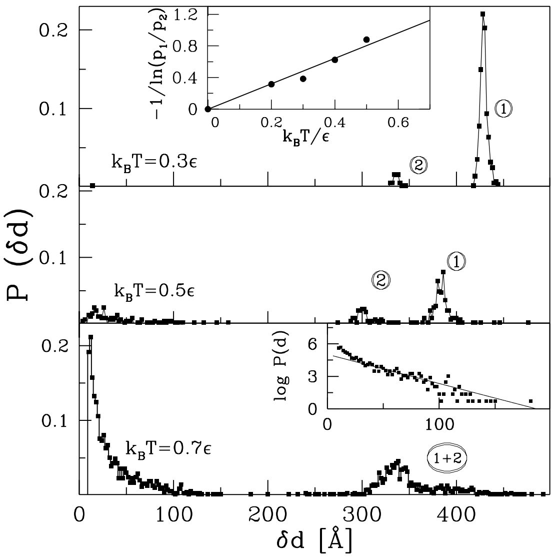
Obecnie wracamy do zagadnienia ponownego zwinięcia się białka 1j85 do stanu natywnego i braku tej własności dla pozostałych białek. Na podstawie wyników dla termicznego rozwijania wnioskujemy, iż własność ta jest skorelowana z możliwością rozwinięcia się tego białka, podczas gdy pozostałe białka wykazują dużą stabilność termiczną. Otrzymane wyniki dla białka 1j85 są zgodne z pracą teoretyczną [170] i doświadczalną [169], gdzie poprzez podgrzewanie lub w warunkach równowagowych pokazano słabą stabilność 1j85.

7.2.4 Czasy zaciskania się węzłów

W niniejszym paragrafie scharakteryzujemy czasy oczekiwania, δt, pomiędzy kolejnymi skokami końców węzła, wykonując symulacje rozciągania dla różnych wartości temperatury zredukowanej 0, 0.2, 0.3, 0.4, 0.5 i 0.7. Najwygodniejszym sposobem obserwacji czasów oczekiwania jest przedstawienie ich w funkcji odpowiadającego im przemieszczenia wektora prędkości $\delta d = v_p \delta t$ (w szczególności przemieszczenia te nie są w istotny sposób zależne od prędkości v_p). Otrzymaliśmy, iż dla $T = 0$ mechanizm zaciskania węzła jest deterministyczny, tak więc węzeł pozostaje w stanie natywnym bardzo długo, a gdy rozwinięcie osiągnie wartość δd=400 Å, węzeł dokonuje jednego skoku bezpośrednio do zaciśniętej konformacji. W momencie skoku rdzeń węzła składa się tylko z części struktury natywnej białka, która nie została jeszcze rozwinięta. Ten sposób rozwijania nazywamy ścieżką 1, której będzie odpowiadał histogram po prawej stronie paneli z rysunku 7.10. Rozciąganie dla coraz wyższych temperatur wprowadza czynnik stochastyczny do skoków węzła, w wyniku czego obserwujemy pojawienie się alternatywnej ścieżki rozwijania oznaczonej numerem 2. W ścieżce 2 węzeł zaciska się dla odległości δd=340 Å; jest to wcześniej niż całkowite rozwinięcie białka. Rysunek 7.10 pokazuje rozkład prawdopodobieństw zaciśnięcia się węzła poprzez ścieżkę 1 lub 2 (histogramy po prawej stronie). Na tym samym rysunku wzrost histogramu po lewej stronie odpowiada przypadkowym skokom końców węzła. Stosunek prawdopodobieństw wyboru tych ścieżek można zapisać następująco

$$\frac{p_1}{p_2} = exp\Big(- \frac{\Delta F}{k_B T}\Big),$$

gdzie ΔF ma interpretację bariery energii swobodnej dla przejścia między ścieżką 1 a ścieżką 2. Tę zależność wykreśliliśmy w wewnętrznym panelu na samej górze rysunku 7.10. Otrzymaliśmy, iż $k_B/\Delta F \simeq 1.6$. Wraz ze wzrostem temperatury skoki na ścieżkach stają się krótsze i zazwyczaj są poprzedzone kilkoma skokami o dużo mniejszej długości $\delta d <$ 100 Å. Te dodatkowe skoki pojawiają się znacząco często od temperatury $k_B T/\epsilon = 0.3$, tworząc wyraźny histogram po lewej stronie środkowego panelu 7.10. Powyżej $k_B T/\epsilon$=0.5 histogramy odpowiadające ścieżkom 1 i 2 rozmywają się. Dla $k_B T/\epsilon$=0.7 krótkie skoki końców węzła tworzą wyraźny histogram (po lewej stronie najniższego panelu), który można opisać rozkładem Poissona $P(\delta d) = \alpha^{-1} exp(\alpha \cdot \delta d)$. Potwierdzeniem tej własności jest liniowa zależność między $\log P(\delta d)$ a δd pokazana na panelu wewnętrznym, gdzie dane zostały dopasowane do linii prostej z współczynnikiem $\alpha \approx 0.27$.



Rysunek 7.10: Rozkład odległości oczekiwania δd na skok lewego końca węzła z białka 1j85 dla różnych temperaturach T. Symbole w kółkach oznaczają ścieżki rozwijania 1 i 2. Szczegółowy opis paneli zawarty jest w tekście.

Rozdział 8

Aplikacja do symulacji rozciągania białek wraz z bazą danych

Niniejszy rozdział zwięźle przedstawia konstrukcję bazy danych o własnościach mechanicznych białek oraz współpracującą z nią stronę internetową *BSDB* (Biomolecule Stretching Data Base), która w szybki i przejrzysty sposób umożliwia przeglądanie zasobów bazy. W rozdziale tym opisujemy też korzystanie z oprogramowania do samodzielnego rozciągania nowych biomolekuł, które jest integralną częścią *BSDB*.

Baza wraz z serwerem obliczeniowym w szczególności zawiera i podsumowuje wyniki przedstawione w rozdziałach 6 oraz 7 i jest dostępna na stronie internetowej *www.ifpan.edu.pl/BSDB*. Dokładny opis jej działania został zawarty w pracy IX.

8.1 Baza danych

Elektroniczną bazą danych nazywamy uporządkowany zbiór informacji oraz program komputerowy gromadzący i przetwarzający dane (DataBase Management System — DBMS). Dane w bazie są posegregowane i ułożone według kryterium ważności, tak aby mogły być one łatwo przeszukiwane.

Baza danych BSDB powstała w celu upublicznienia wyników dotyczących własności mechanicznych ponad 7500 białek, uzyskanych w symulacjach rozciągania w ramach modelu Go na podstawie przeglądu bazy *PDB*. Lista najsilniejszych białek i kilka ich najciekawszych zachowań zostały podane w rozdziale 6, jednakże zaprezentowanie całego zbioru wraz z wykresami korelacyjnymi w zależności między innymi od długości, klasy, architektury, topologii, czy też pokazanie statystyk jest nie możliwe na kilku stronach. Według nas najprzejrzystszym i najwygodniejszym sposobem prezentacji uzyskanych wyników jest przedstawienie ich w postaci internetowej bazy danych, którą skonstruowaliśmy i w tym rozdziale opisujemy.

Należy ponadto zauważyć, że liczba deponowanych struktur w *PDB* rośnie lawinowo. Zatem, aby baza była trwałym narzędziem, musi być ona systematycznie aktualizowana. Tak więc do strony obsługującej bazę dołączyliśmy oprogramowanie, które po pierwsze pozwala na samoaktualizację bazy (w zadanych odstępach czasu), a po drugie — na jej aktualizację przez użytkowników bądź autorów. Instrukcja obsługi tego oprogramowania jest opisana w paragrafie 8.3. Bazę danych umieściliśmy na osobnym komputerze, tak aby „nigdy" nie zabrakło w niej miejsca. Taki projekt wymagał zastosowania języka MySQL [175], który obsługuje zdalne bazy danych i jest wystarczająco szybki.

Konstrukcja bazy

Konstrukcja bazy polega na odpowiednim podziale zasobów informacji i zapisaniu ich w tabelach w postaci danych tekstowych jak też poprzez całe pliki binarne (np. obrazy). Niniejsza baza składa się z 3 rodzajów tabel, z których każda pełni inną funkcję.

Główna tabela, I, zawiera wyniki symulacji rozciągania białek opisane w rozdziale 6. Wyniki te są posegregowane w 19 kolumnach, przy czym w pierwszej kolumnie znajduje się indeks podstawowy, zawierający kod PDB białka oraz numer łańcucha. Takie rozwiązanie wydaje się optymalne, gdyż w tabeli mogą być przedstawione wyniki dla różnych łańcuchów tego samego białka, ale nigdy nie może się powtórzyć dwa razy ten sam łańcuch. Dane z tabeli I są wyświetlane na stronie *BSDB* po wyborze kodu białka.

Tabela II zwiera pliki wynikowe (*.out), konformacje białek otrzymane podczas rozciągania (*.sav) i wykresy sił (*.gif). Dane te nie są prezentowane na stronie z wyjątkiem wykresu siły, jednakże użytkownik może je pobrać poprzez stronę internetową bazy.

Tabela III została stworzona jako element pomocniczy do symulacji nowych białek nie umieszczonych w tabeli I podczas pierwszego jej załadowania. Przechowuje ona pliki białek wraz z informacją o parametrach rozciągania, które czekają na wykonanie nowej symulacji. Dane pozostają w tabeli do momentu ukończenia symulacji, a następnie są usuwane.

Omówione powyżej tabele zostały dodatkowo uzupełnione o informacje na temat struktury białka zgodnie z klasyfikacją *CATH* dostępną pod adresem [31]. Posiadanie tej informacji umożliwi użytkownikowi porównanie rozkładu wartości siły w zależności od struktury białka.

Aktualizacja bazy

Dużym atutem bazy jest możliwość jej samoaktualizacji, a także aktualizacji przez administratorów oraz użytkowników, po wcześniejszym sprawdzeniu wiarygodności wyników.

Baza danych jest aktualizowana na dwa sposoby. Pierwszy polega na dodawaniu nowych wyników uzyskanych eksperymentalnie lub z pełnoatomowych symulacji wraz z publikacjami, w których się one ukazały. Odbywa się to z poziomu strony internetowej poprzez odpowiednią funkcję PHP oraz pocztę elektroniczną wysyłaną bezpośrednio do administratora bazy. Nie istnieje możliwość umieszczenia tego rodzaju danych w bazie bezpośrednio przez użytkowników, ze względu na zapewnienie wiarygodności wyników jak i wykluczenie umieszczenia tych samych rezultatów pod różnymi rekordami. W chwili obecnej, oprócz danych z własnych obliczeń, baza zawiera wyniki z symulacji pełnoatomowych **Survey of all–atom–obtained results** i otrzymane w eksperymentach **Survey of experimental results**. Informacje te są przedstawione w postaci tabel wraz z odnośnikami do prac, z których pochodzą.

Należy zauważyć, iż baza zachęca użytkowników do wysyłanie wszelkich sugestii dotyczących aktualizacji danych, nowych sposobów rozciągania białek i także procedur związanych z obsługą baz danych, gdyż ma to na celu zapewnienie możliwie najbardziej komfortowego i wydajnego sposobu korzystania z bazy.

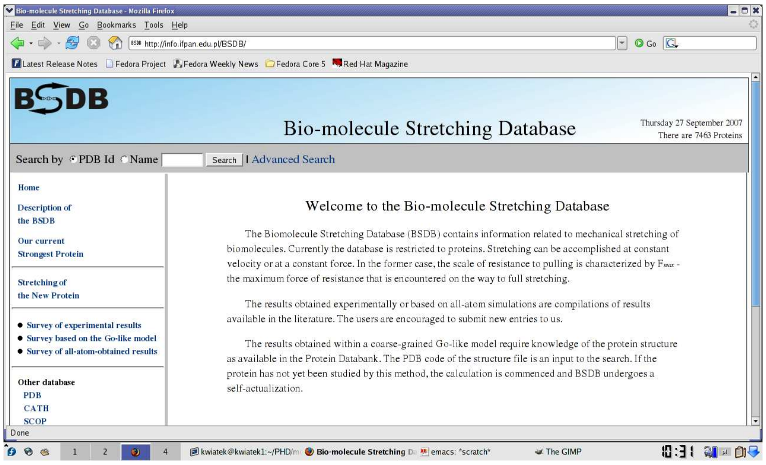
Drugi sposób to aktualizacja bazy po każdorazowym wykonywaniu przez użytkowników nowych symulacji za pomocą dołączonego oprogramowania. Proces ten jest w pełni zautomatyzowany. Wszystkie parametry symulacji użytkownik wybiera na stronie internetowej. Schemat uruchamiania nowej symulacji jest podany w paragrafie 8.3. Do tej aktualizacji jest używana tabela III. Proces jest uruchomiany poprzez odpowiedni program napisany w języku PHP, który pobiera odpowiednie pliki z bazy i przetwarza je dalej za pomocą skryptów tak, aby były one zrozumiałe dla programów wykonujących symulacje. Ten sam program sprawdza też pojawienie się na serwerze plików wynikowych symulacji. W przypadku znalezienia takiego pliku, program uzupełnia bazę o nowe dane i czyści tymczasową tabelę III. Oczekujące białka wraz z odpowiednim programem są automatycznie co godzinę wstawiane do systemu kolejkowego na serwerze obliczeniowym. W jednej chwili mogą być wykonywane cztery oddzielne symulacje.

Ponadto, w ściśle określonych odstępnach czasu (co pół roku), zaprogramowana jest samoaktualizacja bazy o wszystkie nowe białka które podczas tego okresu zostały zdeponowane w PDB.

Procedura pobierania danych z baz $PDB, CATH$, wypełnianie programów do symulacji odpowiednimi parametrami, uruchomienie systemu kolejkowego, obróbka wyników i wstawianie ich do tabel lub wysyłanie do użytkownika jest zautomatyzowane i kontrolowane za pomocą skryptów, łączących różnego rodzaju języki programowania (MySQL, PHP, fortran, Smongo [175, 176]).

8.2 Przeglądanie bazy danych — strona internetowa

Baza danych jest integralną częścią strony internetowej o nazwie *Bio-molecule Stretching DataBase* (*BSDB*). Jej głównym zadaniem jest w przejrzysty i szybki sposób udostępniać informację o własnościach mechanicznych białek i umożliwić rozbudowę bazy. Strona główna jest pokazana na rysunku 8.1 i została napisana w języku HTML 4.01 [176] do którego są dołączone skrypty w języku Java [177]. Konstrukcja strony jest przejrzysta, tak więc poniżej opiszemy tylko dwa najważniejsze jej elementy.

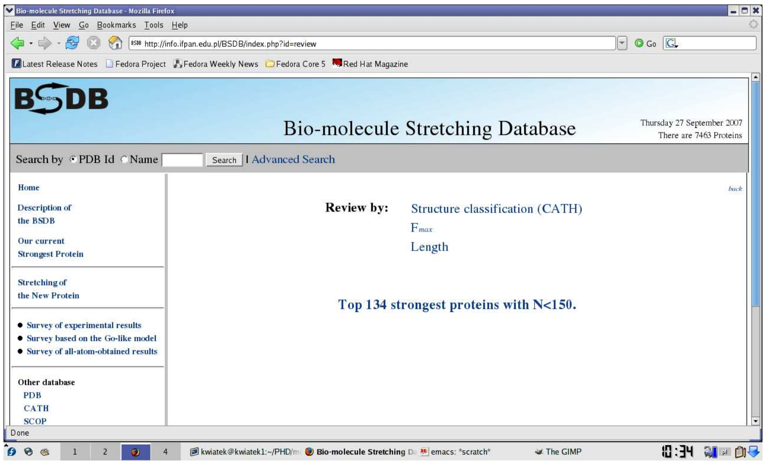


Rysunek 8.1: Strona tytułowa bazy *BSDB*.

Wyszukiwanie białek

Strona umożliwia wyszukiwanie białek na dwa sposoby. Pierwszy z nich polega na wpisaniu konkretnego kodu białka lub nazwy potocznej w pasku wyszukiwania.

Drugi sposób to wyszukiwanie zaawansowane poprzez funkcję *Advanced Search*, pokazaną na rysunku 8.2. Umożliwia ona wyszukanie pojedynczego białka lub zbioru białek odpowiadających zadanym kryteriom. Tak więc możliwe jest wyszukiwanie według zakresu wartości maksymalnej siły, klasyfikacji *CATH*, długości białka lub części kodu *PDB*. Takie wyszukiwanie umożliwia bezpośrednie sporządzanie statystyk o własnościach mechanicznych białek. Wyszukiwanie odbywa się poprzez zapytanie wysłane do bazy, zgodnie z formularzem wypełnionym przez użytkownika. W wyniku strona wyświetli pierwszych 25 białek spełniających dane zapytanie. Jeżeli ilość znalezionych białek jest większa niż 25, to na dole okna używanego pojawia się lista odnośników do kolejnych partii białek. Wybranie kolejnej partii białek spowoduje przesłanie przez adres URL informacji o elementach zapytania i numerze wybranej partii białek. Ponowne przesyłanie całego zapytania opóźniłoby proces wyszukiwania. Należy zauważyć, iż w każdej chwili użytkownik może obejrzeć wyniki dla konkretnego białka. Ten sposób wyszukiwania jest wygodny gdy użytkownik zbiera informację o białkach posiadających konkretną cechę lub gdy nie zna kodu białka.



Rysunek 8.2: Przegląd danych zawartych w *BSDB*.

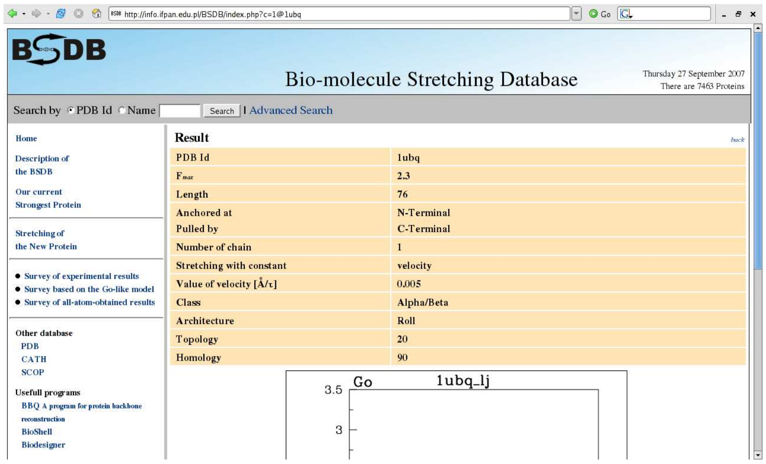
Przeglądanie bazy

Własności mechaniczne białek mogą być przeglądane według trzech kryteriów. Po pierwsze, według wartości maksymalnej siły. Wybranie tej opcji spowoduje wyświetlenie białek w kolejności od najmniejszej do największej wartości siły. Po drugie, w zależności od długości białka, czyli od najdłuższego do najkrótszego wraz z wartością siły. Po trzecie, zgodnie z przyporządkowaną klasą według klasyfikacji *CATH*. Wybór tej opcji spowoduje wyświetlenie listy wierszy, gdzie każdy wiersze odpowiada kolejnemu białku wraz z jego klasyfikacją *CATH* i wartością maksymalnej siły. Wiersze są ułożone w kolejności od najmniejszej do największej numeracji klasy białek.

Szczegółowe wyniki

Charakterystyka własności mechanicznych dla konkretnego białka jest dostępna z wielu podstron poprzez kliknięcie na niebieski odnośnik w postaci kodu *PDB* białka.

Informacje o białku są przedstawione w postaci tabeli pokazanej na rysunku 8.3 wraz z wykresem siły w funkcji przemieszczenia $F(d)$. Pod wykresem są umieszczone odnośniki umożliwiające pobranie plików z wynikami symulacji.



Rysunek 8.3: Schemat przedstawienia danych szczegółowych na stronie internetowej *BSDB*.

8.3 Serwer do rozciągania białek

Serwer do rozciągania białek jest integralną częścią strony internetowej do uzupełniania bazy *BSDB* o nowe wyniki symulacji. Znajduje się on w Instytucie Fizyki Polskiej Akademii Nauk i dysponuje narzędziami numerycznymi do rozciągania biomolekuł w modzie ze stałą prędkością lub stała siłą. Serwer umożliwia rozciąganie białek nie zawartych w *BSDB* lub kolejnych łańcuchów białek już tam zdeponowanych. Symulacja rozciągania białek jest wykonywana przy użyciu dynamiki molekularnej dla optymalnej wersji modelu Go $\{6-12, C, M3, E^o\}$, w której konkretne parametry symulacji (np. prędkość ciągnięcia) mogą być dobrane przez użytkownika. W przyszłości planujemy powiększyć oprogramowania o rozciąganie tandemu białek, kompleksów, struktury DNA i białek wraz z DNA. Poniżej przedstawiamy procedurę według której są wykonywane symulacje rozciągania nowych białek.

Uruchomienie serwera

Procedurę do rozciągania białek użtkownik może uruchomić korzystając ze strony *BSDB* poprzez wybranie opcji **Stretching of a New Protein**, a następnie wpisanie kodu PDB badanego białka i określenie charakteru symulacji (rozciąganie ze stałą prędkością lub siłą). W wyniku tego serwer pobiera z bazy PDB plik żądanego białka i sprawdza jego poprawność.

Weryfikacja pliku PDB

Pliki PDB są sprawdzane pod kilkoma względami. Po pierwsze, uniemożliwione jest rozciąganie białek w kompleksie (np. z DNA, RNA lub innymi białkami). Sprawdzana jest poprawność pliku uwzględniając listę synonimów i tablic zawierających wyjątki od standardowego zapisu współrzędnych atomów w białku. Wykrywane jest między innymi istnienie zaburzenia współzależnego (kilku współrzędnych dla tego samego atomu) i kompletność sekwencji aminokwasów wraz ze wszystkimi atomami z łańcucha głównego. W tych dwóch ostatnich przypadkach użytkownik może zrezygnować z symulacji lub samodzielnie naprawić plik (np. za pomocą programu BBQ [178]) i przesłać poprzez stronę poprawny plik w formacie PDB. Plik będzie ponownie sprawdzony przez serwer. Jednakże należy zauważyć, iż pliki PDB mogą zawierać błędy, których nie udało nam się przewidzieć. W tej sytuacji takie białko zostanie zwrócone tylko z komunikatem o błędzie. Na tym etapie sprawdzane jest też czy białko posiada węzeł – jeśli tak, to czas symulacji jest odpowiednio krótszy. Ostatnim elementem jest określenie zgodności formatu współrzędnych z oprogramowaniem do rozciągania. W przypadku pozytywnego zakończenia powyższych testów, serwer (oprogramowanie na nim zawarte) sprawdza czy dane białko jest już w bazie, a jeśli tak, to który jego łańcuch.

Określenie warunków rozciągania

W kolejnym kroku użytkownik jest proszony o wypełnienie tablicy określającej warunki symulacji, która zawiera: prędkość, siłę mostków siarczkowych, wybór łańcucha i kierunek rozciągania (numeru aminokwasu który będzie zakotwiczony, oraz za który będzie ciągnięte białko), częstość wypisywania konfiguracji białka podczas dynamiki. Warunkiem koniecznym przejścia tego etapu jest podanie poprawnego swojego adresu elektronicznego (e-mail).

Zatwierdzenie przyjętego zlecenia

Po wykonaniu powyższych czynności następuje sprawdzenie przez serwer sensowności wybranych parametrów i wyświetlenie ich na ekranie. Następnie użytkownik proszony jest o ich sprawdzenie. Po wciśnięciu przycisku **Next** następuje wstawienie białka do systemu kolejkowego i wysłanie do użytkownika pocztą elektroniczną informacji o przewidywanym czasie zakończenia symulacji.

Wyniki otrzymane z symulacji — aktualizacja bazy

W wyniku zakończenia symulacji w zależności od wyboru opcji do rozciągania powstaje kilka plików wynikowych. Jednakże w wyniku każdej symulacji dla rozciągania ze stałą prędkością baza $BSDB$ jest aktualizowana o: wartość maksymalnej siły wraz z jej położeniem, wykres siły w funkcji przemieszczenia (*.gif), scenariusze zrywania kontaktów (*.out), konformację białka podczas rozciągania (*.sav). Powyższe dane są wprowadzone do bazy po zatwierdzeniu ich przez program będący częścią oprogramowania, które dodatkowo weryfikuje poprawność wyników. Po wprowadzeniu danych do bazy stają się one także dostępne dla użytkownika, który może je bezpośrednio pobrać ze strony. Należy zauważyć, iż maksymalna wartość siły jest wyszukiwana numerycznie, tak więc w przypadku bardzo skomplikowanych wykresów $F(d)$ wartość F_{max} może być błędnie odczytana. Jednakże każdy użytkownik ma możliwość pobrania oryginalnych danych (*.out) i odczytania wartości siły samodzielnie. Natomiast pliki wynikowe (*.sav) z konformacją białka podczas rozciągania zawierają tylko położenie atomów C^{α}, zapisane w formacie PDB.

W przypadku zakończenia symulacji przed czasem lub uzyskania niepoprawnych wyników, użytkownik, który wprowadził dane białko, otrzymuje drogą elektroniczną komunikat o nieudanym rozciąganiu, a administrator informację o błędach.

Rozdział 9

Przewidywanie kolejności rozrywania kontaktów w modelu GNM

W rozdziale tym przewidujemy kolejność rozrywania kontaktów w strukturze białka używając modelu sieci gaussowskiej (GNM) na podstawie chwilowej jego konformacji. Przewidywanie opiera się na tezie, iż zerwanie kontaktu między parą aminokwasów i, j jest skorelowane z wystąpieniem minimum (bądź nawet zniknięciem) wartości średnich kwadratowych fluktuacji $\langle(\Delta R_i)^2\rangle$, $\langle(\Delta R_j)^2\rangle$ w modzie o największym okresie drgań. Wyniki takiego prognozowania porównujemy z diagramem scenariuszowym otrzymanym z symulacji rozciągania białka.

Szczegółowa analiza modów drgań w modelu GNM użyta do przewidywania zrywania kontaktów dla 27. domeny tytny i zielonego białka GFP, została opisana w pracy [V].

9.1 Wstęp

W rozdziałe 6 zajmowaliśmy się najbardziej powszechnym rodzajem dużych konformacyjnych zmian jakie występują w białkach, czyli rozwijaniem mechanicznym. Wspominaliśmy już, iż sam mechanizm rozwijania lub dużych zmian konformacyjnych może wynikać z różnych powodów i jest bardzo powszechny w procesach biologicznych. Mamy tutaj na myśli np. zjawisko allosterii [32], które polega między innymi na ruchach domen białkowych, częściowym rozwijaniu się i zmianach konformacyjnych jej elementów. Oba te procesy mogą posłużyć do wyjaśnienia np. przejść konformacyjnych w kinazach adentylate, które zostały opisane w pracach [179, 180]. Ponadto częściowe rozwijanie i ponowne zwijanie [181] (ang. proteinquakes) może opisywać także zmiany konformacyjne w miozynach. Należy także tutaj wymienić ruchy funkcjonalne w dużych strukturach biologicznych jak podjednostki rybozomowe [182, 183], białka opiekuńcze [184] i wiele innych. Poznanie i zrozumienie makroskopowych konformacyjnych zmian jest obecnie nieosiągalne dla pełnoatomowych dokładnych symulacji, ze względu na bardzo długi czas obliczeń. Pewnym ułatwieniem byłoby posiadanie chwilowych konformacji podczas zachodzących zmian, jednakże otrzymanie ich konwencjonalnymi technikami krystalograficznymi jest bardzo trudne. Tym niemniej okazuje się, iż eksperymentalnie obserwowane ruchy funkcyjne w takich układach jak rybosom czy białka opiekuńcze może być odtworzone przy użyciu modelu sieci gaussowskiej (GNM) [185] na podstawie stanu natywnego białka.

W rozdziale tym pokazujemy, iż lukę w zrozumieniu dużych konformacyjnych zmian można wyjaśnić poprzez połączenie modelu sieci gaussowskiej z konformacjami chwilowymi uzyskanymi z dynamiki molekularnej, np. podczas rozciągania białek w prostym modelu Go. Należy jednak zauważyć, iż zastosowanie samej stochastycznej dynamiki nie byłoby wystarczające, gdyż oferuje ona tylko pewną liczbę trajektorii a nie pełen zespół statystyczny, który jest konieczny do przewidzenia drgań termicznych. Natomiast model GNM jest w stanie – nawet na podstawnie jednej trajektorii rozciągania – przewidzieć ruch elementów struktury z drgań termicznych poszczególnych reziduów.

Przyjrzyjmy się przez chwilę warunkom stosowalności modelu sieci gaussowskiej. Wersja modelu GNM dla białek zakłada istnienie prostych harmonicznych parametrów sił do opisu drgań wokół położenia równowagi. Tak więc model GNM jest stosowany do analizy zachowania się struktury w jej energetycznym minimum, o którym zakłada się, iż jest jej stanem natywnym wyznaczonym krystalograficznie. Wydawałoby się, iż ten kwazi-statyczny, harmoniczny opis jest stosowalny jedynie blisko stanu natywnego i mógłby nie być poprawny w przypadku dużych zaburzeń konformacyjnych. Jednakże, po pierwsze, ostatnio pojawia się coraz więcej przekonywujących rezultatów, iż model GNM poprawnie przewiduje kierunek nawet dużych zmian konformacyjnych [186, 187]. Po drugie, etapy mechanicznego rozciągania białek mogą być zinterpretowane jako ruch z jednego lokalnego minimum do drugiego poprzez przekroczenie bariery energetycznej (w której pewne natywne kontakty muszą zostać zerwane) pomiędzy tymi dwoma konformacjami. Tak więc, białko odpowiada na zewnętrzną mechaniczną siłę poprzez ruch zgodnie z najniższą częstością, aż do momentu pewnego krytycznego położenia, które znajduje się w obszarze odpowiadającym naszemu założeniu o liniowej elastyczności układu. Wraz ze wzrostem amplitudy przemieszczenia, obszar nieliniowej odpowiedzi układu zostaje osiągnięty i w konsekwencji natywne kontakty są rozrywane. Innymi słowy, białko rozwija się lokalnie/częściowo, gdy elastyczne naprężenia są zbyt duże. Następnie zachodzi relaksacja białka do nowego lokalnego minimum, dla którego znowu założenia o modelu sieci guassowskiej są poprawne.

Na podstawie powyższych rozważań konstruujemy następującą hipotezę. Podczas rozciągania białka naprężenie pomiędzy danymi dwoma reziduami i, j będącymi w natywnym kontakcie powinno przekraczać pewną krytyczną progową wartość, i wtedy ich średnie kwadratowe fluktuacje $\langle(\Delta R_i)^2\rangle$ w jednym z modów o największym okresie drgań powinny zbiegać do zera. Natomiast średnie kwadratowe fluktuacje odległości $\langle(\Delta r_{ij})^2\rangle$ między natywną parą i, j powinny fluktuować ze stałą i małą wartością amplitudy, aż do zerwania kontaktu, widocznego jako nagły bardzo duży wzrost. Tak więc, aby przetestować tę tezę i pokazać, iż duże zmiany konformacyjne mogą być badane poprzez połączenie metody GNM z dynamiką molekularną w ramach modelu Go, dokonujemy następująceego teoretycznego eksperymentu.

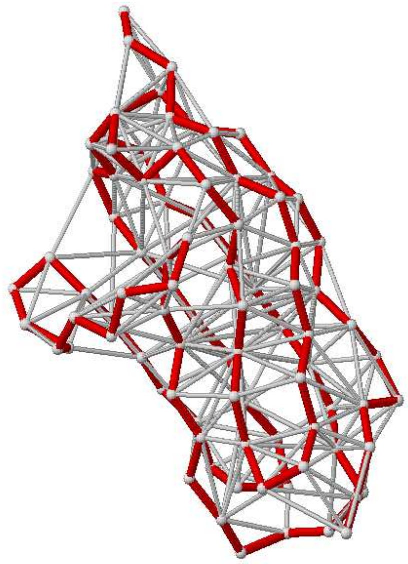
Wykonujemy pojedynczą symulację rozciągania białka ze stałą prędkością używając dynamiki molekularnej wraz z modelem Go o parametrach $\{6-12, C, M2, E^o\}$, opisanym w rozdziale 4. W wyniku dostajemy zbiór chwilowych konformacji dla kolejnych przemieszczeń wektora prędkości (kolejnych chwil czasu); najmniejsza różnica w czasie odpowiada przemieszczeniu się terminalnych atomów o wartość 0.5 Å podczas rozciągania. Wybieramy najbardziej interesujące chwilowe struktury białka i dokonujemy ich analizy wykorzystując model GNM. Analizujemy amplitudy fluktuacji reziduów w modach o największym okresie drgań pod kątem zidentyfikowania kontaktów, które pękną w następnych chwilach czasu. Każdą prognozę porównujemy z rzeczywistym mechanizmem zmian strukturalnych podczas rozciągania w modelu Go. Symulację rozciągania przeprowadzamy w stanie o zerowej temperaturze, aby wyeliminować dodatkowe fluktuacje i możliwość odtworzenia się natywnych i nienatywnych kontaktów [120]. Przyjęcie takich warunków zdecydowanie upraszcza analizę fluktuacji węgli C^α w modelu GNM. Należy jednak pamiętać, iż wzrost fluktuacji termicznych zawsze zmniejsza siłę oporu i wartość rozciągnięcia, dla którego pękają kontakty. Jednakże, zgodnie z naszymi wcześniejszymi wynikami z pracy [IV] i podrozdziału 6.2.4, zdarzają się białka, których sam sposób rozwijania zależy od temperatury. Do obecnego eksperymentu wykorzystaliśmy 27. domenę tytyny (1tit), zielone białko GFP(1emb) i 10 innych (1pga, 1crn, 1ubq, 1aj3, 1ula, 1aqb, 5pti, 1dz3, 1ido i 5rsa), dla których główna droga rozwijania nie zależy od wyboru temperatury.

9.2 Model sieci gaussowskiej

Model sieci gaussowskiej jest kolejnym przykładem modelu gruboziarnistego. Pierwsza jego wersja powstała w ramach rozwijania teorii o przypadkowych elastycznych sieciach polimerów [188, 189] do opisu fluktuacji połączeń i węzłów wewnątrz sieci. Następnie w roku 1997 model GNM został przystosowany do opisu fluktuacji w białku przez Bahar i Ermana [190, 14], którzy opisali drgania atomów poprzez pojedyncze harmoniczne siły zgodnie z założeniem zaproponowanym przez Tirion [15].

Konstrukcja modelu sieci gaussowskiej w białku

W modelu tym, podobnie jak w przypadku gruboziarnistego modelu Go, molekułę białka przedstawia się poprzez węgle C^{α}. Następnie wyznacza się mapę kontaktów, określającą które atomy C^{α} są w natywnym kontakcie. Najczęściej jest tu stosowane kryterium jednorodnego odcięcia R_c=7.5Å [185, 190, 14, 191], które opisaliśmy w podrozdziale 3.1.3 i powstałą z niego mapę oznaczyliśmy M_c. Jednakże w tym przypadku sieć kontaktów uwzględnia także wszystkie wiązania wzdłuż łańcucha, które znajdą się w zasięgu danego promienia od rozważanego atomu C_i^{α}. W wyniku otrzymujemy sieć połączeń, której przykładowa postać jest pokazana dla tytyny na rysunku 9.1.



Rysunek 9.1: Reprezentacja struktury białka (łańcuch główny oznaczony jest czerwonym kolorem) poprzez sieć połączeń używaną do konstrukcji modelu sieci gaussowskiej (szare połączenia).

Następnie zakładamy, iż atomy w sieci są połączone identycznymi elastycznymi sprężynami. Prowadzi to do powstania elastycznej sieci, reprezentującej strukturę białka w stanie natywnym i odpowiada to sieci przypadkowego polimeru, z tym wyjątkiem, że polimer tworzyłby bardziej obszerną sieć połączeń.

Niech $\mathbf{R}_i$ oznacza położenie węgla i, oraz $\Delta\mathbf{R}_i$ jego fluktuacje. Wektor łączący atom C_i^{α} z C_j^{α} oznaczmy jako $\mathbf{R}_{ij} = \mathbf{R}_j - \mathbf{R}_i$, a jego długość poprzez $r_{ij} = |\mathbf{R}_j - \mathbf{R}_i|$. Ponadto, niech $\Delta\mathbf{R}_{ij}$ określa jednoczesne przemieszczenie się $\mathbf{R}_{ij}$ od wartości średniej danej przez $\mathbf{R}_{ij}^o$. Dodatkowo wprowadzamy Δr_{ij}=$\sqrt{(\Delta\mathbf{R}_{ij})^2}$, gdzie $(\Delta\mathbf{R}_{ij})^2 = (\Delta\mathbf{R}_{ij}^T \cdot \Delta\mathbf{R}_{ij})$. Zgodnie z powyższym, ruch atomów w elastycznej sieci dany jest potencjałem harmonicznym

$$V(r_{ij}) = \frac{1}{2}g(\mathbf{R}_{ij} - \mathbf{R}_{ij}^0)^2 H(R_c - r_{ij}) = \frac{1}{2}g(\Delta\mathbf{R}_{ij})^2 H(R_c - r_{ij}), \tag{9.1}$$

gdzie g jest uniwersalną jednorodną stałą sprężystości, a $H(x)$ funkcją Heaviside'a

$$H(x) = \begin{cases} 1 & \text{gdy } x \geq 0 \\ 0 & \text{gdy } x < 0 \end{cases} \tag{9.2}$$

Wprowadźmy ponadto N-wymiarowy wektor fluktuacji

$$\{\Delta\mathbf{R}\} = \begin{bmatrix} \Delta\mathbf{R}_1 \\ \Delta\mathbf{R}_2 \\ \vdots \\ \Delta\mathbf{R}_N \end{bmatrix}$$

o składowych $\Delta\mathbf{R}_i$ dla wszystkich N atomów. Całkowita energia sieci składającej się z N węzłów jest sumą energii danych równaniem 9.1 dla wszystkich oddziałujących par węzłów (atomów), i jest dana wyrażeniem

$$V_{cal} = \frac{g}{2}\{\Delta\mathbf{R}^T\}K\{\Delta\mathbf{R}\}, \tag{9.3}$$

gdzie K reprezentuje macierz Kirchoff'a natywnych kontaktów o wymiarze $N \times N$ z elementami pozadiagonalnymi równymi -1, jeśli atomy C_i^α i C_j^α są w kontakcie, i zero w przeciwnym wypadku. Elementy diagonalne określają sumę elementów pozadiagonalnych w i-tym wierszu wzięte ze znakiem przeciwnym.

Średnie zmiany w położeniach reziduów w białku można wyrazić poprzez funkcję korelacji $\langle \Delta\mathbf{R}_i^T \cdot \Delta\mathbf{R}_j \rangle$ pomiędzy przemieszczeniami pary reziduów i, j lub poprzez średnie kwadratowe fluktuacje $\langle (\Delta\mathbf{R}_i)^2 \rangle = \langle \Delta\mathbf{R}_i^T \cdot \Delta\mathbf{R}_i \rangle$ dla pojedynczego reziduum i. Dla wszystkich par i, j zachodzi

$$\langle \Delta\mathbf{R}_i \cdot \Delta\mathbf{R}_j \rangle = \frac{\int (\Delta\mathbf{R}_i^T \cdot \Delta\mathbf{R}_j) exp(-V_{cal}/k_B T) d\{\Delta\mathbf{R}\}}{\int exp(-V_{cal}/k_B T) d\{\Delta\mathbf{R}\}}.$$

Powyższe dość skomplikowane wyrażenie można następnie ściśle przekształcić, zgodnie z pracami [188, 189], do dużo prostszej postaci

$$\langle \Delta\mathbf{R}_i \cdot \Delta\mathbf{R}_j \rangle = \frac{3k_B T}{2g}(K^{-1})_{ij}, \tag{9.4}$$

gdzie $(K^{-1})_{ij}$ jest elementem (i, j) odwrotności macierzy kontatków K. Średnie kwadratowe fluktuacje $\langle (\Delta\mathbf{R}_i)^2 \rangle$ i-tego reziduum w białku dane są poprzez i-ty diagonalny element K^{-1} w równaniu (9.4). Natomiast średnie kwadratowe fluktuacje odległości $r_{ij} = |\mathbf{R}_i - \mathbf{R}_j|$ pomiędzy dwoma atomami C_i^α i C_j^α można wyrazić następująco

$$\langle(\Delta r_{ij})^2\rangle = \langle(\Delta\mathbf{R}_i - \Delta\mathbf{R}_j)^2\rangle = \frac{3k_BT}{2g}\big[(K^{-1})_{ii} + (K^{-1})_{jj} - 2(K^{-1})_{ij}\big]. \tag{9.5}$$

Powyższe przekształcenia są możliwe jeśli zauważy się, iż macierz połączeń została zdefiniowana tak, że suma elementów w każdym wierszu (lub kolumnie) wynosi zero. W wyniku tego wyznacznik K jest równy 0 i otrzymujemy macierz osobliwą, dla której może być wyliczona tylko tzw. macierz pseudoodwrotna, poprzez rozkład na osobliwe wartości własne (ang. singular value decomposition method). Macierz pseudooodwrotną K możemy zapisać jako K^{-1}=$U(\Lambda^{-1})U^T$, gdzie U jest macierzą złożoną z wektorów własnych $u_i, i = 1, \ldots, N$ macierzy K, a Λ jest diagonalną macierzą wartości własnych K. Wszystkie wartości własne λ_i dla macierzy K są dodatnie [188].

Średnie kwadratowe fluktuacje położenia dla każdego C^α dane równaniem (9.4) są odpowiednikiem czynnika termicznego Debaya-Wallera, który jest mierzalny eksperymentalnie poprzez rozpraszanie promieniowania X w strukturach krystalograficznych białek i zdeponowany w pliku PDB danego białka jako termiczny tzw. czynnik B. Zależność pomiędzy czynnikami B_i i średnimi kwadratowymi fluktuacjami atomu i dana jest równaniem

$$B_i = \frac{8\pi^2}{3}\langle(\Delta R_i)^2\rangle. \tag{9.6}$$

Powyższa zależność jest używana do wyznaczenia stałej sprężystości g, która jest jawnie widoczna w równaniu 9.4. Jest ona traktowana jako dodatkowy parametr, dobierany tak, aby otrzymać najlepsze dopasowanie danych teoretycznych dla stanu natywnego do eksperymentalnie mierzonego czynnika B. Wartość bewzględna g nie zmienia faktycznego rozkładu fluktuacji reziduów, a jedynie je jednorodnie przeskalowuje. Rozkład amplitudy drgań wzdłuż sekwencji aminokwasów w białku uzyskany w modelu GNM zazwyczaj wykazuje bardzo dobrą zgodność z danymi doświadczalnymi [192]. Wyniki te potwierdza kilka innych wcześniejszych prac nad układami biologicznymi [193, 194] używających modelu GNM, gdzie pokazano, iż dynamika białka jest nieczuła na detale modelu i dobór parametrów, a najbardziej istotne mody o najdłuższych okresach drgań (np. te które odpowiadają sztywnym powolnym ruchom całych podjednostek) zależą przede wszystkim od kształtu białka [195]. Poprawność i szerokie zastosowanie modelu GNM zostało ostatnio dodatkowo potwierdzone serią eksperymentów, takich jak te z wymianą wodoru na deuter [196], czy też wynikami z pomiarów NMR pokazującymi kierunkowe ruchy indywidualnych wiązań peptydowych i poprzez zaburzenia widoczne w wielokrotnych strukturach [185, 186, 197, 198].

Rozkład na wartości własne — najszybsze i najwolniejsze mody drgań

Konstrukcja modelu GNM, jak i podane powyżej wyniki prac [195, 196, 197, 198, 185, 186] pokazują, że białko można uważać za ciało sprężyste, i na tej podstawie w szczególności scharakteryzować funkcjonalne ruchy w układach biologicznych. Uzyskanie ich w pełnoatomowych symulacjach jest zbyt czasochłonne, natomiast poniżej pokazujemy, że mogą być one wyznaczone poprzez kombinacje modelu GNM i dynamiki molekularnej wraz z modelem gruboziarnistym. Poniżej zwięźle pokazujemy technikę rozkładu na wartości własne i wyznaczanie z niej kolejnych modów drgań.

Macierz kontaktów K^{-1} możemy zapisać jako sumę wkładów od wszystkich indywidualnych modów drgań [14] wyrażanych przez wektor własny $\mathbf{u}_k$ opisujący k-ty mod normalny

$$K^{-1} = \sum_k{}' \lambda_k^{-1} \mathbf{u}_k \mathbf{u}_k^T. \tag{9.7}$$

Znak prim przy sumie oznacza, że zerowe wartości własne – mające interpretację sztywnych ruchów centrum masy układu – są wyeliminowane z sumowania. Natomiast i-ty składnik wektora własnego określa amplitudę ruchów fluktuacyjnych i-tego atomu C^α w białku wywieraną w k-tym modzie. Jeśli uporządkujemy wartości własne od najmniejszej do największej, to największy wkład do równania (9.7) będzie mieć rozkład pierwszych niezerowych wartości własnych λ_k. Wartości te odpowiadają najmniejszym niezerowym częstościom drgań. Z drugiej strony te najwolniejsze mody odgrywają najważniejszą rolę w badaniu zmian struktury białka, gdyż ich wkład do średnich kwadratowych fluktuacji skaluje się jak λ_k^{-1}. Potwierdzeniem tego są wyniki prac z ostatnich lat, które pokazują, iż większość ważnych/funkcjonalnych ruchów w białkach [199, 200, 201] oraz ruchy dużych biologicznych struktur (takich jak rybozom [182, 183]), które są powiązane z ich biologiczną funkcją, mogą być z powodzeniem zidentyfikowane poprzez tylko kilka najmniejszych niezerowych częstości. Kolejne ukazujące się prace [187] bazujące na tym samym rozumowaniu dodatkowo poszerzają zakres stosowalności modelu, m.in. do badania zmiany konformacji białka pomiędzy strukturą „otwartą" i „zamkniętą", lub zamiany domen w białkach (ang. domain swapping). Z drugiej strony najszybsze mody wskazują, które pojedyncze atomy w strukturze wykonują najszybsze oscylacje.

9.3 Wyniki — Analiza rozciągania tytyny w modelu GNM

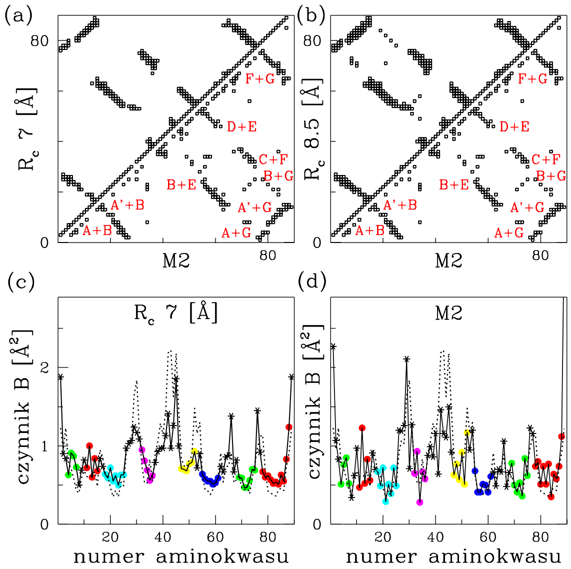
Przypomnijmy, iż 27. domena tytny o kodzie 1tit jest białkiem z klasy β, składającym się z ośmiu nici β oznaczonych A(4-7), A'(11-15), B(18-25), C(32-36), D(47-52), E(56-61), F(69-75) i G(78-88). Jej struktura, wraz z oznaczeniami kolejnych nici β, jest pokazana na rysunku 4.1 w podrozdziale 4.1.3.

9.3.1 Porównanie mapy kontaktów w modelu Go i GNM

Zgodnie z rozważaniami z podrozdziału 3.1.3 istnieje wiele sposóbów wyznaczenia mapy kontaktów. W przypadku modelu Go pokazaliśmy, iż najlepszą zgodność z doświadczeniem otrzymuje się dla mapy $M3$ wyznaczonej na podstawie przykrywań się powiększonych sfer van der Waalsa. Natomiast wcześniejsze prace [192] nad modelem GNM pokazują, iż najlepszą zgodność z doświadczeniem daje technika jednorodnego odcięcia o R_c=7.3Å, którą oznaczamy M_c. W tej samej pracy pokazano jednakże, iż dopasowanie wartości R_c w dużym stopniu zależy od kształtu i rozmiaru białka. Taki stan rzeczy wymaga znalezienia wspólnej optymalnej mapy, która byłaby najlepsza dla rozważanego przez nas białka 1tit.

W tym celu po pierwsze porównaliśmy mapę kontaktów M_c dla wartości promienia odcięcia od 6.5Å co 0.5Å do 17Å z mapą pełnoatomową $M2$. Po drugie, dla każdej uzyskanej mapy kontaktów policzyliśmy średnie kwadratowe odchylenie dla każdego atomu C_i^α w białku używając modelu GNM i następnie porównaliśmy je z doświadczalnie wyznaczonymi czynnikami B_i. Rysunek 9.2 pokazuje podobieństwo między mapami natywnych kontaktów uzyskanych technikami M_c7, $M_c8.5$ (powyżej diagonali) i $M2$ (poniżej diagonali) oraz na ich podstawie otrzymane $\langle(\Delta\mathbf{R}_i)^2\rangle$ w ramach modelu GNM.

Otrzymaliśmy, iż dla wartości R_c=8.5Å w technice M_c zostały odtworzone wszystkie natywne kontakty uzyskane w technice $M2$, jednakże kosztem ogromnej ilości nierzeczywistych dodatkowych kontaktów, zwłaszcza wokół łańcucha białka. W tym przypadku korelacja między teoretycznymi czynnikmi B_i i doświadczalnymi wynosi 0.71%. Najlepszą zgodność B_i otrzymujemy dla wartości R_c=7 Å i wynosi ona 0.77%. Należy jednak zauważyć, iż dla R_c=7Å zgodność z mapą pełnoatomową wynosiłaby 96% gdyby rozważyć tylko kontakty natywne tworzone przez wstęgi β. Te kontakty są najważniejsze i w modelu GNM, gdzie w stanie natywnym charakteryzują się one małymi drganiami, i dla modelu Go, gdzie od nich zależy scenariusz rozwijania się białka. Tak więc, skoro różnica w mapach kontaktów M_c7 i $M2$ jest nieznaczna a zastosowanie



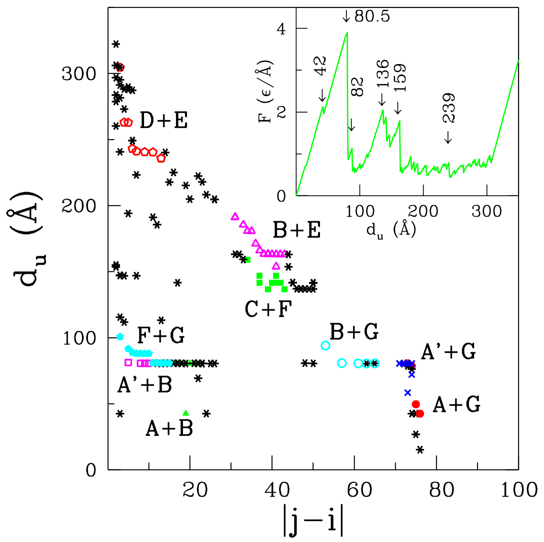
Rysunek 9.2: Górne panele: mapy natywnych kontaktów otrzymane techniką M_c dla odcięcia R_c równego 7 lub 8.5 Å(powyżej diagonali), i techniką $M2$ (poniżej diagonali). Dolne panele: temperaturowy czynnik B (średnie kwadratowe odchylenie) wyznaczony na podstawie mapy M_c7 i mapy $M2$ w modelu GNM. Symbole literowe oznaczają wstęgi β, pokazane na rysunku 4.1.

innej techniki wyznaczenia mapy w modelu GNM lub Go obniża zdecydowanie ich zgodność z danymi doświadczalnymi, dochodzimy do wniosku, iż optymalnym wyborem jest zastosowanie dwóch różnych map: mapy M_c7 dla modelu GNM i mapy $M2$ dla modelu Go. W powyższych obliczeniach korzystaliśmy z g=3.8 kcal mol^{-1} Å^{-2}, który dawał najlepszą zgodność danych teoretycznych z eksperymentalnie wyznaczonym czynnikiem B_i. Przypomnijmy, że w podrozdziale 4.2 pokazaliśmy, iż dla modelu Go najlepszą zgodność z doświadczeniem otrzymuje się używając mapy $M3$, jednakże w przypadku białka 1tit różnica między $M3$ i $M2$ jest niewielka.

9.3.2 Mody o najmniejszych niezerowych częstościach

Rozciąganie tytyny

Wyniki symulacji rozciągania 1tit w temperaturze $\tilde{T} = 0$ różnią się od wyników w temperaturze pokojowej pojawieniem się małego maksimum na wzrastającej krzywej siły i innymi wartościami odległościami rozrywania kontaktów. Jednakże sposób rozwijania jest bardzo podobny. Wykres siły w funkcji przemieszczenia $F(d)$ i scenariusz rozrywania kontaktów dla rozciągania 1tit dla $\tilde{T}$=0 jest pokazany na rysunku 9.3, jednakże nie omawiamy go, odsyłając czytelnika do rozdziału 4.1.3, gdzie znajduje się dokładny opis symulacji dla $\tilde{T}$=0.3.

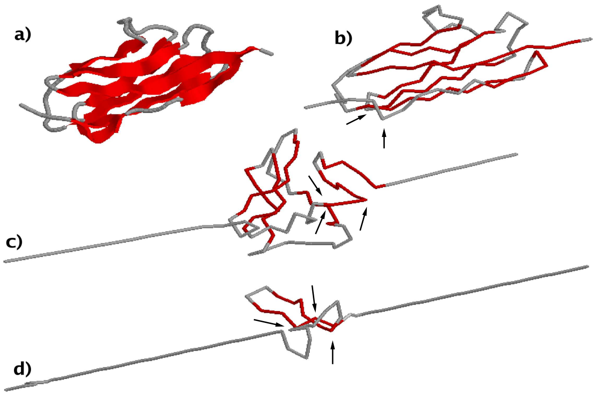


Rysunek 9.3: Scenariusz pękania kontaktów dla białka 1tit (dla $\tilde{T}$=0) w funkcji odległości między końcami białka d_u. Symbole literowe wskazują w jakich strukturach drugorzędowych pękają kontakty. Gwiazdki oznaczają kontakty poza strukturą drugorzędową. Wszystkie oznaczenia są takie same jak na rysunku 4.1. Panel wewnętrzny przedstawia wykres siły w funkcji przemieszczenia, strzałki wskazują wybrane konformacje białek użyte do analizy modów drgań.

Obecnie zajmiemy się omówieniem wyników przewidywania rozrywania kontaktów na podstawie pięciu najbardziej charakterystycznych konformacji tytyny wskazanych strzałkami na rysunku 9.3. Są to: pierwsze naderwanie – stan przejściowy (d_u=42Å), stan tuż przed maksimum największego oporu (d_u=80.5Å), konformacja struktury pomiędzy maksimami (d_u=82Å), drugie maksimum (d_u=159Å) i stan odpowiadający rozerwaniu ostatniej spinki β D+E (d_u=239Å). Konformacja białka w tych chwilach czasu wraz z zaznaczonymi miejscami w strukturze, które według prognozy mają być rozerwane, jest pokazana na rysunku 9.4. W poniższych wynikach używamy stałej sprężystości g=1 kcal mol^{-1}Å^{-2}, tak więc absolutna amplituda ruchów w każdym modzie jest arbitralna.

Pierwsze naderwanie (d_u=42Å)

Rysunek 9.5 przedstawia średnie kwadratowe odchylenie $(\Delta\mathbf{R}_i)^2$ jako funkcję położenia aminokwasów wzdłuż sekwencji w modzie o najmniejszej niezerowej częstości, policzone na podstawie równań (9.4) i (9.7). Otrzymaliśmy, iż amplituda fluktuacji dla kilku reziduów jest znacznie większa niż w pozostałej części struktury. Największe drgania są widoczne dla pary natywnych aminokwasów 1-77(G) i 14(A')-89(G) co wskazuje, iż kontakt ten został zerwany. Przewidywania te zgadzają się z diagramem scenariuszowym 9.3, gdzie dla d_u=40 kontakt został rozerwany. Kolejne duże drganie widzimy dla reziduów 25, 26, 64, 65, 66, 67, 75 i 76, jednakże ich pochodzenie



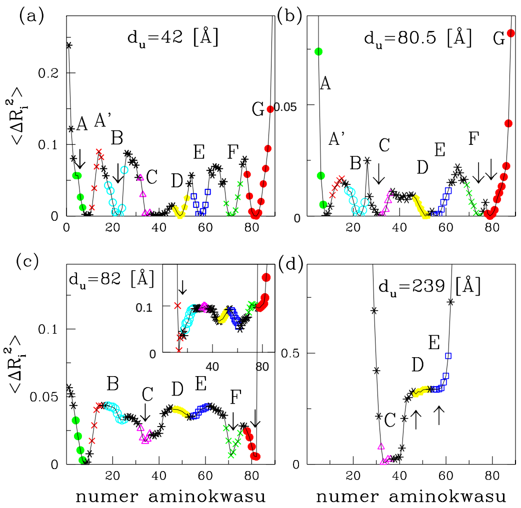
Rysunek 9.4: a) Wstążkowa reprezentacja struktury białka w stanie natywnym. Panele b), c) i d) pokazują konformacje struktury białka (reprezentacja łańcuchowa) dla wybranych etapów rozciągania 1tit ze stałą prędkością, gdzie b) stan przejściowy (d_u = 42 Å), c) konformacja po rozerwaniu mechanicznego imadła (d_u=80.5 Å), d) konformacja przed rozerwaniem ostatniej spinki β (d_u=239 Å). Strzałki wskazują prognozowane w modelu GNM miejsca rozerwania kontatków.

jest inne. Aminokwasy te należą do nieustrukturyzowanej części białka lub nie tworzą kontaktów z pozostałą częścią struktury, w wyniku czego charakteryzują się dużą wewnętrzną swobodą niezależnie od etapu rozciągania. Mod o najmniejszej niezerowej częstości pokazuje także kilka charakterystycznych minimów $\langle(\Delta\mathbf{R}_i)^2\rangle \simeq 0$, gdzie naprężenie w strukturze jest największe. Najmniejsze fluktuacje wykazuje atom 23 z nici B, który jest w kontakcie wodorowym z atomem A(4). Zgodnie z przewidywaniami opartymi na modelu GNM kontakt ten zostanie zerwany w następnym kroku (rysunek 9.3). Niezerowe fluktuacje reziduum 4 wynikają ze swobodnych ruchów jego otoczenia. Minima wewnątrz spinek β G+F i B+E wynikają z narastającej w tej chwili czasu siły ścinającej pomiędzy tymi dwoma płaszczyznami ciągniętymi w przeciwnych kierunkach.

Struktura tuż przed największą zmianą konformacyjną w tytynie (d_u=80.5 Å)

Konformacja ta według symulacji rozciągania prowadzi do rozerwania kontaktów między spinkami β A+G, A'+G, A+B, B+G i F+G. Rozkład wartości $\langle(\Delta\mathbf{R}_i)^2\rangle$ dla aminokwasów z konformacji białka w d_u = 80.5 Å w najwolniejszym modzie jest pokazany na rysunku 9.5(b). Po charakterze fluktuacji widzimy, że nić A nie tworzy już kontaktów z resztą białka, natomiast szczegółowa analiza A' i G pokazuje bimodalny rozkład fluktuacji. Część reziduów ze wstęgi G charakteryzuje się dużymi odchyleniami od położenia równowagi, natomiast druga część zerowymi fluktuacjami, które to określają położenie miejsc zawiasowych. Miejsca zawiasowe wynikają z narastających bardzo dużych naprężeń w nierozerwalnych wiązaniach wodorowych tworzonych przez aminokwasy 78, 79, 80 i 81 ze wstęgi G i aminokwasami z 73, 74 i 75 ze wstęgi F. Zgodnie z przyjętą przez nas hipotezą kontakty te powinny zostać rozerwane jako pierwsze w kolejnej chwili czasu. Przewidywania te są zgodne z diagramem scenariuszowym dla rozwijania 1tit, gdzie po pierwsze dla d_u=80.5Å istnieje połowa natywnych kontaktów między F+G (fluktuacje dla swobodnych reziduów są znaczące), a po drugie dla d_u=81Å wymienione powyżej kontakty z miejsc zawiasowych zostają rozerwane.

Podobne własności obserwujemy dla reziduów dla nici A i B pomiędzy którymi kontakty właśnie zostały rozerwane. Z kształtu fluktuacji tych reziduów widać, iż wstęga A jest już całkowicie swobodna (fluktuacje atomów 4-7 są tak duże, iż nie mieszczą się w skali), natomiast rezidua B18 i B25 są nadal nieruchome gdyż tworzą kontakt ze wstęgą E. Ponadto analiza fluktuacji w trzecim modzie pokazuje wyraźnie, iż całe wstęgi A i A' i część wstęg G i B nie tworzy już kontaktów, a ich amplitudy drgań są znacznie większe niż pozostałych reziduów. Drugi mod w tym przypadku nie dawał tak jednoznacznej odpowiedzi jak trzeci mod.



Rysunek 9.5: Średnie kwadratowe fluktuacje $\langle(\Delta\mathbf{R}_i)^2\rangle$ aminokwasów w modach o najmniejszych niezerowych częstościach uzyskane z modelu GNM na podstawie wyróżnionych konformacji d_u podczas rozciągania 1tit. Symbole literowe i kolorowe oznaczenia są takie same jak na rysunku 9.3 i odnoszą się do aminokwasów tworzących kontakty natywne przez struktury drugorzędowe. Strzałki wskazują miejsca gdzie fluktuacje są najmniejsze. Panel a) odpowiada stanowi przejściowemu (d_u=42Å); b) konformacja w trakcie zrywania mechanicznego imadła (d_u = 80.5Å); c) struktura białka po zerwaniu imadła (d_u = 82Å): w pierwszym modzie wewnętrzny panel i w drugim modzie główny panel; d) struktura przed zerwaniem ostatniej spinki β.

Konformacja struktury pomiędzy maksimami (d_u=82Å)

Po rozerwaniu mechanicznego imadła niemalże jednocześnie zostaje rozerwanych około 50% natywnych kontaktów, aż do momentu napotkania na kolejny duży opór w białku. Ten duży opór rozumiemy jako stan metastabilny białka pochodzący od prostopadłego ułożenia do kierunku rozciągania wiązań wodorowych pomiędzy spinkami β C+F, B+E i D+E. Ta konformacja doprowadzi do mechanizmu ścinającego tych β wstęg zakończonego powstaniem drugiego maksimum na krzywej $F(d)$. Od odległości d_u=82Å do d_u=100Å obserwujemy sukcesywne pękanie pozostałych kontaktów dla F+G. Dla rozciągnięcia d_u>100 aż do d_u=120Å obserwujemy rozerwanie tylko trzech kontaktów w elementach struktury drugorzędowej.

Rysunek 9.5(c) pokazuje wkład do średnich kwadratowych fluktuacji reziduów od pierwszego modu (panel wewnętrzny) i od drugiego modu dla d_u=82Å. Dla 98% struktury obserwujemy bardzo wysoki poziom $\langle(\Delta\mathbf{R}_i)^2\rangle$, który wynika z rozerwania ponad 50% kontaktów w białku. Głębokie ostre minimum w wykresie fluktuacji ($\langle(\Delta\mathbf{R}_i)^2\rangle \simeq 0$) tworzy pozostałe 2% reziduów, aminokwas 13 w kontakcie z 17, który zgodnie z diagramem scenariuszowym zostanie zerwany przed drugim maksimum. Jednakże dla struktury białka rozerwanej w połowie analiza drugiego modu jest bardziej przydatna do prognozowania rozrywania kontaktów, gdzie wyraźnie najmniejsze fluktuacje mają rezidua tworzące wiązania wodorowe w F+G (oznaczone strzałkami). Kontakty te zostaną rozerwane wcześniej niż te wskazane przez pierwszy mod. Fakt, iż kontakt, który będzie rozerwany wcześniej jest widoczny w drugim modzie a nie pierwszym, może wynikać z ułożenia równoległego osi kontaktu 13+17 do kierunku rozciągania. Aż do momentu rozerwania F+G kształt wszystkich fluktuacji pozostaje taki sam. Przewidywanie to jest zgodne z opisem we wcześniejszym paragrafie.

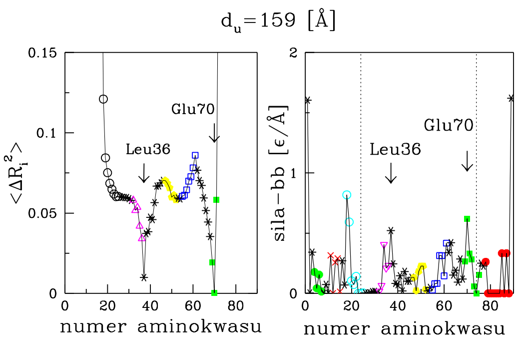
Konformacja białka tuż przed drugim maksimum (d_u=136)

Drugie maksimum na krzywej $F(d)$ przedstawione na rysunku 9.3 pochodzi z rozerwania większości kontaktów pomiędzy wstęgami β C+F i B+E i kontaktów na zewnątrz struktury drugorzędowej. Dla rozciągnięcia d_u>120Å aż do d_u=142Å obserwujemy dwa głębokie minima dla reziduów F(72) i F(73) i niemalże zerowe fluktuacje dla aminokwasów 31, 32 i 33 z nici C (dane nie pokazane). Drugi mod charakteryzuje się dwoma ostrymi minimami rozdzielonymi reziduami o pewnych nieregularnych fluktuacjach odpowiadających nie rozerwanej strukturze białka. Minima te odpowiadają pierwszemu lub drugiemu reziduum, które jest nadal w natywnym kontakcie i na którym skupia się siła rozrywająca.

Warto zauważyć, że w konformacjach w których ponad połowa tytyny została rozwinięta, krzywa $\langle(\Delta\mathbf{R}_i)^2\rangle$ powstała dla swobodnych końców białka przyjmuje kształt fali stacjonarnej. W szczególności mody nieparzyste 3, 5 i 7 odpowiadają własnościom oscylatora harmonicznego o amplitudzie drgań znacznie większej niż amplituda drgań dla reziduów będących nadal w kontakcie i są one widoczne na końcach białka. Natomiast mody 2, 4 i 6 posiadają te same własności, ale w pobliżu nie rozwiniętej natywnej części białka.

Konformacja białka w i tuż po drugim maksimum (d_u=159 Å)

Drugie maksimum na krzywej $F(d)$ jest na samym szczycie rozszczepione na dwa maksima. Pierwsze z nich omówiliśmy powyżej. Drugie maksimum pochodzi z rozerwania kontaktów poza strukturą drugorzędową dla $d_u \simeq 158$ Å i pomiędzy B+E dla $d_u \simeq 160$ Å, oznaczonych odpowiednio gwiazdkami i trójkątami na rysunku 9.3. Wartość $\langle(\Delta\mathbf{R}_i)^2\rangle$ w pierwszym najwolniejszym modzie dla rozciągnięcia d_u=159 Å jest pokazana na rysunku 9.6, gdzie dwa głębokie minima odpowiadają aminokwasom leucynie z numerem 36 i glutaminianowi z numerem 70. W następnym przemieszczeniu wektora prędkości kontakt ten zostanie zerwany.



Rysunek 9.6: Lewy panel: średnie kwadratowe odchylenie $\langle(\Delta\mathbf{R}_i)^2\rangle$ aminokwasów w konformacji opowiadającej przemieszczeniu d_u=159Å. Prawy panel: wartości siły pochodzące od naprężeń wzdłuż łańcucha, wywierane na każdy aminokwas, także dla przemieszczenia d_u=159Å. Aminokwasy wewnątrz pionowych przerywanych linii odpowiadają natywnej nie rozciągniętej strukturze. Strzałki na lewym panelu wskazują najmniejsze fluktuacje, strzałki na prawym panelu największe siły, w obydwu przypadkach są to te same rezidua.

Struktura białka przed rozerwaniem ostatniej spinki β (d_u=239Å)

Wraz z rozerwaniem C+F i B+E w drugim maksimum zostają sukcesywnie rozrywane pozostałe elementy, aż do pozostania konformacji z jedną spinką β D+E i kilkoma krótkozasięgowymi kontaktami. W symulacji rozciągania kontakty między D+E zaczynają pękać dla odległości między końcami łańcucha $d_u \simeq 235$Å. Do analizy drgań modów w ramach modelu GNM wybraliśmy konformację odpowiadającą d_u=239Å, dla której fluktuacje w pierwszym niezerowym modzie o najmniejszej częstości są pokazane na rysunku 9.5. Rozkład $\langle(\Delta\mathbf{R}_i)^2\rangle$ wzdłuż sekwencji pokazuje dwa głębokie główne minima i dwa szerokie lokalne minima. Chcielibyśmy tutaj wyjaśnić, iż minima główne pochodzą od krótkozasięgowych kontaktów od reziduum 33 do 40, tak więc naprężenie pomiędzy nimi jest silniejsze (głębsze minimum). Tym niemniej widać wyraźnie, iż dwa szerokie minima odpowiadają naprężeniom między aminokwasami wewnątrz nici β, B oraz E. Analiza diagramu scenariuszowego 9.3(d) z dynamiki molekularnej potwierdza nasze prognozy, iż kontakty między reziduami 49, 50, 51 i 52 z nici D z reziduami 55, 56 i 57 z nici E zostaną rozerwane wraz z przemieszczeniem się wektora prędkości. Duże odchylenia od położenia równowagi dla pozostałych aminokwasów, które na rysunku 9.6 miały małe drgania, są także poprawne.

Podsumowując, powyżej przedstawiliśmy struktury dla których przewidywanie daje raczej jednoznaczne wskazanie pary i kolejności rozrywania kontaktów tak jak dla d_u=159Å. Należy jednak podkreślić, iż w pozostałych przypadkach kolejne mody raczej pozwalają zidentyfikować klaster kontaktów, niż kilka par kontaktów.

9.3.3 Lokalizacja naprężeń w strukturze białka

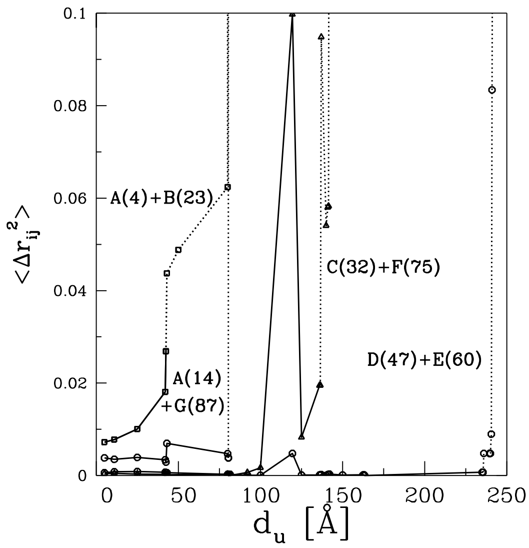
W stanie podstawowym całkowita siła na wszystkie atomy C^α jest równa zero. Przyłożenie zewnętrznej siły pochodzącej od rozciągania ze stałą prędkością powoduje powstanie kilku ognisk naprężeń na reziduach w stanie natywnym. Używając modelu Go, wyznaczyliśmy superpozycję składowych sił oraz cząstkowe siły dla potencjału Lennarda-Jonesa, siłę przenoszoną przez łańcuch białka, wywieraną na każde reziduum w białku podczas rozciągania. Stwierdziliśmy, że wartość siły wzdłuż łańcucha białka jest największa dla tych reziduów, gdzie $\langle(\Delta\mathbf{R}_i)^2\rangle$ jest najmniejsze. Przykład rozkładu wartości siły wzdłuż łańcucha i odpowiadająca mu wartości $\langle(\Delta\mathbf{R}_i)^2\rangle$ dla każdego C^α w modzie o najmniejszej niezerowej częstości dla np. $d_u \simeq 159$ są pokazane na rysunku 9.6. Tak więc największa wartość siły wywierana na nierozwiniętą strukturę białka jest zlokalizowana na reziduach 36, 70 i 17, dla $\langle(\Delta\mathbf{R}_i)^2\rangle \simeq 0$ w porównaniu do pozostałych reziduów. Należy jednak zauważyć, iż duża wartość siły dla atomu C^α o numerze

17 pochodzi od krótkozasięgowego kontaktu wewnątrz zakrętu, którego drgania są widoczne na rysunku 9.5(d), więc nie zaburza to naszej hipotezy.

Analiza pozostałych chwilowych konformacji białka podczas rozciągania prowadzi do tych samych wniosków, jednakże nie jest tak jasna jak ta przedyskutowana powyżej. W przypadku mniej rozwiniętego białka, jednoznaczna korelacja między największą wartością siły wzdłuż łańcucha a $\langle(\Delta \mathbf{R}_i)^2\rangle \simeq 0$ jest widoczna dopiero w kolejnych modach o najmniejszych częstościach lub ich złożeniu.

9.3.4 Fluktuacje średniego kwadratowego odchylenia między reziduami

Obecnie zajmiemy się analizą fluktuacji średniego kwadratowego odchylenia odległości $\langle(\Delta r_{ij})^2\rangle$ między reziduami będącymi w natywnym kontakcie, danych równaniem (9.5). Rysunek 9.7 pokazuje wartość średniego kwadratowego odchylenia odległości dla czterech par natywnych kontaktów jako funkcję d_u podczas symulacji rozciągania, gdzie zerwanie kontaktu jest widoczne jako gwałtowny wzrost wartości $\langle(\Delta r_{ij})^2\rangle$. Obserwujemy, iż wartość ta do momentu zerwania kontaktu fluktuje z małą i stałą amplitudą. W sytuacji gdy rezidua będące w kontakcie nie tworzą dodatkowych połączeń z resztą aminokwasów, wielkość $\langle(\Delta r_{ij})^2\rangle$ wzrasta gwałtownie, jednorazowo, tak jak w przypadku A+G dla d_u=85Å, lub stopniowo A(4)+B(23). Kolejnym wynikiem jest zgodność dużych wartości $\langle(\Delta r_{ij})^2\rangle$ dla kontaktów pękających po pierwszym maksimum, gdy następuje relaksacja struktury białka do nowego położenia równowagi. Ta reorientacja struktury związana jest z pękaniem i tworzeniem się na nowo kontaktów, nawet w temperaturze pokojowej. Opisane zachowanie dotyczy przede wszystkim kontaktów z β wstęg F+G i C+F. Przykładowe zachowanie się jednej pary C(32)+F(75) jest zobrazowane na rysunku 9.7, gdzie dla odległości d_u od 100Å do 125Å widać wyraźny okresowy wzrost $\langle(\Delta r_{ij})^2\rangle$. Zgodnie z diagramem scenariuszowym 9.1 kontakt ten nadal istnieje, gdyż wartość odległości r_{ij} jest nadal mniejsza niż wartość progowa określająca istnienie kontaktu $l_{e,ij}$, zdefiniowana w podrozdziale 3.1.1. Tym niemniej obserwujemy jego jednoznaczne zerwanie dla d_u = 139Å. Aby powyższa analiza była kompletna zauważmy, iż fluktuacje odległości pomiędzy dwoma najbliższymi sąsiadami wzdłuż sekwencji, oddziałującymi poprzez potencjał harmoniczny w modelu Go, są niemalże zerowe w porównaniu z fluktuacjami odległości atmów odziałujących poprzez potencjał Lennarda-Jonesa.



Rysunek 9.7: Średnie kwadratowe fluktuacje odległości $\langle(\Delta\mathbf{R}_{ij})^2\rangle$ pomiędzy parami aminokwasów i, j w funkcji przemieszczenia d_u na podstawie struktury z symulacji rozciągania w modelu Go. Linie ciągłe oznaczają istnienie kontaktu, linie przerywane odpowiadają fluktuacjom już rozerwanego tego samego kontaktu. Litery wraz z numerami wskazują z jakiej nici β pochodzi aminokwas o danym numerze.

Podsumowanie

W rozdziale tym po pierwsze pokazaliśmy, iż chwilowe struktury otrzymane podczas symulacji rozciągania białka ze stałą prędkością mogą być użyte do analizy drgań w modelu sieci gaussowskiej. Po drugie, na przykładzie kilku najbardziej znaczących chwilowych konformacji białka 1tit, pokazaliśmy, iż minima w rozkładzie średnich kwadratowych fluktuacji $\langle(\Delta\mathbf{R}_i)^2\rangle$ w modach o najmniejszych niezerowych częstościach wskazują, które natywne kontakty pękną dla kolejnego wzrostu odległości między końcami białka. Jednakże należy tu podkreślić, iż w przypadku pozostałych stanów rozrywanie tytyny i przede wszystkim bardziej skomplikowanych białek, jak np. zielonego białka GFP, przewidywanie nie jest już tak jednoznaczne. W takich przypadkach jesteśmy w stanie raczej przewidzieć klaster kontaktów, z których tylko niektóre pary pękną w następnym kroku. W takich sytuacjach niezbędna jest konfrontacja otrzymanych wyników z rzeczywistą symulacją, np. modelem Go. W przypadku gdy taka symulacja nie jest możliwa alternatywnym podejściem jest uwzględnienie niejednorodnej stałej sprężystości w modeu GNM. Okazuje się bowiem, iż w pewnym stopniu ta degeneracja może być usunięta poprzez zastosowanie niejednorodnej wartości stałej sprężystości g. Tak więc, po wprowadzeniu tylko trzech różnych wartości g, zlikwidowaliśmy dwie z sześciu degeneracji dla d_u=42 Å. Jednakże należy mieć też na uwadze, iż każde rozerwanie kontaktu jest związane z przejściem pomiędzy własnościami elastycznymi a nieelastycznymi. Tak więc analiza modów normalnych wynikająca z opisu idealnie harmonicznych odchyleń nie może do końca poprawnie opisać zmian w czasie rozrywania natywnego kontaktu. Ponadto model GNM, w postaci używanej w niniejszej pracy, dostarcza nam tylko informacji o amplitudzie drgań, nie mówiąc nic o ich kierunku (kierunek drgań mógłby być uzyskany z anizotropowej [14] wersji modelu GNM). Jednakże nawet małe rozciągnięcie tytyny prowadzi do dużej liczby zerowych wartości własnych (około 30, gdzie bryła sztywna wykazuje tylko sześć), co oznacza, iż następuje rozkład na kilka niezależnych regionów struktury.

Ponadto pokazaliśmy, iż najniższe niemalże zerowe fluktuacje są zgodne z największą siłą wywieraną na daną parę aminokwasów przez naprężenia w łańcuchu liczone w ramach modelu Go.

Rozdział 10

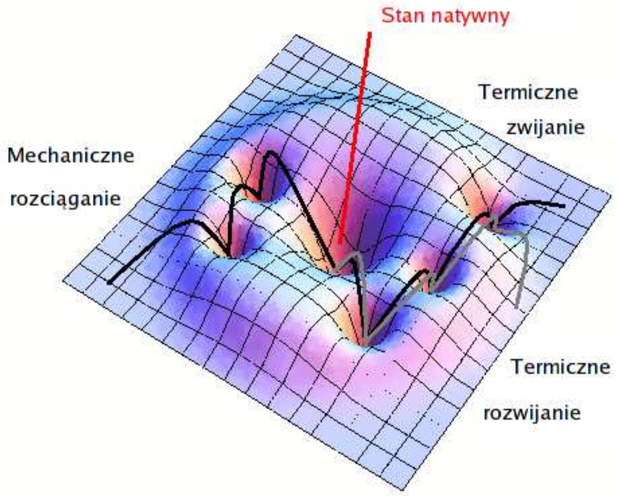
Termiczne rozwijanie białek

W rozdziale tym charakteryzujemy proces termicznego rozwijania białek. Definiujemy czas termicznego rozwijania i określamy jego niskotemperaturowe zachowanie. Porównujemy proces termicznego rozciągania z termicznym zwijaniem opisanym w części 5, a następnie z mechanicznym rozwijaniem dyskutowanym w rozdziale 6. Wyniki tych badań zawarte są w pracy [II].

10.1 Wstęp

Konformacyjne zmiany w białkach, jak już dobrze wiemy, występują w wielu procesach: termicznym zwijaniu które dyskutowaliśmy w rozdziale 5, mechanicznym rozwijaniu omówionym w rozdziałach 6 i 9, chemicznej denaturacji [202] czy też termicznie indukowanemu rozwijaniu. Procesy te w większości wymagają obserwacji zachodzących zmian w dużych okresach czasu, które nie są osiągalne w konwencjonalnych pełnoatomowych symulacjach. Ponadto takie procesy jak termiczne zwijanie czy rozwijanie wymagają ogromnej statystyki do określenia prawidłowości zjawiska. Tak więc mechanizmy te mogą być wydajnie badane jedynie przy pomocy modeli gruboziarnistych. Dodatkowym ważnym argumentem jest to, iż np. gruboziarnisty model typu Go bazuje na topologii stanu natywnego, co umożliwia zastosowanie go do badania różnych zjawisk i określenia zależności pomiędzy nimi, i dla wyznaczenia ich wspólnej zależności od parametrów takich jak np. temperatura. Schematyczny krajobraz energetyczny wiążący procesy termicznego zwijania, rozwijania i mechanicznego rozciągania jest pokazany na rysunku 10.1.

W rozdziale tym skupimy się na omówieniu ostatniego z wymienionych powyżej procesów — termicznego rozwijania. Zjawisko to jest często wykorzystywane w badaniach teoretycznych do interpretacji stanów przejściowych w białku podczas procesu zwijania [27]. Eksperymentalnie stan przejściowy jest badany przez inżynierię białkową a następnie analizę zmiany tak zwanego czynnika ϕ (ang. ϕ-value analysis) [203, 204]. Wartość czynnika ϕ zawiera informację o tym, czy podmiana danego aminokwasu na inny zmienia stan przejściowy dla procesu zwijania. Podstawą do teoretycznego badania stanu przejściowego poprzez symulacje termicznego rozwijania jest oczekiwanie, iż droga termicznego rozwijania powinna być odwrotna do zwijania termicznego i że stan przejściowy jest bezpośrednio/szybko osiągalny ze stanu natywnego, w szczególności dla wysokich i bardzo wysokich temperaturach, rzędu nawet $200°C$. Tym niemniej należy mieć na uwadze, że dla tak wysokich temperatur powierzchnia energii swobodnej może ulec modyfikacji. Dodatkowym czynnikiem motywującym badanie tego procesu jest próba połączenia zachowania się układu w stanie przejściowym z dużymi zmianami konformacyjnymi [29, 30] podczas rozwijania w ewolucji białek [28].



Rysunek 10.1: Schematyczny krajobraz energetyczny dla termicznego zwijania oraz mechanicznego rozwijania.

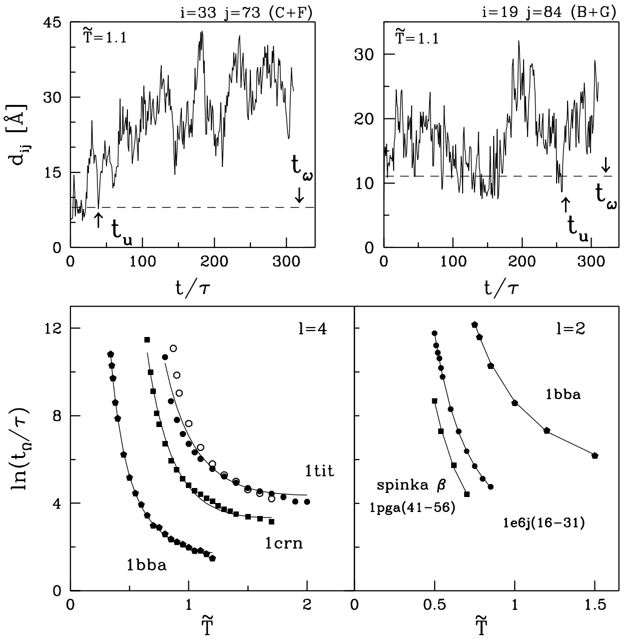
W niniejszym rozdziale badamy proces termicznego rozwijania w funkcji termicznego zwijania i mechanicznego rozciągania w ramach modelu gruboziarnistego Go o parametrach $\{6{-}12, C, M2, E^o\}$ i uogólnioną jego formą zawierającą reprezentacje grup bocznych o parametrach $\{C^{\alpha-\beta}, C, M2, E^o\}$, opisaną w rozdziale 3. Procedury symulacji mechanicznego rozwijania i termicznego zwijania zostały opisane w rozdziałach 4 i 5. Przypomnijmy, że w symulacjach mechanicznego rozwijania do jednego z końców białka w stanie natywnym jest przyłożony wektor prędkości, który wyznacza kierunek rozciągania ze stałą prędkością, podczas którego rejestrujemy rozrywanie kontaktów w funkcji odległości między końcami białka d_u. W procesie termicznego zwijania białko w stanie początkowym jest w konformacji przypadkowej, z której zgodnie ze swoją funkcją energii podąża do stanu natywnego, określonego utworzeniem wszystkich natywnych kontaktów. W symulacjach termicznego rozwijania konformacją startową jest stan natywny białka. Molekuła podlega fluktuacjom termicznym, które zwłaszcza w wysokiej temperaturze prowadzą do sukcesywnego zrywania kontaktów i rozwinięcia białka. Białko uznajemy za rozwinięte gdy wszystkie natywne kontakty zostaną jednocześnie zerwane po raz pierwszy. Kryterium istnienia kontaktu jest takie samo dla każdego z powyższych mechanizmów i dane jest zależnością $r_{ij}<l_{e,ij}$, gdzie $l_{e,ij}=1.5\sigma_{ij}$ zgodnie z definicją z rozdziału 3.1.1. Do badania używamy 27. domeny tytny (1tit), krambiny (1crn), peptydu krowiej trzustki (1bba), spinki β z białka G (1pga), i helisy α z białka (1bba).

10.2 Własności termicznego rozwijania

Proces termicznego rozwijania charakteryzujemy poprzez czasy rozwijania i własności diagramów scenariuszowych wprowadzonych w podrozdziale 4.1.3 w zależności od różnych temperatur.

Górne panele na rysunku 10.2 pokazują przykładowe zachowanie się odległości d_{ij} między aminokwasami z kontaktu natywnego w bardzo wysokiej temperaturze $\tilde{T}=1.1$ w zależności od czasu. Linia przerywana wyznacza odległość progową $l_{e,ij}$ dla danego kontaktu, poniżej której kontakt istnieje, a powyżej jest zerwany. Obserwujemy przypadkowe chwilowe a czasami okresowe zrywanie i powstawanie kontaktu spowodowane termicznymi fluktuacjami, co jest widoczne poprzez przekraczanie wartości progowej $l_{e,ij}$ w kolejnych chwilach czasu. Jednakże znajdujemy też dobrze widoczny i zależny od pary natywnego kontaktu czas t_u, dla którego kontakt pęka na dobre, przynajmniej w zakresie czasu rozważanego w symulacji. Wartość t_u jest zazwyczaj mniejsza dla kontaktów dalekozasięgowych i większa dla średnio/krótkozasięgowych, a wręcz niemożliwa do wyznaczenia dla kontaktów lokalnych. Na podstawie charakterystycznego zachowania się poszczególnych d_{ij} w funkcji czasu określiliśmy definicję czasu rozwijania t_ω, jako czasu koniecznego do zerwania wszystkich nielokalnych kontaktów $l>4$ w danym białku, gdzie l oznacza różnice wzdłuż sekwencji między i, j. Przykładowa wartość t_ω dla białka 1tit jest także pokazana na górnych panelach rysunku 10.2. W przypadku rozważania pojedynczych elementów struktury drugorzędowej, jak helisa czy spinka β, t_Ω wyznaczamy dla $l=2$.

Wartość czasu t_ω może się zmieniać od trajektorii do trajektorii, tak więc najbardziej typowy czas t_Ω wyznaczyliśmy jako medianę. Należy podkreślić, iż nie każda trajektoria prowadzi do pełnego rozwinięcia, zatem czas określony przez wartość średnią mógłby nie być adekwatny do rzeczywistości. Lewy dolny panel z rysunku 10.2



Rysunek 10.2: Górne panele: zależność od czasu wartości odległości d_{ij} między przykładową parą natywnych aminokwasów w temperaturze $\tilde{T}$=1.1 (lewy panel: atomy C^α 33 i 37 z nici C i F; prawy panel: atomy 19 i 84 z nici B i G dla białka 1tit). Dolne panele lewy (prawy): zależność t_Ω od temperatury $\tilde{T}$, otrzymana na podstawie kryterium z $l = 4(l = 2)$. Poniżej temperatury $\tilde{T}$=0.8 dane otrzymane są na podstawie co najmniej 201, a powyżej co najmniej na podstawie 500 trajektorii. Pełne czarne symbole są otrzymane w modelu Go a puste symbole w modelu z grupami bocznymi. Linie odpowiadają dopasowaniu do zależności $t_\Omega/\tau = exp(A/\tilde{T}^2 - B/\tilde{T} + C)$, współczynniki dopasowania i stałe są podane w tekście.

pokazuje wartość t_Ω w zależności od temperatury dla białek 1crn, 1bba, 1tit z modelu Go i dla 1tit z modelu Go z grupami bocznymi. Znaleźliśmy, iż wraz z obniżaniem temperatury $\tilde{T}$, t_Ω rośnie szybciej niż wynikałoby to z prawa Arrheniusa $\sim \exp(1/t_\omega)$ – zależność ta może sugerować także rozbieżność dla skończonego $\tilde{T}_O$ zgodną z prawem Vogela-Fulchera $\sim \exp(1/(t_\omega - \tilde{T}_O))$. Jednakże czasy rozwijania t_Ω mogą być z bardzo dobrą dokładnością dopasowane do zależności $t_\Omega/\tau = exp(A/\tilde{T}^2 - B/\tilde{T} + C)$, gdzie A, B, C są współczynnikami dopasowania i odpowiednio wynoszą (10.592, 10.492, 6.976) dla 1tit, (8.381, 9.887, 6.258) dla 1crn, (1.391, 0.943, 1.548) dla 1bba z kryterium l =4, (5.259, 1.576, 4.916) dla 1bba z l=2, (2.329, 0.671, 0.659) dla spinki z l=2, (3.363, 2, 143, 2.585) dla helisy z kryterium l=2. Współczynnik dopasowania dla każdego przypadku wynosi co najmniej 0.987. Trochę gorsze dopasowanie dostajemy dla zależności $exp[exp(D/\tilde{T})]$.

W tym przypadku $\tilde{T}_O$ wynosiłoby 0.56, 0.44 i 0.22 z barierą energetyczną 2.6, 2.3 i

1.5 odpowiednio dla 1tit,1crn i 1bba. Jednakże należy tu podkreślić, iż dla skończonych obiektów jakimi są pojedyncze białka, taka rozbieżność powinna mieć miejsce tylko dla $\tilde{T}$=0. Podobny problem występuje także przy dyskusji czasów zwijania, które w ogólności układają się w kształt U w zależności od temperatury (jak dyskutowalismy w rozdziale 5) z rozbieżnością w zerze, a w nieskończoności z zachowaniem zgodnym z prawem Arrheniusa.

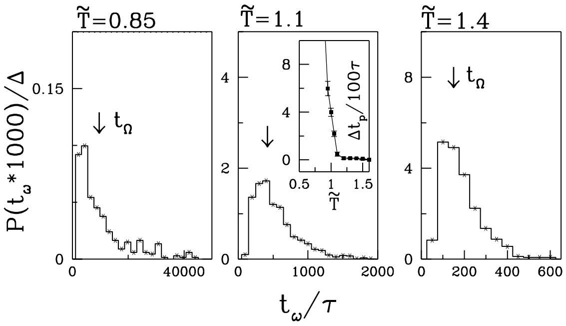
Powyższe rozważania mogłyby sugerować, iż charakter rozbieżności może zależeć od parametru l. Tym niemniej, zmiana parametru l w zakresie od 4 do 10 pokazała nieznaczny wpływ na t_Ω. Jednakże otrzymaliśmy, że obniżenie l poniżej 4 ma znaczący wpływ na wartość t_ω. Prawy panel rysunku 10.2 pokazuje zależność czasów rozwijania dla l=2 i l=4 dla białka 1bba. Otrzymane wyniki sugerują fizyczne znaczenie wartości parametru l=4, jako odróżniającego pomiędzy lokalnymi i nielokalnymi kontaktami [205]. Z punktu widzenia symulacji komputerowych rozważanie l<4 jest bardzo czasochłonne dla białek większych niż 1bba, a ponadto okazuje się nie mieć uzasadnienia w chemicznej denaturacji [202].

Dolny prawy rysunek 10.2 pokazuje zależność czasów rozwijania t_Ω od temperatury dla pojedynczej stabilnej spinki β i helisy α. W tym przypadku rozbieżność w $\tilde{T}_O$ jest jeszcze bardziej widoczna.

10.2.1 Charakterystyczna temperatura w termicznym rozwijaniu

Badania, jakie opisaliśmy w rozdziałach 5 oraz 6 pokazały i potwierdziły istnienie charakterystycznej temperatury dla procesu zwijania i mechanicznego rozciągania. W przypadku zwijania jest to temperatura, w której kinetyka procesu zwijania jest optymalna, i w szczególności dla 1tit w modelu $\{6-12, C, M2, E^o\}$ i $\{C^{\alpha-\beta}, 6-12, C, M2, E^o\}$, $\tilde{T}_{min}$ wynosi odpowiednio 0.28, 0.4. W przypadku procesu mechanicznego rozwijania temperatura ta została wyznaczona przez [129], jako temperatura powyżej której z wykresu siły w funkcji przemieszczenia znikają maksima siły i rozwijanie postępuje zgodnie modelem „dżdżownicy” (ang. worm-like chain) [206]. Jest to temperatura, od której własności czysto entropowe determinują rozwijanie białka. Dla białka 1tit ta zredukowana temperatura w modelu typu C^α i C^β odpowiednio wynosi 0.8 [129] i 0.9 [II]. Jednym z celów niniejszej pracy jest odpowiedzenie na pytanie, czy istnieje charakterystyczna temperatura T_Ω dla procesu termicznego rozwijania.

Rysunek 10.3 pokazuje histogramy czasów rozwijania t_ω w trzech różnych zredukowanych temperaturach $\tilde{T}$ dla tytyny. Rozkłady te charakteryzują się określonymi kształtami dla zależności położenia medialnego czasu t_Ω od czasu najbardziej prawdo-



Rysunek 10.3: Rozkład czasów rozwijania dla białka 1tit w trzech zredukowanych temperaturach 0.85, 1.1 i 1.4. Strzałki wskazują medialną wartość czasu. Wewnętrzny panel panelu środkowego pokazuje wartość różnicy pomiędzy czasem medialnym a maksimum w rozkładzie czasów t_ω w funkcji $\tilde{T}$.

podobnego i dla $\tilde{T}$=1.1 czasy te pokrywają się. Powyższe charakterystyczne zależności pozwalają nam na określenie definicji charakterystycznej temperatury w procesie termicznego rozwijania. Poniżej tej temperatury średni i najbardziej prawdopodobny czas różnią się znacząco, sugerując pojawianie się szerokiego rozkładu skal czasu. Powyżej proces jest niemalże natychmiastowy. Stwierdziliśmy, iż w przypadku białka 1tit charakterystyczna temperatura wynosi T_ω=1.1.

10.3 Porównanie termicznego rozwijania z termicznym zwijaniem

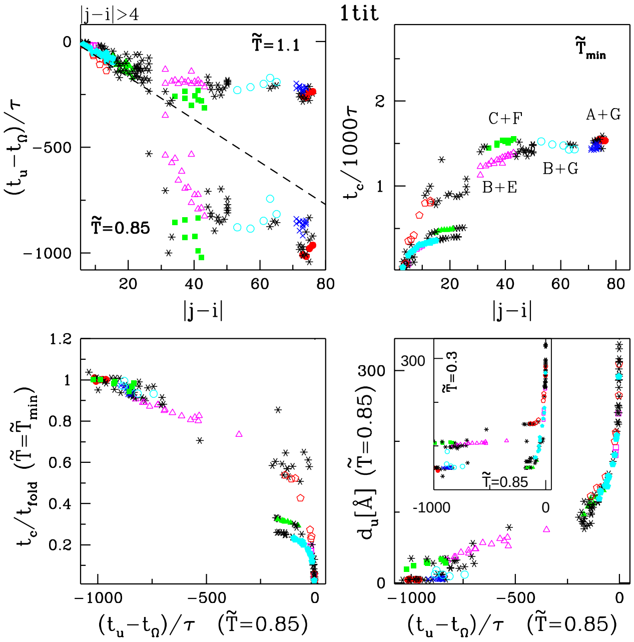
Scenariusz termicznego rozwijania

Górny lewy panel na rysunku 10.4 pokazuje wynik uśredniania serii scenariuszy rozwijania tytyny dla temperatury $\tilde{T}_\Omega$=1.1 i dla przykładowej niższej temperatury $\tilde{T}$=8.5. Ten uśredniony scenariusz odpowiada średniemu czasowi pękania t_U=$\langle t_u \rangle$ konkretnych natywnych kontaktów w funkcji ich odległości wzdłuż sekwencji $|j - i|$. Procesy rozwijania mogą przebiegać w bardzo długich skalach czasu, zwłaszcza w niższych temperaturach. Zatem, aby uzyskać szczegółową informację o zależności między kolejnymi wydarzeniami, czas pęknięcia został wyrażony poprzez wartość $t_U - t_\Omega$. Tak więc powyżej wzmiankowany panel pokazuje diagram scenariuszowy jako funkcję $t_U - t_\Omega$ od $|j-i|$. Otrzymaliśmy, iż powyżej $\tilde{T}_\Omega$ termiczne fluktuacje rozrywają kontakty niemalże jednocześnie, niezależnie od ich odległości sekwencyjnej $|j-i|$. Odwrotne zachowanie obserwujemy poniżej $\tilde{T}_\Omega$ gdzie scenariusz rozwijania przybiera charakterystyczny wzór oznaczający pękanie elementów struktury w ustalonej kolejności. Kontakty pomiędzy nićmi β C+F i A+G pękają znaczenie później niż te pomiędzy B+E i B+G. Otrzymana kolejność zrywania kontaktów jest zgodna z wynikami dla nanosekundowej długiej pełnoatomowej symulacji [117] rozwijania tego samego białka w stanie o temperaturze 450 K.

Termiczne zwijanie a rozwijanie

Większość wydarzeń pokazana dla scenariusza termicznego rozwijania dla temperatury $\tilde{T}$=0.85 przebiega w dokładnie odwrotnej kolejności do powstawania kontaktów podczas termicznego zwijania w temperaturze najszybszego zwijania $\tilde{T}_{min}$. Scenariusz termicznego zwijania jest pokazany na górnym prawym panelu rysunku 10.4. Wykres krzyżowy pomiędzy dwoma scenariuszami znajduje się na lewym dolnym panelu rysunku 10.4. Wykres krzyżowy pokazuje, iż termiczne rozwijanie jest monotoniczną funkcją zwijania, choć procesy te nie są swoją ścisłą odwrotnością. Tym niemniej należy podkreślić, iż wydarzenia które są najbardziej istotne dla poszukiwania stanu przejściowego (końcowego etapu przed zwinięciem i początkowego etapu podczas rozwijania) są niemalże liniowo antyskorelowane. Dodatkową liniową antykorelację stanu przejściowego obu procesów otrzymaliśmy także w uogólnionym modelu Go z grupami bocznymi (dane nie pokazane). Wyniki te jakościowo zgadzają się z rezultatami symulacji pełnoatomowej dla spinki β z białka 1pga, gdzie antykorelacja była badana pomiędzy charakterystycznym zbiorem 16 aminokwasowej spinki β, a rozwijanie było przeprowadzona w T=350 K

Przypomnijmy, iż proces zwijania uważamy za zakończony gdy wszystkie natywne kontakty zostaną utworzone po raz pierwszy jednocześnie. Kolejność tworzenia się kontaktów podczas zwijania dla $\tilde{T}_{min}$ jest pokazana na prawym górnym panelu 10.4 w funkcji $|j-i|$. Czas ustalenia się kontaktu, t_c, odpowiada średniemu czasowi powstania kontaktu po raz pierwszy. Czas ten wyznaczamy na podstawie co najmniej 500 trajektorii. Analiza procesu zwijania tytyny pokazała, iż 1tit może się zwijać przechodząc jedną z dwóch możliwych dróg zwijania. W pierwszym przypadku, który jest widoczny dla 24% trajektorii, kontakty między wstęgami C—F ustalają się w ponad dwa razy dłuższym czasie niż czas potrzebny do stworzenia grupy wiązań dla A—G i A'—G. W drugim przypadku pozostałe 76% trajektorii charakteryzuje się zwijaniem, w którym grupa kontaktów między C—F powstaje najpierw, a następnie zostają utworzone wiązania dla A—G. Scenariusz pokazany na prawym górnym panelu rysunku 10.4 pokazuje główną drogę zwijania i jest on znacząco różny od tego podobnego scenariusza łączącego dwa sposoby zwijania, pokazanego na rysunku 5.2 w rozdziale 5. Podobna sytuacja ma miejsce także podczas zwijania tytyny w modelu z grupami bocznymi, gdzie 82% trajektorii odpowiada głównej takiej samej drodze zwijania, dla 11% trajektorii druga ścieżka zwijania jest także zgodna z modelem jedno kulkowym, natomiast w pozostałych 7% obserwujemy trzeci sposób rozwijania. Tak więc antykorelacja dyskutowana w powyższym paragrafie jest zgodna z główną drogą zwijania białka tytyny.



Rysunek 10.4: Górny lewy i prawy panel przedstawia diagram scenariuszowy rozwijania i zwijania dla białka 1tit. Scenariusz rozwijania jest pokazany dla dwóch temperatur, gdzie linia przerywana oddziela pękanie kontaktów w każdej z nich. Scenariusz powstawania kontaktów odpowiada temperaturze najszybszego zwijania $\tilde{T}_{min}$. Dolny lewy panel przedstawia zależność pomiędzy scenariuszem zwijania a rozwijania w wyżej wymienionych temperaturach. Dolny prawy panel pokazuje zależność pomiędzy czasem pękania kontaktów podczas mechanicznego a termicznego rozwijania białka. Mechaniczne rozciąganie jest przeprowadzone dla temperatury $\tilde{T}$=0.3 i $\tilde{T}$=0.85 — odpowiednio dla danych z wewnętrznego i głównego panelu.

10.4 Termiczne a mechaniczne rozwijanie

Proces mechanicznego rozciągania białek ze stałą prędkością badaliśmy w rozdziale 6. Dla celów bieżącej dyskusji przypomnijmy, iż w tym przypadku diagram scenariuszowy pokazuje wielkość rozciągnięcia d_u dla jakiego kontakt jest rozerwany po raz ostatni w funkcji $|j-i|$. Przypomnijmy także, iż rozciąganie mechaniczne w temperaturze pokojowej nie jest odwrotnością zwijania, natomiast może być interpretowane jako jego odwrotność dopiero w temperaturze dla której maksima siły na wykresie $F(d)$ zanikają, w obszarze entropowym. Dolny prawy panel z rysunku 10.4 pokazuje wykres krzyżowy czasów rozrywania kontaktów podczas mechanicznego i termicznego rozwijania tytyny w temperaturze $\tilde{T}$=0.85, gdzie wyraźnie widzimy liniową zależność pomiędzy procesami. Kolejność wydarzeń w obu mechanizmach jest niemal identyczna, jednakże okresy czasu pomiędzy ich ustaleniem się nie skalują się liniowo. Wyjątkiem są wydarzenia blisko stanu natywnego. Temperatura zredukowana $\tilde{T}$=0.85 wyznacza granicę entropową, w jakiej było przeprowadzone mechaniczne rozciąganie, jednakże jest ona poniżej charakterystycznej temperatury dla procesu termicznego rozwijania, $\tilde{T}_{\Omega}$. Panel wewnętrzny na tym samym rysunku pokazuje, iż w przypadku gdy proces mechanicznego rozciągania będzie miał miejsce poniżej swojej charakterystycznej temperatury, np. w temperaturze pokojowej $\tilde{T}$=0.3, a termiczne rozwijanie w $\tilde{T}$=0.85, to procesy te bynajmniej nie będą skorelowane.

Podsumowanie

W niniejszym rozdziale określiliśmy definicję czasu termicznego rozwijania $\tilde{T}_{\Omega}$. Pokazaliśmy szybszą niż wynikałoby to z prawa Arrheniusa rozbieżność czasów rozwijania w niskich temperaturach. Wykazaliśmy istnienie charakterystycznej temperatury dla procesu termicznego rozwijania, poniżej której scenariusz rozwijania kontaktów zależy od odległości sekwencyjnej $|j-i|$ pomiędzy parą i,j w strukturze białka, i gdzie pojedyncze czasy rozwijania układają się w szeroki rozkład. Ponadto pokazaliśmy, iż długie czasy tworzenia się kontaktów podczas termicznego zwijania są antyskorelowane z krótkimi czasami pękania kontaktów podczas rozwijania.

Reasumując pokazaliśmy, iż prosty, opierający się na topologii białka w stanie natywnym dynamiczny model Go, poprawnie odtwarza kolejność zmian konformacyjnych w 27 domenie tytyny znanych z doświadczenia i symulacji pełnoatomowych, w procesach termicznego zwijania, termicznego rozwijania i mechanicznego rozciągania.

Rozdział 11

Podsumowanie

W niniejszej rozprawie pokazaliśmy, że procesy termicznego zwijania i rozwijania białek, jak też ich mechanicznego rozciągania, mogą być badane przy pomocy modeli gruboziarnistych typu Go.

Po pierwsze, poprzez porównanie wyników teoretycznych z doświadczalnymi, wyłoniliśmy optymalną formę modelu Go do rozciągania białek spośród 504 różnorakich wariantów. Wykazaliśmy, iż ten optymalny model Go zdefiniowany jest poprzez parametry $\{6-12, C, M3, E^o\}$ wprowadzone w rozdziale 3.

Najbardziej ambitnym celem rozprawy było wykonanie przeglądu własności mechanicznych struktur białek zawartych w bazie PDB. Spośród ok. 7500 kompletnych struktur wyłoniliśmy 134 najsilniejsze białka. Otrzymane wyniki pozwoliły nam na scharakteryzowanie własności mechanicznych białek. Wykazaliśmy, że znane do tej pory najbardziej stabline mechanicznie białka wykazują opór co najmniej o połowę słabszy od najsilniejszego wyznaczonego przez nas białka 1c4p. Ponadto wykazaliśmy, że własności mechaniczne zależą od klasy, architektury i topologii struktury białka. Pokazaliśmy, że grupa najsilniejszych białek charakteryzuje się tylko 13 typami architektury i 6 typami topologicznymi. Wyznaczyliśmy i opisaliśmy sześć najbardziej charakterystycznych motywów w białkach odpowiedzialnych za bardzo duży opór mechaniczny. Zidentyfikowaliśmy, trzy nowe rodzaje imadła mechanicznego, które są odpowiedzialne za bardzo duży opór mechaniczny białek podczas rozciągania.

Wyniki które uzyskaliśmy podczas przeglądu bazy PDB umieściliśmy w specjalnie w tym celu skonstruowanej przez nas bazie danych $BSDB$. Do bazy tej zostało dołączone oprogramowanie, które umożliwa każdej zainteresowanej osobie samodzielne wykonanie obliczeń numerycznych. Należy tutaj podkreślić, że otrzymane przez nas wyniki znalazły już pierwsze potwierdzenie w pracach doświadczalnych. Grupa prof. Mariano Carrion z *Cajal Institute, CSIC* w Madrycie pokazała doświadczalnie (dane jeszcze nie opublikowane), że białka 1g1k i 1aoh, zajmujące pozycje 3 i 4 na naszej liście, rzeczywiście są ponad dwukrotnie silniejsze niż 27. domena tytyny.

Kolejnym ważnym rezultatem tej pracy jest wyjaśnienie własności mechanicznych białek z węzłami. Pokazaliśmy, że proces zaciskania węzłów w białkach jest istotnie różny od mechanizmu, który występuje w polimerach. Węzły białkowe zaciskają się i w końcowym stanie jeden z końców zawsze zatrzymuje się w charakterystycznym ostrym zakręcie lub na początku helisy. W większości przypadków miejsca te są determinowane poprzez obecność glicyny lub proliny.

Ponadto pokazaliśmy, że wyniki otrzymane w modelach Go mogą być wykorzystane do przewidywania dużych zmian konformacyjnych w białkach za pomocą modelu sieci gaussowskiej GNM. Wynik ten może umożliwić badanie różnego rodzaju zmian konformacyjnych w dużych układach biologicznych których struktury w stanie przejściowym nie są dostępne, a otrzymanie ich z modeli pełnoatomowych nie jest możliwe.

Ostatnim elementem naszej pracy było scharakteryzowanie procesu termicznego rozwijania białek. Określiliśmy czasy termicznego rozwijania białek oraz ich niskotemperaturowe zachowanie. Wykazaliśmy istnienie charakterystycznej temperatury dla procesu termicznego rozwijania, poniżej której scenariusz rozwijania kontaktów zależy od odległości sekwencyjnej $|j-i|$ pomiędzy parą i, j w strukturze białka, i gdzie pojedyncze czasy rozwijania układają się w szeroki rozkład. Ponadto pokazaliśmy, że długie czasy tworzenia się kontaktów podczas termicznego zwijania są antyskorelowane z krótkimi czasami pękania kontaktów podczas rozwijania. Pokazaliśmy, że prosty, opierający się na topologii białka w stanie natywnym, dynamiczny model Go, poprawnie odtwarza kolejność zmian konformacyjnych w 27. domenie tytyny znanych z doświadczenia i symulacji pełnoatomowych, w procesach termicznego zwijania, termicznego rozwijania i mechanicznego rozciągania.

Dodatek A

Tablice

Tablica A.1: 134 najsilniejsze białka z $40 \leq N \leq 150$. Tytyna i ubikwityna są dodane dla porównania. λ oznacza położenie maksimum siły w procentach, BnMA oznacza kształt siły w fukcji rozciągania d. Kolejność uporządkowania białek n o tej samej wartości F_{max}, w ramch błędu obliczenia, jest przypadkowa. Wskaźnik ^{S} oznacza białka z mostkami siarczkowymi, których zerwanie podczas symulacji rozciągania było uniemożliwione. Puste miejsca w kolumnie $CATH$ oznaczają brak dostępnej klasyfikacji tych białek.

n	PDB	$F_{max}[\epsilon/\text{Å}]$	λ[%]	CATH	BnMA
1	1c4p	5.1	18	3.10.20	M
2	1qqr	5.1	19	3.10.20	M
3	1g1k	3.9	6	2.60.40	MA
7	1aoh	3.8	6	2.60.40	MA
4	1c76	3.8	17	3.10.20	MA
5	1c77	3.8	25	3.10.20	MA
6	1c79	3.8	25	3.10.20	MA
8	1c78	3.7	25	3.10.20	MA
9	2sak	3.7	18	3.10.20	MA
10	1nam	3.7	9	2.60.40	MA
11	1so9	3.6	18	2.60.370	MA
12	1ppx	3.5	40	3.90.79	B2M
13	1ssn	3.5	34	3.10.20	B1M
14	1rnz^s	3.4	45	3.10.130	B3M
15	1ie5^s	3.4	14	2.60.40	MA
16	1b88	3.4	7	2.60.40	MA
17	3rsk^s	3.4	45	3.10.130	B2MA
18	1npu	3.4	8	2.60.40	MA
19	2ncm	3.3	8	2.60.40	MA
20	1anu	3.3	7	2.60.40	MA
21	1rlf	3.3	9	3.10.20	MA
22	1eaj	3.3	8	2.60.40	MA
23	1oo2	3.3	26	2.60.40	MA
24	1h5b^s	3.2	8	2.60.40	MA
25	1i9e	3.2	4	2.60.40	MA

n	PDB	$F_{max}[\epsilon/\text{Å}]$	$\lambda[\%]$	CATH	BnMA
26	1mvf	3.1	12	2.60.40	MA
27	1f5w	3.1	7	2.60.40	MA
28	1sp0	3.1	16	2.60.370	MA
29	1amx	3.1	31	2.60.40	B1MA
30	1i3o	3.1	49	3.40.50	B2MA
31	1tfp	3.1	21	2.60.40	MA
32	1ves	3.1	6		MA
33	1sn0	3.1	21	2.60.40	MA
34	1oau	3.1	5	2.60.40	MA
35	1sn2	3.0	21	2.60.40	MA
36	1oax	3.0	5	2.60.40	MA
37	1oar	3.0	5	2.60.40	MA
38	1pun	3.0	41	3.90.79	B2M
39	1j05	3.0	5	2.60.40	MA
40	1lve^s	3.0	6	2.60.40	MA
41	1fvc	2.9	5	2.60.40	MA
42	1p7e	2.9	14	3.10.20	MA
43	1jhl	2.9	6	2.60.40	MA
44	1gke	2.9	27	2.60.40	MA
45	1etb	2.9	26	2.60.40	B1MA
46	1vhp	2.9	3	2.60.40	MA
47	1c08^s	2.9	6	2.60.40	MA
48	1i3v	2.9	4	2.60.40	MA
49	1pga	2.9	3	3.10.20	MA
50	1l2n	2.9	10	3.10.20	MA
51	1dfu	2.9	41	2.40.240	B1M
52	1yn4	2.9	5		MA
53	1bmz	2.9	21	2.60.40	MA
54	1kip^s	2.9	6	2.60.40	MA
55	1qd0	2.9	4	2.60.40	MA
56	1ivl^s	2.9	7	2.60.40	MA
57	1tyr	2.9	32	2.60.40	B1MA
58	1ic4	2.9	6	2.60.40	MA
59	1bz8	2.9	28	2.60.40	B1MA
60	1vfb^s	2.8	6	2.60.40	MA
61	1wtl	2.9	4	2.60.40	MA
62	1jrk	2.9	32	3.90.79	B2MA
63	1sok	2.9	25	2.60.40	B1MA
64	1bvk	2.9	5	2.60.40	MA
65	1ie4	2.8	24	2.60.40	MA
66	2try	2.8	32	2.60.40	B1MA
67	1e5a	2.8	26	2.60.40	MA
68	1kir^s	2.8	6	2.60.40	MA
69	1iik	2.8	26	3.10.130	MA
70	1v5o	2.8	16	3.10.20	MA
71	1k53	2.8	26	3.10.20	MA

n	PDB	$F_{max}[\epsilon/\AA]$	$\lambda[\%]$	CATH	BnMA
72	1ttr	2.8	26	2.60.40	MA
73	1kiq^s	2.8	6	2.60.40	MA
74	1oaq	2.8	3	2.60.40	MA
75	2imm	2.8	5	2.60.40	MA
76	1m94	2.8	12	3.10.20	MA
77	1kot	2.8	43	3.10.20	B2MA
78	1dvy	2.8	22	2.60.40	B1MA
79	1i8k^s	2.8	7	2.60.40	MA
80	1h8c	2.8	10	3.10.20	MA
81	1n4x^s	2.6	8	2.60.40	MA
82	1wit	2.8	6	2.60.40	MA
83	1gnu	2.8	44	3.10.20	B3MA
84	1em7	2.8	14	3.10.20	MA
85	1kgi	2.8	22	2.60.40	MA
86	1wiu	2.7	5	2.60.40	MA
87	1gko	2.7	21	2.60.40	B1MA
88	2dlf	2.7	5	2.60.40	MA
89	1p4i^s	2.7	4	2.60.40	MA
90	43ca^s	2.7	6	2.60.40	MA
91	43c9	2.7	6	2.60.40	MA
92	1tvd	2.7	5	2.60.40	MA
93	2igd	2.7	7	3.10.20	MA
94	4lve	2.7	5	2.60.40	MA
95	1igd	2.7	7	3.10.20	MA
96	1hz6	2.7	10	3.10.20	M
97	1a2y	2.7	5	2.60.40	MA
98	1j1x^s	2.7	6	2.60.40	MA
99	1fmf	2.7	64	3.40.50	B3MA
100	1yn5	2.7	5		MA
101	1ap2	2.7	4	2.60.40	MA
102	1eo6	2.7	36	3.10.20	MA
103	1k26	2.7	32	3.90.79	B2MA
104	1ieh	2.7	13	2.60.40	MA
105	1hz5	2.6	25	3.10.20	MA
106	1qp1	2.7	5	2.60.40	MA
107	1sn5	2.7	20	2.60.40	MA
108	1f2x	2.7	2	2.60.40	MA
109	1gb4	2.7	12	3.10.20	MA
110	1ugm	2.7	33	3.10.20	B2MA
111	1eta	2.7	30	2.60.40	MA
112	1tum	2.6	37	3.90.79	B2MA
113	1ufy	2.6	14	3.30.1330	MA
114	2rox	2.6	22	2.60.40	B1MA
115	1py9^s	2.6	4	2.60.40	MA
116	1rbj^s	2.55	26	3.10.130	B2M
117	1bzd	2.6	31	2.60.40	MA

n	PDB	$F_{max}[\epsilon/\text{Å}]$	$\lambda[\%]$	CATH	BnMA
118	5lve	2.6	5	2.60.40	MA
119	1f86	2.6	23	2.60.40	B1MA
120	1od9^s	2.6	9	2.60.40	MA
121	1w19	2.6	39		B1MA
122	1vjk	2.5	23	3.10.20	MA
123	1ttc	2.5	29	2.60.40	MA
124	1mfw	2.5	40	3.10.20	B1MA
125	1pav	2.5	35	3.30.110	B1MA
126	1eie	2.5	54	3.10.130	B2MA
127	1put	2.5	7	3.10.20	MA
128	1mg4	2.5	35	3.10.20	B1MA
129	1lm8	2.5	26	3.10.20	B1MA
130	1ui9	2.5	11	3.30.1330	MA
131	1nme	2.5	14	3.40.50	B2MA
132	1nvi	2.5	6	3.10.20	MA
133	1lqb	2.5	23	3.10.20	B1MA
134	1mel	2.5	3	2.60.40	MA
	1tit	2.1	4	2.60.40	MA
	1ubq	2.2	6	3.40.50	MA

Tablica A.2: Wartości sił dla białek ze zbioru S134 otrzymane w modelu z grupami bocznymi. Notacja jest taka sama jak w tablicy A.1.

n	PDB	N	$F_{max}[\epsilon/Å]$	L_n[Å]	L_m[Å]	L_f[Å]
1	1c4p	137	9.7	50.4	219.2	516.8
2	1g1k	143	9.0	43.5	158.1	539.6
3	1qqr	138	8.9	52.3	210.5	520.6
4	1aoh	138	8.4	34.0	142.1	520.6
5	1pun	129	8.2	34.2	299.4	486.4
6	1tum	129	8.1	33.7	295.5	486.4
7	1ie5	107	7.8	51.5	167.5	402.8
8	1kot	119	7.7	12.7	283.8	448.4
9	1so9	131	7.6	40.6	120.0	494.0
10	1c76	136	7.5	29.0	155.5	513.0
11	1yn4	99	7.5	35.4	128.3	372.4
12	2sak	121	7.5	31.9	169.5	456.0
13	1v80	76	7.4	37.1	133.1	285.0
14	1rlf	90	7.4	42.4	58.8	338.2
15	1c78	136	7.4	27.2	206.6	513.0
16	1npu	116	7.4	40.6	136.5	437.0
17	1ppx	129	7.3	36.0	280.2	486.4
18	1v5o	102	7.3	63.8	187.5	383.8
19	1c77	136	7.2	27.2	206.7	513.0
20	1c79	136	7.2	27.2	208.7	513.0
21	1ssn	136	7.2	9.5	244.0	513.0
22	1sp0	131	7.0	40.6	180.1	494.0
23	1pgx	83	6.9	62.7	134.9	311.6
24	2ncm	99	6.8	42.1	56.9	372.4
25	1hz6	72	6.8	41.6	127.2	269.8
26	1k26	156	6.8	46.1	289.3	589.0
27	1b9r	105	6.7	29.1	120.5	395.2
27	1pav	78	6.6	13.0	180.5	292.6
29	1jrk	156	6.5	42.2	215.0	589.0
30	1gnu	117	6.6	5.0	263.6	440.8
31	1vhp	117	6.5	40.8	109.1	440.8
32	1b88	114	6.5	39.0	120.8	429.4
33	1vjk	88	6.4	31.0	115.2	330.6
34	1eo6	117	6.4	22.1	235.5	440.8
35	1sn5	130	6.4	21.5	167.5	490.2
36	1pus	129	6.4	36.1	269.2	486.4
37	1oau	122	6.4	43.4	116.2	459.8
38	1h8c	82	6.4	39.5	118.5	307.8
39	1tvd	116	6.3	38.0	112.9	437.0
40	1ves	113	6.2	35.1	108.8	425.6
41	1oar	122	6.2	43.5	112.8	459.8
42	1sn0	130	6.2	21.5	180.0	490.2

n	PDB	N	$F_{max}[\epsilon/\text{Å}]$	L_n[Å]	L_m[Å]	L_f[Å]
43	1nme	146	6.2	52.8	347.1	551.0
44	1h5b	113	6.1	40.5	118.2	425.6
45	1oo2	119	6.1	12.9	166.5	448.4
46	1i3v	129	6.1	40.4	103.6	486.4
47	1i9e	115	6.1	49.5	63.4	433.2
48	1sn2	130	6.1	21.6	90.4	490.2
49	1w19	147	6.1	21.6	348.2	554.8
50	1m94	73	6.0	27.4	37.1	273.6
51	1km7	100	6.0	35.8	94.7	376.2
52	1ugm	113	6.0	21.9	136.2	425.6
53	1w29	146	6.0	19.4	206.5	551.0
54	1kgi	127	6.0	19.2	98.3	478.8
55	1wiu	93	6.0	39.0	49.2	349.6
56	1oax	122	5.9	43.7	56.7	459.8
57	1hz5	72	5.9	33.0	77.5	269.8
58	1k53	72	5.9	32.5	77.5	269.8
59	1mfw	107	5.9	12.4	140.4	402.8
60	1bvk	108	5.8	39.7	50.7	406.6
61	1fmf	137	5.8	18.8	319.1	516.8
62	1ivl	107	5.8	36.5	50.8	402.8
63	1wtl	108	5.8	40.4	53.7	406.6
64	1anu	138	5.8	24.0	35.4	520.6
65	1oaq	120	5.7	40.6	49.5	452.2
66	1ieh	135	5.7	51.1	113.3	509.2
67	1eaj	126	5.7	41.6	56.5	475.0
68	1f86	115	5.7	12.1	79.2	433.2
69	1i3o	144	5.7	40.4	0.0	543.4
70	1lqb	118	5.7	38.9	130.4	444.6
71	1bm7	127	5.6	12.2	76.2	478.8
72	1dfu	94	5.6	13.9	121.8	353.4
73	1f5w	126	5.6	41.4	65.9	475.0
74	1eta	127	5.6	12.6	125.4	478.8
75	1lsl	113	5.5	100.1	275.5	425.6
76	1bz8	126	5.5	20.7	121.2	475.0
77	1pga	56	5.5	26.5	30.4	209.0
78	1kip	107	5.5	36.3	46.1	402.8
79	43c9	113	5.5	36.5	47.5	425.6
80	2try	127	5.5	10.8	124.3	478.8
81	1nam	116	5.5	35.2	47.6	437.0
82	1ufy	122	5.5	30.8	80.2	459.8
83	1fvc	109	5.5	40.7	50.9	410.4
84	2igd	61	5.5	40.6	49.0	228.0
85	1tyr	127	5.5	10.6	124.9	478.8
86	1mel	148	5.4	38.0	45.3	558.6
87	1mg4	113	5.4	5.9	120.8	425.6
88	1igd	61	5.4	40.4	50.4	228.0

n	PDB	N	$F_{max}[\epsilon/Å]$	L_n[Å]	L_m[Å]	L_f[Å]
89	1tbe	76	5.4	33.5	48.0	285.0
90	1gb4	57	5.4	28.9	34.9	212.8
91	1p7e	56	5.3	26.3	31.5	209.0
92	1nvi	81	5.3	34.1	45.5	304.0
93	1ui9	122	5.3	27.9	60.9	459.8
94	1wit	93	5.2	39.3	50.4	349.6
95	1ttc	127	5.2	12.8	126.1	478.8
96	1mvf	135	5.2	38.7	140.3	509.2
97	1bzd	127	5.2	11.0	123.4	478.8
98	1i8k	107	5.2	34.3	46.4	402.8
99	1qp1	107	5.2	36.7	46.5	402.8
100	2dlf	113	5.2	39.3	54.1	425.6
101	4lve	114	5.2	39.8	52.9	429.4
102	1lm8	106	5.2	43.3	137.7	399.0
103	1bmz	127	5.2	11.5	74.6	478.8
104	1tfp	130	5.2	12.1	78.9	490.2
105	1dvy	124	5.1	12.4	120.0	467.4
106	1l2n	81	5.1	36.0	54.8	304.0
107	1ie4	127	5.1	15.3	97.3	478.8
108	1pqe	126	5.1	33.4	72.8	475.0
109	1kmt	141	5.0	44.5	71.5	532.0
110	1c08	107	5.0	35.3	41.9	402.8
111	1kir	107	5.0	36.0	44.6	402.8
112	2imm	114	5.0	39.9	52.9	429.4
113	1py9	116	5.0	40.6	51.0	437.0
114	1f2x	135	4.9	40.0	53.3	509.2
115	1gko	127	4.9	11.6	79.5	478.8
116	1kiq	107	4.9	36.2	47.8	402.8
117	1vfb	107	4.9	36.6	44.6	402.8
118	1em7	56	4.8	25.9	31.8	209.0
119	1rnz	125	4.8	34.1	99.7	471.2
120	1lve	122	4.8	39.2	51.6	459.8
121	2rox	127	4.8	14.0	87.8	478.8
122	1gke	120	4.8	15.1	99.8	452.2
123	1etb	127	4.8	8.6	89.7	478.8
124	5lve	114	4.8	36.8	53.4	429.4
125	1ic4	107	4.7	35.4	48.9	402.8
126	1com	127	4.7	25.7	53.0	478.8
127	1jf8	131	4.7	17.3	214.8	494.0
128	1dvt	115	4.7	12.2	79.1	433.2
129	1jhl	108	4.6	39.1	50.5	406.6
130	1amx	127	4.6	9.5	84.2	478.8
131	1n4x	113	4.5	33.7	52.4	425.6
132	1a2y	107	4.5	37.0	157.7	402.8
133	1qd0	128	4.5	40.2	45.7	482.6
134	1rbj	124	4.1	32.0	247.7	467.4

Tablica A.3: Najsilniejsze białka z N>150. Kolejność ustawienia białek w tablicy dla tego samego F_{max} jest arbitralna. Symbol n_D oznacza liczbę domen w białku. Kolumny D_1 i D_2 zawierają klasyfikację $CATH$ dla pierwszej i drugiej domeny w bialku. Wartości siły oznaczone symbolem $*$ pochodzą z szybkiego rozciągania i przeskalowania zgodnie z formułą $F_{max,slow} = 0.86\ F_{max,fast}$ -0.13

	PDB	N	$F_m[\epsilon/Å]$	$L_n[Å]$	$L_m[Å]$	$L_f[Å]$	wzór	n_D	D_1	D_2
1	1bg2	323	4.8	17.0	133.3	1223.6	B1MA	1	3.40.850.10	
2	2sli	679	3.8	50.6	224.3	2576.4	B1MA	3	2.60.120.200	2.120.10.10
3	1tmo	794	3.7	44.0	2309.7	3013.8	B4MA	4	3.40.50.740	3.40.228.10
4	1bfd	523	3.6	62.1	124.1	1983.6	MA	3	3.40.50.1220	3.40.50.970
5	2plc	274	3.6	9.4	150.4	1037.8	B1MA	1	3.20.20.190	
6	1clc	541	3.6 *	39.6	162.3	2052.0	MA	2	2.60.40.10	1.50.10.10
7	1ile	821	3.6 *	40.4	2372.3	3116.0	B5MA	3	1.10.730.10	3.90.740.10
8	5ptd	296	3.5	7.4	199.6	1121.0	B2MA	1	3.20.20.190	
9	1thv	207	3.5	23.4	38.4	782.8	MA	1	2.60.110.10	
10	1cpy	421	3.5 *	40.5	1216.3	1596.0	B4MA	2	3.40.50.1820	1.10.287.410
11	1kas	411	3.5 *	17.6	54.2	1568.0	MA	2	3.40.47.10	
12	1bif	432	3.5 *	54.3	1110.8	1637.8	B4MA	2	3.40.50.300	3.40.50.1240
13	1ctn	538	3.5 *	60.1	1024.5	2040.6	B6MA	3	2.60.40.10	3.20.20.80
14	1dot	686	3.4 *	28.1	2237.5	2603.0	B4MA	4	3.40.190.10	
15	1cbg	490	3.4 *	29.0	268.3	1868.2	B1MA	1	3.20.20.80	
16	1auq	208	3.4	20.9	389.3	786.6	B3MA	1	3.40.50.1820	
17	2ng1	293	3.4	14.0	778.5	1109.6	B5MA	2	1.20.120.140	3.40.50.300
18	1lpp	534	3.4 *	49.5	1543.3	2025.4	B3MA	1	3.40.50.1820	
19	1bs9	207	3.3	20.8	475.8	782.8	B3MA	1	3.40.50.1820	
20	1cgt	684	3.3 *	55.7	1291.5	2595.6	B6MA	4	3.20.20.80	2.60.40.1180
21	1dmr	779	3.3 *	24.8	2249.4	2956.4	B5MA	4	3.40.50.740	3.40.228.10
22	1qpg	415	3.2 *	14.3	1210.3	1573.2	B5MA	2	3.40.50.1260	3.40.50.1270
23	8ohm	435	3.2 *	43.9	1116.8	1648.2	B3MA	3	3.40.50.300	3.40.50.300
24	1zxq	192	3.2	79.2	451.7	726.2	B3MA	2	2.60.40.10	
25	1a8h	500	3.2 *	46.3	1577.4	1996.2	B6MA	3	3.40.50.620	2.170.220.10
26	1ciu	683	3.2 *	58.7	1566.4	2591.6	B8M	4	3.20.20.80	2.60.40.1180
27	1cyg	680	3.1 *	63.6	1267.1	2580.2	B6MA	4	3.20.20.80	2.60.40.1180
28	1ciy	577	3.1 *	41.9	84.4	2188.2	MA	3	1.20.190.10	2.100.10.10
29	1hcz	250	3.1	41.7	160.3	946.2	B1MA	2	2.60.40.830	2.40.50.100
30	1fsz	334	3.1 *	62.0	174.6	1265.6	B1MA	2	3.40.50.1440	3.30.1330.20
31	1cex	197	3.1	28.1	436.9	744.2	B3MA	1	3.40.50.1820	
32	1bag	425	3.1 *	13.7	508.5	1611.2	B3MA	2	3.20.20.80	2.60.40.1180
33	1gca	309	3.0 *	43.4	115.8	1170.8	MA	2	3.40.50.2300	
34	1cii	602	3.0 *	38.5	1783.2	2283.2	B2M	3	1.20.250.10	3.30.305.10
35	1avk	620	3.0 *	36.3	1051.9	2352.2	B6MA	3	3.10.450.40	3.10.450.40
36	1edg	380	3.0 *	31.4	272.5	1443.2	B2MA	1	3.20.20.80	
37	1ra9	159	3.0	13.9	193.2	600.4	B1MA	1	3.40.430.10	
38	1chd	198	3.0	8.5	299.3	748.6	B2MA	1	3.40.50.180	
39	1fts	295	2.9	10.5	779.7	1117.2	B2MA	2	1.20.120.140	3.40.50.300
40	1bk0	329	2.9	29.2	555.6	1246.6	B2MA	1	2.60.120.330	

Tablica A.4: Białka podobne homologicznie w zbiorze S134

nazwa	kod PDB
1 domena β streptokinazy	1c4p, 1qqr
2 domena kohezyny celulozy	1aoh, 1anu
z Clostridium thermocellum	
3 staphylokinase, sakstar variant	1ssn, 2sak, 1c76, 1c77, 1c78, 1c79
4 mutator mutt protein	1ppx, 1pun, 1pus, 1tum
5 APOCOX11(cytochrome C oxidase assembly protein ctag)	1so9, 1sp0
6 transthyretin (prealbumin)	2rox, 1f86, 1bz8, 1oo2
transthyretin (prealbumin)	1gko, 1tfp, 1tyr, dvy
7 transthyretin	1ttc, 1eta, 1etb, 1bmz
8 transthyretyna ludzka	1bzd, 2try
9 transthyretin szczurza	1gke, 1kgi, 1ie4
10 transthyretin (morski leszcz)	1sn0, 1sn2, 1sn5
11 wirus coxsackie i receptor domeny D1 adenowirusa CAR	1eaj, 1f5w
12 mysie monoclonal antyciało D1.3 (VH or VL domain)	1kip, 1kiq, 1a2y, 1kir
13 Caspase-3 (duża podjednostka)	1nme, 1i3o
14 murine, mysi receptor komórki T (TCR)v domena α	1b88, 1nam, 1h5b
15 domena b1 białka L z peptostreptococcus magnus	1hz5, 1hz6, 1k53
16 gabaa receptor associated protein gabarap	1gnu, 1kot
17 Immunoglobulin E, FV domain, SPE-7	1oau, 1oax, 1oaq, 1oar
18 Immunoglobulin G, VL domain	1ivl, 2imm
19 Immunoglobulin VL domain (Bence-Jones protein)	1wtl, 1qp1
20 ludzki łńcuch Immunoglobuliny K-4, Len	4lve, 5lve, 1lve
21 twitchin Immunoglobulin superfamily domain, 18th domain	1wit, 1wiu
22 nudix protein from pyrobaculum aerophilum	1jrk, 1k26
23 chorismate mutase from thermus thermophilus	1ufy, 1ui9
24 terminal N domeny doublecortiny z DCLK	1mfw, 1mg4
25 trzecia IgG-binding domena z streptococcal białka G	1igd, 2igd, 1p7e
26 B1,B2 IgG-binding domena z streptococcal białka G	1pgx
27 różne warianty domeny B1, białko streptococcal G	1em7, 1gb4
28 anti-hen egg lysozyme antibody (HYHEL-10) fragment VL,VH	1jhl, 1c08, 1ic4
29 lumazine synthase z mycobacterium tuberculosis	1w19,
30 elongin-B from HIF-pvhl/elongin-c/elongin-b complex	1lqb, 1lm8
31 ribonuclease B lub A	1rbj, 1rnz

Spis tablic

Bibliografia

[1] A. Matouschek, *Curr. Opin. Struct. Biol.* **13**, 98–109 (2003).

[2] S. Prakash and A. Matouschek, *Trends in Biochemical Sciences* **29**, 593–600 (2004).

[3] H. P. Erickson, *Science* **276**, 1090–1090 (1997).

[4] M. Rief, M. Gautel, F. Oesterhelt, J. M. Fernandez, and H. E. Gaub, *Science* **276**, 1109–1112 (1997).

[5] R. Hynes, New York: Springer, 1990.

[6] P. Shrestha and M. Mori, *Tenascin: An Extracellular Matrix Protein in Cell Growth, Adhesion and Cancer*, Austin, Texas: Landes Biosci, 1997.

[7] M. Rief, J. Pascual, M. Saraste, and H. Gaub, *J. Mol. Biol.* **286**, 553–561 (1999).

[8] H. Lu, B. Isralewitz, A. Krammer, V. Vogel, and K. Schulten, *Biophys. J.* **75**, 662–671 (1998).

[9] H.Lu and K. Schulten, *J. Chem. Phys.* **247**, 141–153 (1999).

[10] M. Cieplak, T. X. Hoang, and M. O. Robbins, *Proteins: Struct. Funct.and Genet.* **49**, 114–124 (2002).

[11] F. B. Scheinerman and C. L. B. III, *Proc. Natl. Acad. Sci. (USA)* **95**, 1562–1567 (2001).

[12] H. Abe and N. Go, *Biopolymers* **20**, 1013–1031 (1981).

[13] S. Takada, *Proc. Natl. Acad. Sci. (USA)* **96**, 11698–11700 (1999).

[14] T. Haliloglu, I. Bahar, and B. Erman, *Phys. Rev. Lett.* **79**, 3090–3093 (1997).

[15] M. Tirion, *Phys. Rev. Lett.* **77**, 1905–1908 (1996).

[16] K. Schulten and I. Kosztin, *Lectures in Theoretical Biophysics*, 2000.

[17] Y. Zhou and M. Karplus, *Nature* **401**, 400–403 (1999).

[18] N. V. Dokholyan, S. V. Buldyrev, H. E. Stanley, and E. I. Shakhnovich, *Folding Des.* **3**, 577–587 (1998).

[19] C. Hardin, Z. Luthey-Schulten, and P. G. Wolynes, *Proteins: Struct. Funct. and Genet.* **34**, 281–294 (1999).

[20] C.-E. Chang, T. Shen, J. Trylska, V. Tozzini, and J. A. McCammon, *Biophys. J* **90**, 3880–3885 (2006).

[21] T. X. Hoang and M. Cieplak, *J. Chem. Phys.* **112**, 6851–6862 (2000).

[22] D. K. Klimov and D. Thirumalai, *Proc. Natl. Acad. Sci. (USA)* **97**, 2544 (2000).

[23] M. Cieplak and T. X. Hoang, *Biophys. J.* **84**, 475–488 (2003).

[24] C. Clementi, H. Nymeyer, and J. N. Onuchic, *J. Mol. Biol.* **298**, 937 (2000).

[25] J. Karnicolas and C. L. B. III, *Protein Sci.* **11**, 2351–2361 (2002).

[26] P. Virnau, L. A. Mirny, and M. Kardar, *PLOS Compu. Biol.* **2**, 174–179 (2006).

[27] V. Pande and D. Rokhsar, *Proc. Natl. Acad. Sci. (USA)* **96**, 9062–9067 (1999).

[28] I. Chang, M. Cieplak, J. R. Banavar, and A. Maritan, *Protein Sci.* **13**, 2446 (2004).

[29] A. Li and V. Dagett, *Proc. Natl. Acad. Sci. (USA)* **91**, 10430–10434 (1994).

[30] S. Kazimierski, K. Wong, S. Freund, Y. Tan, A. Fersht, and V. Daggett, *Proc. Natl. Acad. Sci. (USA)* **98**, 4349–4354 (2001).

[31] C. Orengo, A. Michie, S. Jones, D. Jones, M. S. J.M., and Thornton, *Structure* **5**, 1093–1108 (1997).

[32] L. Stryer, *Biochemia, 4th ed.*, Freeman, New York, 1995.

[33] W. Humphereys, A. Dalke, and K. Schulten, *J. Mol. Graph.* **14**, 33 (1996).

[34] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, *Nucl. Acid. Res.* **28**, 235–242 (2000).

[35] M. L. Mansfield, *Nat. Struct. Biol.* **1**, 213–214 (1994).

[36] W. Taylor, *Nature* **406**, 916–919 (2000).

[37] R. Lua and A. Grosberg, *PLOS Comp. Biol.* **2**, 350–357 (2006).

[38] W. Taylor, *Comput. Biol. Chem.* **31**, 151–162 (2007).

[39] P. Virnau, L. A. Mirny, and M. Kardar, *PLOS Comp. Biol.* **2**, 1074–1079 (2006).

[40] A. Navon and A. Goldberg, *Mol. Cell* **8**, 1339–1349 (2001).

[41] X. R. Bao, H. J. Lee, and S. R. Quake, *Phys. Rev. Lett.* **95**, 199901 (1999).

[42] B. Marcone, E. Orlandini, A. L. Stella, and F. Zonta, *J. Phys. A - Mathematical and General* **38**, L15–L21 (2005).

[43] Y. K. P. Virnau and M. Kardar, *J. Am. Chem. Soc.* **127**, 15102 (2005).

[44] P. Pieranski, S. Przybyl, and A. Stasiak, *Eur. Phys. J. E* **6**, 123 (2001).

[45] K. Koniaris and M. Muthukumar, *Phys. Rev. Lett* **66**, 2211–2214 (1991).

[46] K. Koniaris and M. Muthukumar, *J. Chem. Phys.* **95**, 2873–2881 (1991).

[47] V. Katritch, J. Bednar, D. Michoud, R. G. Scharein, J. Dubochet, and A. Stasiak, *Nature* **384**, 142–145 (1996).

[48] L. Tskhovrebova, K. Trinick, J. A. Sleep, and M. Simmons, *Nature* **387**, 308–312 (1997).

[49] M. Gauel and D. Goulding, *FEBS Lett.* **375**, 11–14 (1996).

[50] B. D. Hames, N. M. Hooper, and J. Houghton, *Biochemia*, PWN, Warszawa, 2000.

[51] A. Hershko, A. Ciechanover, H. Heller, A. L. Haas, and I. Rose, *Proc. Natl. Acad. Sci. (USA)* **77**, 1783–1786 (1980).

[52] G. Binnig, C. Quate, and C. Gerber, *Phys. Rev. Lett.* **56**, 9597–9601 (1986).

[53] J. M. Fernandez and H. Li, *Science* **303**, 1674 (2004).

[54] M. Schlierf, H. Li, and J. Fernandez, *Proc. Natl. Acad Sci. (USA)* **101**, 7299 (2004).

[55] G. Lee, L.A.Chrisey, and R.J.Colton, *Science* **266**, 771–773 (1994).

[56] V. T. Moy, E. L. Florin, and H. Gaub, *Science* **266**, 257–259 (1994).

[57] H. Grubmuller, B. Heymann, and P. Tavan, *Science* **271**, 997–999 (1996).

[58] U. Bockelmann, B. Essevaz-Roulet, and F. Heslot, *Phys. Rev. Lett.* **4489-4492**, 79 (1997).

[59] M. Rief, F. Oesterhelt, B. Heymann, and H. E. Gaub, *Science* **257**, 1295–1297 (1997).

[60] K. Mitsui, M. Hara, and A. Ikai, *FEBS Lett.* **385**, 29–33 (1996).

[61] A. F. Oberhauser, P. E. Marszalek, H. P. Erickson, and J. M. Fernandez, *Nature* **14**, 181–185 (1998).

[62] M. Rief, J. M. Fernandez, and H. E. Gaub, *Phys.Rev. Lett.* **81**, 4764–4767 (1998).

[63] H. B. Li and J. M. Fernandez, *J. Mol. Biol.* **334**, 75–86 (2003).

[64] H. Li, W. A. Linke, A. F. Oberhauser, M. Carrion-Vazquez, J. G. Kerkviliet, H. Lu, P. E. Marszalek, and J. M. Fernandez, *Nature* **418**, 98–1002 (2002).

[65] S. B. Fowler, R. B. Best, J. L. T. Herrera, T. J. Rutherford, A. Steward, E. Paci, M. Karplus, and J. Clarke, *J. Mol. Biol* **322**, 841–849 (2002).

[66] M. Carrion-Vazquez, P. E. Marszalek, A. F. Oberhauser, and J. M. Fernandez, *Proc. Natl. Acad. Sci. (USA)* **96**, 11288–11292 (1999).

[67] C. Cecconi, E. A. Shank, C. Bustamante, and S. Marqusee, *Science* **309**, 2057–2060 (2005).

[68] H. Li, A. F. Oberhauser, S. B. Fowler, J. Clarke, and J. M. Fernandez, *Proc. Natl. Acad. Sci. (USA)* **97**, 6527–6531 (2000).

[69] H. P. Erickson, *Proc. Natl. Acad. Sci. (USA)* **91**, 10114–10118 (1994).

[70] L. Li, H. H.-L. Huang, C. L. Badilla, and J. M. Fernandez, *J. Mol. Biol.* **345**, 817–826 (2005).

[71] M. Rief, M. Gautel, M.Schemmel, and H. G. Gaub, *Biophys. J* **75**, 3008–3014 (1998).

[72] K. Watanabe, C. Muhle-Goll, M. S. Z. Kellermayer, S. Labeit, and H. L. Granzier, *Struct. Biol. J.* **137**, 248–258 (2002).

[73] A. F. Oberhauser, C. Badilla-Fernandez, M. Carrion-Vazquez, and J. M. Fernandez, *J. Mol. Biol.* **14**, 181–185 (1998).

[74] Y. Oberdorfer, H. Fuchs, and A. Janshoff, *Langmuir* **16**, 9955–9958 (2000).

[75] M. Rief, M. Gautel, and H. E. Gaub, *Adv. Exp. Med. Biol.* **481**, 129–136 (2000).

[76] K. Watanabe, P. Nair, D. Labeit, M. S. Z. Kellermayer, M. Greaser, S.Labeit, and H. L. Granzier, *J. Biol. Chem.* **277**, 11549–11558 (2002).

[77] R. Law, P. Carl, S.Harper, P. Dalhaimer, D. W. Speicher, and D. E. Discher, *Biophys. J.* **84**, 533–544 (2000).

[78] P. F. Lenne, A. J. Raae, S. M. Altmann, M. Saraste, and J. K. H. Horber, *FEBS Lett.* **476**, 124–128 (2000).

[79] R. Law, P. Carl, S. Harper, D. W. Speicher, and D. E. Discher, *J. Biol. Chem.* **279**, 16410–16410 (2004).

[80] R. Law, G. Liao, S. Harper, G. Yang, D. Speicher, and D. E. Discher, *Biophys. J* **85**, 3286–3293 (2003).

[81] H. Li, A. F. Oberhauser, S. D. Redick, M. Carrion-Vazquez, H. P. Erickson, and J. M. Fernandez, *Proc. Natl. Acad. Sci. (USA)* **98**, 10682–10686 (2001).

[82] K. Ma, L. sing Kan, and K. Wang, *Biochem. J.* **40**, 3427–3438 (2001).

[83] A. Sarkar, S. Caamano, and J. M. Fernandez, *J. Biol. Chem* **280**, 6261–6264 (2005).

[84] D. J. Brockwell, E. Paci, R. C. Zinober, G. Beddard, P. D. Olmsted, D. A. Smith, R. N. Perham, and S. E. Radford, *Nat. Struct. Biol* **10**, 731–737 (2003).

[85] G. Lee, K. Abdi, Y. Jiang, P. Michaely, V. Bennett, and P. E. Marszalek, *Nature* **3**, 246–249 (2006).

[86] L. W. Li, S. Wetzel, A. Pluckthun, and J. M. Fernandez, *Biophys. J.* **90**, 30–32 (2006).

[87] P. Carl, C. H. Kwok, G. Manderson, D. W. Speicher, and D. Discher, *Proc. Natl. Acad. Sci. (USA)* **98**, 1565–1570 (2001).

[88] N. Bhasin, P. Carl, S. Harper, G. Feng, H. Lu, D. W. Speicher, and D. E. Discher, *J. Biol. Chem.* **279**, 45865–45874 (2004).

[89] I. Schwaiger, A. Kardinal, M. Schleicher, A. A. Noegel, and M. Rief, *Nat. Struct. Mol. Biol* **11**, 81–85 (2004).

[90] M. Schlierf and M. Rief, *J. Mol. Biol.* **345**, 497–503 (2005).

[91] M. Schlierf and M. Rief, *Biophys. J* **90**, 33–35 (2006).

[92] M. Carrion-Vazquez, A. F. Oberhauser, T. E. Fisher, P. E. Marszalek, H. Li, , and J. M. Fernandez, *Prog. Biophys. Mol. Biol.* **74**, 63–91 (2000).

[93] G. Yang, C. Cecconi, W. A. Baase, I. R. Vetter, W. A. Breyer, J. A. Haack, B. W. Matthews, F. W. Dahlquist, and C. Bustamante, *Proc. Natl. Acad. Sci. (USA)* **97**, 139–144 (2000).

[94] R. B. Best, B. Li, A. Steward, V. Daggett, and J. Clarke, *Biophys. J.* **81**, 2344–2356 (2001).

[95] M. Carrion-Vazquez, H. Li, H. Lu, P. E. Marszalek, A. F. Oberhauser, and J. M. Fernandez, *Nat. Struct. Biol.* **10**, 738–743 (2003).

[96] C. L. Chyan, F. C. Lin, H. Peng, J. M. Yuan, C. H. Chang, S. H. Lin, and G. Yang, *Biophys. J.* **87**, 3995–4006 (2003).

[97] H. Dietz and M. Rief, *Proc. Natl. Acad. Sci. (USA)* **101**, 16192–16197 (2004).

[98] H. Dietz and M. Rief, *Proc. Natl. Acad. Sci. (USA)* **103**, 1244–1247 (2006).

[99] D. J. Brockwell, S. Godfrey, S. Beddard, E. Paci, D. K. West, P. D. Olmsted, D. A. Smith, and S. E. Radford, *Biophys. J.* **89**, 506–519 (2005).

[100] E. Oroudjev, J. Soares, S. Arcidiacono, J. B. Thompson, S. A. Fossey, and H. G. Hansma, *Proc. Natl. Acad. Sci. (USA)* **30**, 6460–5465 (2002).

[101] S. Furuike and M. Yamazaki, *FEBS Lett.* **498**, 72–75 (2001).

[102] C. G. Baumann, V. A. Bloomfield, S. B. Smith, C. Bustamante, M. D. Wang, and S. M. Block, *Biophys. J.* **78**, 1965–1978 (2000).

[103] F. Oesterhelt, D. Oesterhelt, M. Pfeiffer, A. Engel, H. E. Gaub, and D. J. Muller, *Science* **288**, 143–146 (2000).

[104] H. Janovjak, M. Kessler, D. Oesterhelt, H. Gaub, and D. J. Muller, *EMBO J.* **22**, 5220–5229 (2003).

[105] D. Muller, W. Baumeister, and A. Engel, *Proc. Natl. Acad. Sci. (USA)* **96**, 13170–13174 (1999).

[106] J. Liphardt, B. Onoa, S. Smith, I. Tinoco, and C. Bustamante, *Science* **292**, 733–737 (2001).

[107] B. R. Brooks, R. E. Bruccoleri, B. Olafson, D. States, S. Swaminathan, and M. Karplus, *J. Comput. Chem.* **4**, 187 (1983).

[108] E. Lindahl, B. Hess, and D. van der Spoel, *J. Mol. Mod.* **7**, 306–317 (2001).

[109] A. Krammer, H. Lu, B. Isralewitz, K. Schulten, and V. Vogel, *Proc. Natl. Acad. Sci. (USA)* **96**, 1351–1356 (1999).

[110] A. Krammer, H. Lu, B. Isralewitz, K. Schulten, and V. Vogel, *Proc. Natl. Acad. Sci. (USA)* **96**, 1351–1356 (1999).

[111] M. Gao, M. Wilmanns, and A. K. Schulten, *Biophys. J.* **83**, 3435–3445 (2002).

[112] H. Lu and K. Schulten, *Biophys. J.* **79**, 51–65 (2000).

[113] H.Lu and K. Schulten, *Proteins* **35**, 453–463 (1999).

[114] M. Gao, D. Craig, O. Lequin, I. D. Campbell, V. Vogel, and K. Schulten, *Proc. Natl. Acad. Sci. (USA)* **100**, 14784–14789 (2003).

[115] D. Craig, A. Krammer, K. Schulten, and V. Vogel, *Proc. Natl. Acad. Sci. (USA)* **98**, 5590–5595 (2001).

[116] D. K. Klimov and D. Thirumalai, *Proc. Natl. Acad. Sci. (USA)* **97**, 7254–7259 (2000).

[117] E. Paci and M. Karplus, *Proc. Natl. Acad. Sci. (USA)* **97**, 6521–6526 (2000).

[118] E. Paci and M. Karplus, *J. Mol. Biol.* **288**, 441–459 (1999).

[119] P.-C. Li and D. E. Makarov, *J. Chem. Phys.* **114**, 4826–4832 (2004).

[120] M. Cieplak, T. X. Honag, and M. O. Robbins, *Phys. Rev. E.* **69**, 011912 (2004).

[121] M. Cieplak and P. E. Marszalek, *J. Chem. Phys.* **123**, 194903 (2005).

[122] R. C. Zinober, D. J. Brockwell, G. S. Beddard, A. W. Blake, P. D. Olmsted, S. E. Radford, and D. A. Smith, *Protein Sci.* **11**, 2759–2765 (2002).

[123] L. I. Schiff, PWN, 1977.

[124] W. Kolos, PWN, 1975.

[125] T. Schlick, *Molecular Modeling and Simulation. An Interdisciplinary Guide*, Springer-Verlag, New York, 2002.

[126] D. W. Herrmann, Wydawnictwo Naukowo-Techniczne, 1997.

[127] M. P. Allen and D. J. Tildesley, *Computer simulation of liquids*, Oxford University Press, New York, 1987.

[128] Y. Udea, H. Takatomi, and N. Go, *Biopolymeres* **17**, 1531–1548 (1978).

[129] M. Cieplak, T. X. Hoang, and M. O. Robbins, *Proteins: Struct. Funct. Bio.* **56**, 285–297 (2004).

[130] J. Skolinck and A. Koliński, *Science* **250**, 1121–1125 (1990).

[131] A. Godzik, J. Skolnik, and A. Koliński, *Proc. Natl. Acad. Sci. (USA)* **89**, 2629–2633 (1992).

[132] J. Vinals, A. Koliński, and J. Skolnick, *J. Mol. Biol.* **83**, 2801–2811 (2002).

[133] A. Kolinski and J. Skolnik, *Proteins* **32**, 475–494 (1998).

[134] A. Koliński, P. Rotkiewicz, B. Ilkowski, and J. Skolnik, *Prog. Teor. Phys. (Kyoto)* **138**, 292–300 (2000).

[135] A. Kolinski and J. Skolnik, *Polymer* **45**, 511–524 (2004).

[136] S. Miyazawa and R. L. Jernigan, *J. Mol. Biol.* **256**, 623–644 (1995).

[137] A. Kolinski, A. Godzin, and J. Skolnick, *J. Chem. Phys.* **103**, 4312–4322 (1995).

[138] T. Veitshans, D. Klimov, and D. Thirumalai, *Folding Des.* **2**, 1–22 (1997).

[139] G. Settanni, T. X. Hoang, C. Micheletti, and A. Maritan, *Biophys. J.* **83**, 3533–3541 (2004).

[140] M. Wojciechowski and M. Cieplak, *J. Phys. Cond. Mat.* **19**, 285218 (2007).

[141] D. K. West, P. Olmsted, and E. Paci, *J. Chem. Phys.* **124**, 154909–1 (2006).

[142] M. Cheung, A. E. Gracia, and J. Onuchic, *Proc. Natl. Acad. Sci. (USA)* **99**, 685–690 (2002).

[143] J. N. Onuchic, H. Nymeyer, A. E. Garcia, J. Chahine, and N. D. Socci, *Advan. Protein Chem.* **53**, 87–152 (2000).

[144] J. I. Kwiecinska and M. Cieplak, *J. Phys. Cond. Mat.* **17**, S1565–S1580 (2005).

[145] C. Clementi, M. Vandruscolo, A. Maritan, and E. Domany, *Proteins: Struct. Funct. and Genet.* **37**, 544–553 (1999).

[146] M. Cieplak, T. X. Hoang, and M. O. Robbins, *Proteins: Struct. Funct.and Genet.* **49**, 104–113 (2002).

[147] J. Tsai, R. Taylor, C. Chothia, and M. Gerstein, *J. Mol. Biol.* **290**, 253–266 (1999).

[148] M. Cieplak, A. Pastore, and T. X. Hoang, *J. Chem. Phys.* **122**, 054906 (2005).

[149] V. Sobolev, A. Sorokine, J. Prilusky, E. E. Abola, and M. Edelman, *Bioinformatics* **15**, 327–332 (1999).

[150] V. Munoz and W. A. Eaton, *Proc. Natl. Acad. Sci. (USA)* **96**, 11311–11316 (1999).

[151] J. Shimada, E. L. Kussell, and E. Shakhnovich, *J. Mol. Biol* **308**, 79–95 (2001).

[152] R. Srinivasan and G. D. Rose, *Proteins: Struct. Funct. and Genet.* **22**, 81–99 (1995).

[153] A. Kolinski and J. Skolnik, *Proteins* **18**, 338–352 (1994).

[154] W. Kabsch and C. Sander, *Biopolymers* **22**, 2577–2637 (1983).

[155] R. L. Jernigan and I. Bahar, *Curr. Opin. Struct. Biol.* **6**, 195–209 (1996).

[156] C. Micheletti, J. R. Banavar, A. Maritan, and F. Seno, *Phys. Rev. Lett.* **82**, 3372–3375 (1999).

[157] Y. Cao and H. LI, *Nat. Mater.* **6**, 109–114 (2007).

[158] M. Carrion-Vasquez, A. F. Oberhauser, S. B. Fowler, P. E. Marszalek, S. E. Broedel, J. Clarke, and J. M. Fernandez, *Proc. Natl. Acad. Sci. (USA)* **96**, 3694–3699 (1999).

[159] P. Szymczak and M. Cieplak, *J. Chem. Phys.* (2007).

[160] P. E. Marszalek, H. Lu, H. B. Li, M. Carrion-Vazquez, A. F. Oberhauser, K.Schulten, and J. M. Fernandez, *Nature* **402**, 100–103 (1999).

[161] C. D. Geierhaas, E. Paci, M. Vendruscolo, and J. Clarke, *J. Mol. Biol.* **343**, 1111–1123 (2004).

[162] S. J. Hamill, A. Steward, and J. Clarke, *J. Mol. Biol.* **297**, 165–178 (2000).

[163] I. Chang, M. Cieplak, R. I. Dima, A. Maritan, and J. R. Banavar, *Proc. Natl. Acad. Sci. (USA)* **98**, 14351–14355 (2001).

[164] Y. Z. Ye and A. Godzik, *Nucl. Acid. Res.* **32**, W582–W585 Suppl. 2 (2004).

[165] R. Metzler, W. Reisner, R. Riehn, R. Austin, J. Tegenfeldt, and I. Sokolov, *Europhys. Lett.* **76**, 696–702 (2006).

[166] A. Arend, R. Rzehak, and W. Zimmermann, *Europhys. Lett., w druku* .

[167] J. Barańska, *Ewolucja konformacji i parametrów węzłów zaciskanych na linie idealnej*, Politechnika Poznańska, Wydział Fizyki Technicznej, 2006.

[168] A. Mallam and S. Jackson, *J. Mol. Biol.* **346**, 1409–1421 (2004).

[169] A. Mallam and S. Jackson, *J. Mol. Biol* **359**, 1420–1436 (2006).

[170] S. Wallin, K. B. Zeldovich, and E. I.Shakhnovich, *J. Mol. Biol.* **368**, 884–893 (2007).

[171] M. T. Alam, T. Yamada, U. Carlsson, and A. Ikai, *FEBS Lett.* **519**, 35–42 (2002).

[172] S. Ohta, M. T. Alam, H. Arakawa, and A. Ikai, *Biophys. J.* **87**, 4007–40020 (2004).

[173] P.-G. de Gennes, *Macromolecules* **17**, 703–704 (1984).

[174] I. D. Kuntz, *J. Am. Chem. Soc.* **94**, 4009–4012 (1972).

[175] K. Yank, Wydanie III, Wydawnictwo Helion, 2005.

[176] E. Freeman and E. Freeman, Wydawnictwo Helion, 2007.

[177] M. Moncur, Wydanie IV, Wydawnictwo Helion, 2007.

[178] A. K. D. Gront, S. Kmiecik, *J. Comput. Chem.* **28(9)**, 1593–1597 (2007).

[179] O. Miyashita, J. Onuchic, and P. Wolynes, *Proc. Natl. Acad. Sci. (USA)* **100**, 12570–12575 (2003).

[180] P. Maragakis and M. Karplus, *J. Mol. Biol.* **352**, 807–822 (2005).

[181] K. Itoh and M. Sasai, *Proc. Natl. Acad. Sci. (USA)* **101**, 14736–14741 (2004).

[182] Y. Wang, A. Rader, I. Bahar, and R. Jernigan, *J. Struct. Biol.* **147**, 302–314 (2004).

[183] Y. Wang and R. Jernigan, *Biophys. J.* **89**, 3399–3409 (2005).

[184] N. Ranson, G. Farr, A. Roseman, B. B, W. Fenton, A. Horwich, and H. Saibil, *Cell* **107**, 869–879 (2001).

[185] F. Tama and Y. Sanejouand, *Protein. Eng.* **14**, 1–6 (2001).

[186] G. Song and R. Jernigan, *Proteins: Struct. Funct. and Bioinf.* **63**, 197–209 (2006).

[187] S. Kundu and R.L.Jernigan, *Biophys. J.* **86**, 3846–3854 (2004).

[188] P. Flory, *Proc. Roy. Soc. London Ser A Math Phys Sci* **351**, 351–380 (1976).

[189] A. Kloczkowski, J. Mark, and B. Erman, *Macromolecules* **22**, 1423–1432 (1989).

[190] I. Bahar, A. Atilgan, and B. Erman, *Fold. Des.* **2**, 173–181 (1997).

[191] A. Atilgan, S. Durell, R. Jernigan, M. Demirel, O. Keskin, and I. Bahar, *Biophys. J.* **80**, 505–515 (2001).

[192] S. Kundu, J. Melton, D. Sorensen, and G. Phillips, *Biophys. J.* **83**, 723–732 (2002).

[193] K. Hinsen, *Proteins: Struct. Funct. and Genet.* **33**, 417–429 (1998).

[194] F. Tama, F. Gadea, O. Marques, and Y. Sanejouand, *Proteins: Struc. Funct. and Genet.* **41**, 1–7 (2000).

[195] M. Lu and J. Ma, *Biophys. J.* **89**, 2395–2401 (2005).

[196] I. Bahar, A. Wallqvist, D. Covell, and R. Jernigan, *Biochemistry* **37**, 1067–1075 (1998).

[197] T. Haliloglu and I. Bahar, *Proteins: Struct. Funct. and Genet.* **37**, 654–667 (1999).

[198] N. Temiz, E. Meirovitch, and I.Bahar, *Proteins: Struct. Funct. and Bioinf.* **57**, 468–480 (2004).

[199] O. Keskin, I. Bahar, D. Flatow, D. Covell, and R. Jernigan, *Biochemistry* **41**, 491–501 (2002).

[200] O. Keskin, S. Durell, I. Bahar, R. Jernigan, and D. Covell, *Biophys. J.* **83**, 663–680 (2002).

[201] I. Navizet, R. Lavery, and R. Jernigan, *Proteins: Struct. Funct. and Genet* **54**, 384–393 (2004).

[202] D. Shortle and M. Ackerman, *Science* **293**, 487–489 (2001).

[203] A. Fersht, New Yokr, Freeman, 1998.

[204] Fersht and Sato, *Proc. Natl. Acad. Sci. (USA)* **101**, 7976–7981 (2004).

[205] J. R. Banavar, M.Cieplak, and A. Maritan, *Phys. Rev. Lett.* **93**, 238101 (2004).

[206] M. Doi and S. Edwards, Oxford University Press, 1986.

Podziękowania

Z największą przyjemnością pragnę podziękować mojemu promotorowi prof. dr. hab. Markowi Cieplakowi z Instytutu Fizyki PAN za całą przekazaną mi wiedzę, inspirujące i burzliwe dyskusje, które zawsze prowadziły do pogłębienia mojej wiedzy, nieustanną mobilizację, ale zarazem cierpliwość i wyrozumiałość. Chciałabym także podziękować za różnorodne możliwości wyjazdów na konferencje i staże naukowe, dzięki którym mogłam poszerzyć swoje horyzonty i zacząć rozwijać także swoje pomysły.

Chciałabym podziękować całemu Zespołowi Fizyki Biologicznej IFPAN. Po pierwsze dr. Hoangowi Trinh Xuan, za nieocenioną pomoc podczas stawiania pierwszych kroków na początku mojego doktoratu i zawsze miłe słowo. Składam podziękowania dr hab. Ewie Sobczak oraz dr. hab. Andrzejowi Sienkiewiczowi i dr Ani Niedźwiedzkiej za nieustanne wsparcie naukowe, wiele cennych wskazówek, a także motywowanie do pracy. Chciałabym także podziękować dr. Michałowi Wojciechowskiemu za cenne dyskusje naukowe, jak też krytyczne przeczytanie manuskryptu niniejszej pracy. Dziękuję mgr. Szymonowi Niewieczerzałowi za zawsze miłe słowo, optymizm, oraz bardzo miło spędzone ostatnie dwa lata mojego doktoratu. Chciałabym także podziękować mgr Izie Rutkowskiej i mgr. Wojtkowi Puławskiemu za serdeczną atmosferę i ciekawe rozmowy, które są niezbędnym uzupełnieniem pracy naukowej. Serdeczne podziękowania składam Bartkowi Witkowskiemu, z którym bardzo dobrze i owocnie współpracowaliśmy przy konstrukcji bazy danych.

Spośród pozostałych pracowników Istytutu Fizyki PAN ogromne podziękowania składam przede wszystkim dr. Pawłowi Janiszewskiemu i dr. Piotrowi Białkoziewiczowi. Chciałabym Wam podziękować za nieocenioną pomoc przy wykonaniu przeglądu białek, tworzeniu bazy danych, stron internetowych, oraz zawsze miłe, pogodne i ciepłe słowa. Chciałabym także podziękować dr hab. Krystynie Jabłońskiej za zawsze miłe słowo i wsparcie moich wszelkich wyjazdów. Składam podziękowania także całemu zespołowi SL1 za miłą naukową atmosferę podczas mojego doktoratu.

Chciałabym podziękowąć dr. Piotrowi Szymczakowi i dr. Piotrowi Sułkowskiemu z Wydziału Fizyki UW za wspólne badanie nieznanych dotąd zawęźlonych białek, niesamowicie cenne dyskusje naukowe, bardzo mile spędzany czas i nieustanne napawanie mnie optymizmen. Chciałabym także podziękować mgr. Janowi Sufczyńskiemu ze cenne uwagi na temat manuskryptu i nieustającą pogodę ducha.

Jestem wdzięczna dr Justynie Barańskiej z Politechniki Poznańskiej dzięki, której poznałam wiele nowych aspektów teorii węzłów.

Pragnę podziękować prof. dr hab. Andrzejowi Kolińskiemu z wydziału chemii UW za wprowadzenie mnie w tematykę bioinformatyki, udostępnienie klastra obliczeniowego bez którego niemożliwe byłoby wykonanie przeglądu białek, oraz wspacie związane z wyjazdem do Iowa State University, oraz dr Dominikowi Grontowi za zawsze chętne dokształcanie mnie w tematyce bioinformatyki i miłe dyskusje.

I would like to express my gratitude to prof. Robert Jernigan, dr hab. Andrzej Kloczkowski and dr. Taner Sen from The Jernigan Laboratory for introducing me to GNM theory and for great discussions during my stay at their laboratory. I would like to thank especially dr hab. Andrzej Kloczkowski for his great hospitality during my vistit at Ames. I spent there wonderful time.

Serdeczne podziękowania składam mojemu mężowi Piotrowi za wszelaką pomoc, cierpliowść, wyrozumiałość i także ogromne poświęcenie, dzięki któremu mogłam pisać tę pracę w miłej rodzinnej atmosferze. Chciałabym Mu także podziękować za krytyczne przeczytanie niniejszego manuskryptu. Ogromne podziękowania składam całej mojej rodzinie, której tę pracę dedykuję. Chciałabym podkreślić, iż w trakcie tych czterch lat było mi dane spotkać i współpracować z dużą liczbą osób, których nie sposób tutaj wymienić, tak więc do Was wszystkich kieruję słowa wdzięczności i podziękowania.

Badania przeze mnie prowadzone w ramach niniejszej pracy były sfinansowane przez grant promotorski Ministerstwa Nauki i Szkolnictwa Wyższego N202 021 31/0739.