

MAREK CIEPLAK, JOANNA I. SUŁKOWSKA

*Instytut Fizyki, PAN
Al. Lotników 32/46, 02-668 Warszawa
e-mail: sulek@piomar.com.pl*

ROZCIĄGANIE MOLEKUŁ BIAŁEK – PORÓWNYWANIE ICH WŁASNOŚCI MECHANICZNYCH

DLACZEGO ROZCIĄGANIE?

W ostatniej dekadzie pojawiły się możliwości technologiczne badania pojedynczych biomolekuł metodami optycznymi i mechanicznymi. Mechaniczne manipulowanie biomolekuł najczęściej realizowane jest poprzez zaczepienie jednego końca (lub innego fragmentu) molekuly na podłożu, a drugiego końca na ostrzu przymocowanym do elastycznej dźwigni mikroskopu siły atomowej i poprzez przesuwanie dźwigni ze stałą prędkością. W ten sposób zbadano różne rodzaje kwasów nukleinowych oraz ponad 55 białek. W wyniku pomiaru w mikroskopie siły atomowej uzyskuje się wykres siły w funkcji przemieszczenia ostrza. Gdyby ten wykres był liniowy, to świadczyłoby to o harmonicznym zachowaniu, jak w ponad 300-letnim prawie Hooke'a. Tymczasem wykresy takie są dużo bardziej skomplikowane i często mają postać zębatej krzywej o wielu maksimach. Kształt wykresu zawiera informacje o mechanicznych właściwościach białka. Wydobycie informacji o szczegółowych zdarzeniach mikroskopowych prowadzących do pojawienia się maksimum zazwyczaj wymaga interpretacji teoretycznej i symulacji. Wynika z nich, że duże maksima siły związane są z jednoczesnym zrywaniem wielu wiązań – najczęściej wodorowych. Silnie powiązane jednostki struktury noszą nazwę imadła mechanicznego.

Charakterystyczną miarę siły oporu białka przeciw rozciąganiu stanowi maksymalna siła $F_{\max}$. Jej wartości mieszczą się w zakresie od kilku do kilkuset pN (pikoniutonów). Na przykład, dla białka C2A $F_{\max}$ wynosi ona 60

pN, a dla superhelikalnej ankiry – 400 pN. Dla szczególnie dobrze badanej domeny I27 tytyny oraz ubikwityny $F_{\max}$ jest rzędu 200 pN. W tym artykule zadajemy pytanie, które białka są silne i dlaczego? Badania doświadczalne dają tylko bardzo fragmentaryczny ogólny „mechaniczny krajobraz” białek, a pełnoatomowe obliczenia chemiczne dotyczą jeszcze mniejszej liczby białek. W dodatku obliczenia te daje się przeprowadzić tylko dla krótkich, nanosekundowych skal czasu, co oznacza konieczność rozpatrywania niefizycznie dużych prędkości przesuwania ostrza, co zmienia charakter i wielkość sił. Zbadanie możliwych mechanicznych zachowań białek jest ważne dla ukierunkowania dalszych badań doświadczalnych oraz dla zrozumienia procesów mechanicznych zachodzących w komórkach. Niedawno udało się nam wyznaczyć wartości $F_{\max}$ dla 7510 białek i skorelować te wartości z atrybutami struktury. Użyte wyniki zostały wysłane do publikacji i niniejszy artykuł stanowi ich podsumowanie.

Należy dodać, że mikroskopu siły atomowej zaczęto ostatnio używać również do badania zwijania się białek (FERNANDEZ i LI 2004, SCHLIERF i współaut. 2004). W tym wypadku położenie ciągnącego ostrza dostosowuje się dynamicznie tak, żeby siła działająca na białko była stała. Najpierw białko jest rozciągane mechanicznie przy użyciu dużej siły, a następnie monitoruje się (skokowe zazwyczaj) zmiany długości białka po znacznym zredukowaniu siły powodującym zwijanie. Wykazaliśmy (SZYMCAK i CIEPLAK 2006),

$$E_p(\{\mathbf{r}_i\}) = V^{BB} + V^{NAT} + V^{NON} + V^{CHIR}.$$

W tym wzorze V^{BB} oznacza potencjał harmoniczny utrzymujący kolejne atomy C^α w pobliżu równowagowej odległości 3,8 Å. Kontakty natywne opisane są przez efektywne potencjały Lennarda-Jonesa:

$$V^{NAT} = \sum_{ij} 4 \varepsilon [(\sigma_{ij}/r_{ij})^{12} - (\sigma_{ij}/r_{ij})^6],$$

gdzie r_{ij} oznacza odległość pomiędzy C^α z aminokwasów i oraz j . Parametry długości σ_{ij} wybrane są tak, żeby położenia minimów potencjałów zgadzały się z odpowiadającymi odległościami podanymi w pliku PDB. Potencjał opisujący kontakty nienatywne, V^{NON} , odpowiada odpychającemu rdzeniowi o rozmiarze 4 Å. W najprostszym modelu parametr ε jest jednorodny. Jego wartość odpowiada średnim efektywnym oddziaływaniom niekovalencyjnym ($\varepsilon/k_B = 800-2300$ K, gdzie k_B jest stałą Boltzmanna). Model zawiera także człon, który faworyzuje przyjmowanie konformacji z natywnymi wartościami lokalnych chiralności (KWIECINSKA i CIEPLAK 2005):

$$V^{CHIR} = \sum_{i=2}^{N-2} 1/2 \varepsilon (C_i - C_i^{NAT})^2,$$

gdzie $C_i = (\mathbf{w}_{i-1} \times \mathbf{w}_i) \cdot \mathbf{w}_{i+1} / d_0^3$ i $\mathbf{w}_i = \mathbf{r}_{i+1} - \mathbf{r}_i$. C_i^{NAT} oznacza chiralność i -tego aminokwasu w stanie natywnym. Dodatnia wartość C_i odpowiada prawoskrętności. Człon V^{CHIR} pełni rolę podobną do członów z kątami dwusiecznymi i kątami wiązania w innych modelach.

Modele typu Go mają charakter fenomenologiczny i pozwalają w przybliżony sposób badać przebieg dużych zmian konformacyjnych w białkach. W wypadku zwijania się białek, wbudowana w model własność faworyzowania stanu natywnego nie jest zadawalająca pod względem pojęciowym. Nie jest to jednak problem w wypadku rozciągania, gdyż układ pozostaje w pobliżu stanu natywnego przynajmniej w początkowych stadiach

procesu. W symulacjach rozciągania oba końce białka przyczepione są do sprężyn o stałej sprężystości $k=0.12 \text{ } \varepsilon/\text{\AA}^2$, co jest bliskie właściwościom elastycznym dźwigni używanych w doświadczeniach, jeśli przyjąć 900 K za efektywną wartość ε/k_B . Przy takim wyborze ε proces zwijania jest optymalny w okolicach efektywnej temperatury pokojowej $0.3 \text{ } \varepsilon/k_B$ a jednostka siły, $\varepsilon/\text{\AA}$, odpowiada około 120 pN. Dla takich parametrów wykresy siły, F , w funkcji przemieszczenia, d , dla szeregu białek, np. dla 5 domen tytyny (CIEPLAK i współaut. 2005), czy ubikwityny (CIEPLAK i MARSZALEK 2005) jest bardzo podobny do wykresów uzyskanych doświadczalnie.

Swobodny koniec jednej ze sprężyn jest zakotwiczony zaś swobodny koniec drugiej sprężyny porusza się ze stałą szybkością v_p wzdłuż kierunku równoległego do linii łączącej końce białka w stanie natywnym. Obliczenia wykonujemy dla $v_p = 0.005 \text{ } \text{\AA}/\tau$, gdzie charakterystyczna skala czasu, τ , jest rzędu 0,25 ns (VEITSHANS i współaut. 1997, SZYMCHAK i CIEPLAK 2006) z uwagi na efekty związane z obecnością otaczającej wody. Taka wartość v_p jest więc tylko o dwa rzędy wielkości większa niż szybkość doświadczalna (wybrane białka można badać również dla właściwej wartości v_p). Efekty termiczne wprowadza się poprzez szum Langevina. Tak więc równania ruchu dla każdego atomu C^α mają postać:

$$m\ddot{\mathbf{r}}_i = -g\dot{\mathbf{r}}_i + \mathbf{F}_c + \boldsymbol{\Gamma}_i;$$

rozwiązuje się je przy pomocy algorytmu przewidywania-poprawiania piątego rzędu. $\mathbf{F}_c$ oznacza wypadkową siłę wynikającą z potencjałów opisujących układ. Stała tłumienia γ wynosi $2m/\tau$, a dyspersja sił przypadkowych $\boldsymbol{\Gamma}_i$ wynosi $\sqrt{2\gamma k_B T}$. Siły te naśladują wpływ roztworu wodnego.

TESTOWANIE MODELU

Rycina 1 przedstawia wykres zależności pomiędzy doświadczalnie wyznaczoną wartością $F_{\max}$ i wartością obliczoną w ramach naszego modelu w najprostszej wersji, w której parametr ε jest taki sam dla każdego kontaktu. Stwierdzamy istnienie przybliżonej korelacji o charakterze liniowym, co uzasadnia używanie modelu do porównywania ze sobą białek pod względem mechanicznym. Najlepsza zgodność teorii z doświadczeniem obserwowana jest dla białek o $F_{\max} > 170$ pN. Do zbioru takich białek należą białko L (BROCKWELL i współaut. 2005), domena I27 tyty-

ny (LI i współaut. 2002), tandemowe połączenia I27-I34, I54-I5, ubikwityna, i białko wołowej anhidrazy (ALAM i współaut. 2002). Nie jesteśmy w stanie badać ankiryiny o 24 podjednostkach z uwagi na brak odpowiadającej struktury w PDB. „Stalowe” właściwości tego białka muszą jednak wynikać z ułożenia podjednostek w naturalną podkowę. Złożenia mniejszych liczb podjednostek nie generują kształtu podkowy, co prowadzi do dużo mniejszych sił i w naszym modelu, i w doświadczeniu (LEE i współaut. 2006, LI i współaut. 2006). W szczególności dla złoże-

nia 12 podjednostek $F_{\max}$ jest bliskie wartości uzyskanej dla domeny I27 tytyny.

Do zbioru białek o średniej wartości siły $F_{\max}$ (100–170 pN) należą E2lip3(N-41) (BROCKWELL i współaut. 2003), gdzie wyrażenie w nawiasach oznacza ciągnięcie za terminal N i aminokwas 41, fibronektyna 12 FnIII, i domena I1 tytyny (symbol struktury 1gc1 w PDB). Z tego zbioru najlepiej z trendem liniowym zgadzają się dane dla pierwszego białka zaś dla domeny I1 tytyny uzyskujemy siłę nieco większą niż dla domeny I27 zamiast nieco mniejszą. Do zbioru słabszych białek, o $F_{\max}$ pomiędzy 60 i 100 pN, należą ubikwityna(48-C) (CARRION-VAZQUEZ i współaut. 2003), lizozym T4 (YANG i współaut. 2000), barnaza (BEST i współaut. 2001), pierwsza domena C2 białka synaptotagminu I (C2A; CARRION-VAZQUEZ i współaut. 2000), kalmodulina (CARRION-VAZQUEZ i współaut. 2000), 10-ta i 13-ta domeny fibronektyny, pięć kolejnych domen białka DdFLN, oraz osobno jego czwarta domena 4FLN. Dla tych słabszych białek korelacja z liniowym trendem jest słabsza, przy czym największe odstępstwa są dla 4FLN (SCHWAIGER i współaut. 2004) i lizozymu T4. Najgorzej testy wypadają dla najsłabszych białek ($F_{\max} < 60$ pN): rodzina spektryny (LAW i współaut. 2003), pojedyncza podjednostka ankiry, rybonukleaza H, i E2lip3. Dla tych białek $F_{\max}$ wyznaczone teoretycznie jest na ogół zbyt duże.

Ulepszenia modelu mogą iść w dwóch kierunkach, albo poprzez uwzględnienie dy-

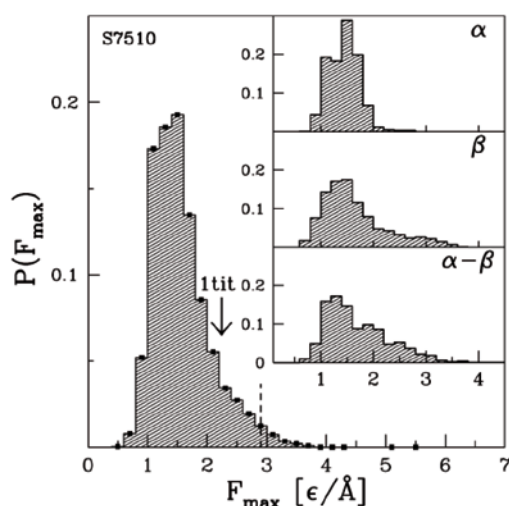
namiki grup bocznych, albo poprzez wprowadzenie niejednorodnych wartości parametru ϵ według odpowiednio uzasadnionego schematu. Ten pierwszy kierunek można zrealizować dołączając do opisu węgle C^β . Okazało się, że nie prowadzi to do jakościowych zmian na analogu Ryc. 1. Inne podejście polega na pozostawieniu opisu opartego na C^α , ale przy powiązaniu wartości parametru ϵ z rodzajami i wielkością atomowych przekrywań badanych przy określaniu jakie pary aminokwasów tworzą kontakty natywne. W tym artykule przedstawimy wyniki uzyskane dla jednorodnej wartości ϵ , gdyż obliczenia dla ulepszonego modelu są w trakcie realizacji.

Należy zaznaczyć, że określenie wartości $F_{\max}$ z danych doświadczalnych jest często utrudnione poprzez fakt, że wykresy F - d zazwyczaj uzyskuje się dla kilku modułów połączonych w tandem, często przy użyciu sztucznych linkerów. Te moduły nie muszą być jednakowe ani też nie muszą się one rozwijać jeden po drugim. Tak więc samo wyznaczenie $F_{\max}$ oparte jest na pewnych założeniach interpretacyjnych.

Innym rodzajem testu naszego podejścia jest badanie wyciągania bakteriorodopsyny z membrany. Podejście typu Go prowadzi do skomplikowanego wykresu F w funkcji d (CIEPLAK i współaut. 2006; ryc. 10), który jest bardzo podobny do wykresu wyznaczonego doświadczalnie (JANOVJAK i współaut. 2003) po przeskalowaniu parametru ϵ .

WYNIKI PRZEGLĄDU BIAŁEK Z PDB

Obecnie przedstawimy wyniki dotyczące wartości $F_{\max}$ dla białek ze zbioru S7510. Ryc. 2 przedstawia rozkład wartości $F_{\max}$



uzyskanych w ramach naszego modelu. Wartości te są w zakresie od 0 do 5,44 i najbardziej prawdopodobne są rzędu 1,5 $\epsilon/\text{\AA}$. Typowy błąd związany ze zmiennością pomiędzy poszczególnymi trajektoriami jest rzędu 0,10–0,15. Dla tytinowej domeny I27 $F_{\max}$ wynosi 2,2 $\epsilon/\text{\AA}$, czyli mniej więcej połowę wartości maksymalnej w S7510. Stwierdzili-

Rycina 2. Rozkład prawdopodobieństwa wartości $F_{\max}$ dla białek ze zbioru S7510.

Kwadratowe symbole dla dużych sił odpowiadają pojedynczym białkom. Strzałka pokazuje wartość $F_{\max}$ dla domeny I27 tytyny. Linia przerywana zaznacza próg, powyżej którego białka należą do zbioru S137 najsilniejszych białek. Panel wewnątrz rysunku przedstawia rozkłady $F_{\max}$ dla klas strukturalnych α , β i α/β . Maksima wszystkich rozkładów są w pobliżu 1,5 $\epsilon/\text{\AA}$.

Tabela 1. 137 najsilniejszych białek według przewidywań najprostszej wersji modelu używanego w tym artykule.

Kolejność uporządkowania białek o tej samej wartości $F_{\max}$, w ramach błędu obliczenia, jest przypadkowa. Uwzględnienie mostków siarczkowych awansuje 1rbj i 1rnz na tej liście oraz powoduje usunięcie 1lsl.

n	PDB	F	λ	CATH	n	PDB	F	λ	CATH	n	PDB	F	λ	CATH
1	1c4p	5.4	14	3.10.20	47	1kgi	3.2	21	2.60.40	93	1k26	3.0	46	3.90.79
2	1qqr	5.1	14	3.10.20	48	1oau	3.2	3	2.60.40	94	1lqb	3.0	29	3.10.20
3	1aoh	4.2	1	2.60.40	49	1vhp	3.2	2	2.60.40	95	1tbe	3.0	6	3.10.20
4	1glk	4.2	2	2.60.40	50	1h8c	3.2	6	3.10.20	96	1gb4	3.0	4	3.10.20
5	1ssn	3.8	33	3.10.20	51	1jfh	3.2	312	3.40.50	97	1rbj	3.0	102	3.10.130
6	1ie5	3.8	14	2.60.40	52	1jrk	3.2	46	3.90.79	98	1tfp	3.0	16	2.60.40
7	1c76	3.7	11	3.10.20	53	1sn5	3.2	17	2.60.40	99	1tyr	3.0	32	2.60.40
8	1ppx	3.7	55	3.90.79	54	1wtl	3.2	3	2.60.40	100	1ves	3.0	2	
9	1c77	3.6	20	3.10.20	55	1amx	3.1	37	2.60.40	101	1vfb	3.0	3	2.60.40
10	1c79	3.6	21	3.10.20	56	1qd0	3.1	2	2.60.40	102	1lsl	3.0	120	
11	1lyn4	3.6	2		57	1ufy	3.1	12	3.30.1330	103	1hz5	3.0	23	3.10.20
12	1c78	3.6	21	3.10.20	58	1mg4	3.1	36	3.10.20	104	1py9	3.0	1	2.60.40
13	2sak	3.6	11	3.10.20	59	1p7e	3.1	2	3.10.20	105	1fmf	3.0	148	3.40.50
14	1v5o	3.6	25		60	1j05	3.1	3	2.60.40	106	2try	3.0	32	2.60.40
15	1sp0	3.6	16		61	1jhl	3.1	4	2.60.40	107	1lve	3.0	3	2.60.40
16	1so9	3.5	16		62	1bmz	3.1	16	2.60.40	108	1gke	3.0	24	2.60.40
17	1sn0	3.5	17	2.60.40	63	1bvk	3.1	7	2.60.40	109	1etb	3.0	21	2.60.40
18	1oo2	3.5	19	2.60.40	64	1c08	3.1	4	2.60.40	110	1i8k	3.0	3	2.60.40
19	1i3v	3.5	2	2.60.40	65	1oar	3.1	3	2.60.40	111	1lct	3.0	19	2.60.40
20	1i9e	3.5	4	2.60.40	66	1tte	3.1	32	2.60.40	112	1pge	3.0	10	2.40.40
21	1nam	3.5	3	2.60.40	67	1pun	3.1	55	3.90.79	113	1gnu	3.0	53	3.10.20
22	1eaj	3.5	5	2.60.40	68	1pus	3.1	54	3.90.79	114	1kot	3.0	54	3.10.20
23	1kiq	3.4	3	2.60.40	69	1wiu	3.1	3	2.60.40	115	1ui9	3.0	8	3.30.1330
24	1f5w	3.4	5	2.60.40	70	1gko	3.1	17	2.60.40	116	1w29	3.0	54	
25	2ncm	3.4	4	2.60.40	71	1n4x	3.1	5	2.60.40	117	1oax	3.0	3	2.60.40
26	1pgx	3.4	9	3.10.20	72	1nvi	3.1	4	3.10.20	118	1qp1	2.9	3	2.60.40
27	1m94	3.4	4	3.10.20	73	1lvc	3.1	4	2.60.40	119	1bz8	2.9	27	2.60.40
28	1anu	3.4	2	2.60.40	74	1ugm	3.1	38		120	1mel	2.9	1	2.60.40
29	1eta	3.4	31	2.60.40	75	1lgl	3.1	5	3.10.20	121	1f2x	2.9	2	2.60.40
30	1kip	3.4	3	2.60.40	76	1lil	3.1	5	2.60.40	122	1em7	2.9	3	3.10.20
31	1sn2	3.4	17	2.60.40	77	1w19	3.1	52		123	1com	2.9	6	3.30.1330
32	1tum	3.4	56	3.90.79	78	1kir	3.1	3	2.60.40	124	1lm8	2.9	36	3.10.20
33	1nme	3.4	90	3.40.50	79	1kmt	3.1	6	2.70.50	125	1dvy	2.9	17	2.60.40
34	1h5b	3.3	6	2.60.40	80	1l2n	3.1	6	3.10.20	126	1f86	2.9	19	2.60.40
35	1npu	3.3	3		81	1dfu	3.1	46	2.40.240	127	1rnz	2.9	104	3.10.130
36	1mvf	3.3	8	2.60.40	82	2rox	3.1	20	2.60.40	128	1tjn	2.9	96	
37	43c9	3.3	3	2.60.40	83	1oag	3.1	2	2.60.40	129	1v80	2.9	6	
38	1hz6	3.3	8	3.10.20	84	1rlf	3.1	4	3.10.20	130	1vjg	2.9	33	
39	1bzd	3.3	32	2.60.40	85	1tvd	3.1	2	2.60.40	131	1wit	2.9	4	2.60.40
40	1a2y	3.3	3	2.60.40	86	2dlf	3.1	3	2.60.40	132	1mfw	2.9	46	3.10.20
41	1km7	3.3	21	3.10.20	87	2imm	3.1	3	2.60.40	133	1ic4	2.9	3	2.60.40
42	1lve	3.3	3	2.60.40	88	1pga	3.1	2	3.10.20	134	2igd	2.9	5	3.10.20
43	1b88	3.3	3	2.60.40	89	1b9r	3.0	2	3.10.20	135	1lve	2.9	4	2.60.40
44	1eo6	3.2	45	3.10.20	90	1pav	3.0	35		136	1ieh	2.9	16	2.60.40
45	1ie4	3.2	21	2.60.40	91	1i3o	3.0	90	3.40.50	137	1dvt	2.9	18	2.60.40
46	1k53	3.2	24	3.10.20	92	1bm7	3.0	15	2.60.40		1tit	2.1	4	2.60.40

śmy, że im większa wartość N, tym większe prawdopodobieństwo uzyskania dużej siły (tu nieudokumentowane a potwierdzone przez symulacje 239 białek o N pomiędzy 153 i 851). Jednak dla każdej wartości N istnieje znaczny rozrzut w wartościach $F_{\max}$ pomiędzy białkami.

W Tablicy 1 przedstawiamy 137 białek (1,8% zbioru S7510), które według naszych obliczeń powinny być najsilniejsze, tzn. o $F_{\max} > 2,9 \text{ e}/\text{\AA}$. Tabela daje listę kodów struktur PDB, wartości N, $F_{\max}$, względne położenie, λ , głównego maksimum, oraz symbol strukturalnej klasyfikacji CATH, jeśli jest dostępny. Parametr $\lambda = (L_m - L_n) / (L_f - L_n)$ jest zdefiniowany

poprzez charakterystyczne wartości odległości L pomiędzy końcami białka: L_n jest wartością natywną, L_m odpowiada położeniu maksimum, a $L_f = (N-1)3,8 \text{ \AA}$ wartości odpowiadającej pełnemu rozwinięciu. Stwierdziliśmy, że rozkład wartości λ w zbiorze S137 ma ostre maksimum przy około 10% choć w zbiorze S7510 rozkład jest dosyć płaski. Oznacza to, że istnienie dużych sił najczęściej się wiąże z pęknięciami zachodzącymi w pobliżu końców białka, tak jak to ma miejsce dla tytyny (LU i współaut. 1998, CIEPLAK i współaut. 2004). Stwierdzamy też, że w wypadku 71% silnych białek na wykresie F-d występuje duże maksimum, po którym następuje słabsze maksimum. Dla 7% słabsze maksimum występuje przed głównym, dla 20% jest kilka maksimum przed maksimum głównym i przynajmniej jedno po, a tylko trzy białka, w tym dwa najsilniejsze, mają tylko jedno maksimum siły.

Białka wymienione w Tabeli 1 niekoniecznie są od siebie odmienne biologicznie, gdyż kilka różnych struktur może być przypisanych do tego samego białka. Dynamika

rozciągania jest wrażliwa na szczegóły strukturalne i stąd na wybór określonego kodu. Stwierdziliśmy, że ze zbioru S137 43 białka nie są ze sobą homologicznie powiązane, natomiast pozostałych 93 należy do 32 różnych grup. W szczególności dwie najsilniejsze struktury 1c4p i 1qqr odpowiadają temu samemu białku – β -domenie streptokinazy (o topologii zwoju UB), która uczestniczy w mechanizmie skrzepliwosci krwi – ale skryzalizowanemu w różnych warunkach. Trzecie najsilniejsze białko, 1aoh, jest w tej samej grupie, co 28-e 1anu – jest to domena kohezyny celulosomu uzyskana z *C. thermocellum*.

ZALEŻNOŚĆ $F_{\max}$ OD TYPU STRUKTURY W ZBIORZE S3813

Schemat klasyfikacji CATH dzieli struktury na kategorie w sposób hierarchiczny: najpierw na klasy (C), potem na architektury (A), topologie (T) i wreszcie na homologię (H). Odpowiada temu podzielony na fragmenty kod numeryczny, czego przykłady widać w Tabeli 1. Rozkład $F_{\max}$ dla S3813 jest niemal taki sam, jak dla S7510. Istnieją jednak znaczące różnice w rozkładach odpowiadających poszczególnym klasom (panel wewnątrz Ryc. 2). Rozkłady dla białek z klas β i α/β posiadają spore ogony w obszarze dużych sił. Tak nie jest dla klasy α , co oznacza, że nie należy oczekiwać, że białka z tej klasy będą mogły mieć duże wartości $F_{\max}$. Potwierdzeniem tego są np. pomiary siły dla polikalmoduliny (CARRION-VAZQUEZ i współaut. 2000). (Nasze obliczenia przewidują, że pewne duże wielodomenowe białka z domenami typu α – 1ile, 2ng1, 1ciy, i 1cii – mogą objawiać duży opór na rozciąganie z uwagi na ruchy ścinające pomiędzy domenami).

Rozważmy teraz dokładniejsze charakterystyki struktury. 80% białek klasy α w S3813 ma architekturę wiązki ortogonalnej. W klasie β 36% to beczułki, 31% kanapki, 13% wstażki, i 13% rolki. W klasie α/β 40% to rolki, a 39% to dwu-warstwowe kanapki. Te rozkłady zmieniają się znacznie, gdy rozważyć zbiór S137 najsilniejszych białek. Żadne z nich nie jest z klasy α . Białka z S137 należą do sześciu architektur. Dwie z nich są szczególnie dobrze reprezentowane: kanapki β (60%) i rolki α/β (30%). Te silne białka należą również

do dziewięciu klas topologicznych. Najczęstsze topologie to immunoglobuliny (2.60.40 w schemacie CATH) i ubikwitynopodobne rolki UB (3.10.20). Pozostałe występujące tu kody to 3.90.79, 3.40.50, 3.10.130, 3.10.1330, 3.10.50, 2.40.40, i 2.40.240. Rozkład $F_{\max}$ dla architektury kanapki β jest w S3813 szeroko rozciągnięty i ma trzy lokalne maksima. Maksimum przy większych siłach odpowiada topologii immunoglobuliny. Na to maksimum składają się głównie homologi białek transportu i immunoglobulin. W wypadku architektury α/β rolki rozkład sił ma dwa maksima. Maksimum przy większych siłach odpowiada topologiom białek ubikwitynopodobnych oraz białek P-30.

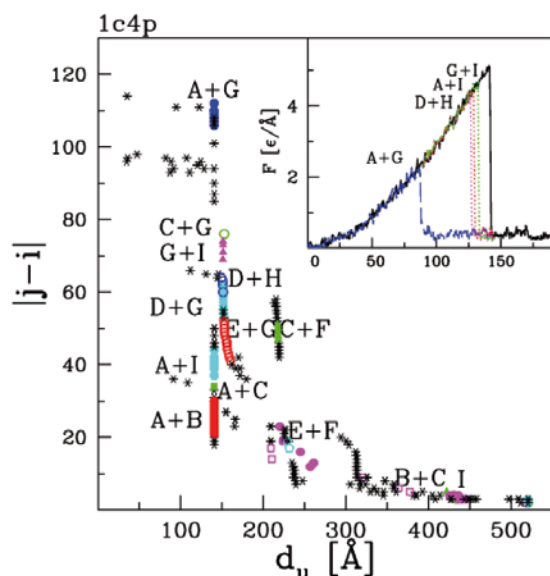
Choć siła białka i jego struktura są ze sobą skorelowane, znaleźliśmy przykłady, w których zadany kod CATH odpowiada dwóm odmiennym zachowaniom dynamicznym, co sugeruje istnienie „niedoróbek” w tym schemacie. Na przykład silne białka 1pga i 1p72 (52-gie i 56-e w Tabeli 1) mają ten sam kod CATH, co słabsze o czynnik 3 1mpe i 1q10. Ten kod to 3.10.20.10, co odpowiada architekturze rolka α/β , topologii ubikwitynopodobne i homologii wiążące immunoglobulinę. Te cztery białka mają $N=56$ i białka słabe różnią się od silnych o mutacje tylko w trzech miejscach (i stąd o promień RMSD 1,9 Å). Jednak mutacje te eliminują z silnych białek kontakty długozasięgowe, co znacznie redukuje siłę oporu na rozciąganie.

DLACZEGO PEWNE BIAŁKA SĄ SILNE?

Mechanizmy rozwijania można wyjaśnić poprzez tzw. diagramy scenariuszowe (CIEPLAK i współaut. 2004), które pokazują przy jakim przesunięciu d_u zadany kontakt jest zerwany po raz ostatni. Kontakt identyfikuje się poprzez odpowiadającą mu odległość sekwencyjną $|j-i|$. Mówimy, że kontakt ten jest zerwany, gdy $r_{ij} > 1.5 \sigma_{ij}$. Zgrupowanie się zrywania kontaktów w pobliżu pewnej wartości d_u oznacza pojawienie się w tym miejscu maksimum siły. Diagram scenariuszowy dla najsilniejszego białka 1c4p pokazany jest na Ryc. 3. Białko to składa się z czterech łańcuchów o podobnych właściwościach i Ryc. 3 dotyczy pierwszego z nich. Łańcuch ten zawiera α helisę (196–210), oznaczona symbolem I, i 8 nici β oznaczonych literami od A

do H, poczynając od końca N. Z Ryc. 3 wnioskujemy, że w pobliżu $d=140$ Å pęka niemal jednocześnie wiele kontaktów. Które z nich odpowiadają „imałdu mechanicznemu”, które trzyma białko najsilniej? Odpowiedzialne kontakty można zidentyfikować teoretycznie drogą usuwania poszczególnych kontaktów z interesującego zestawu i sprawdzania, jaki to ma wpływ na wykresy F w funkcji d . Z panelu w Ryc. 3 wynika, że sercem imadła są kontakty pomiędzy nicią A (segment 158–168) i równoległą G(266–278). Usunięcie tych kontaktów redukuje bowiem $F_{\max}$ o 50%.

Okazuje się, że w zbiorze S137 94% imadeł mechanicznych zbudowanych jest z długich równoległych nici β , które zrywane są przez ruchy ścinające (LU i współaut.

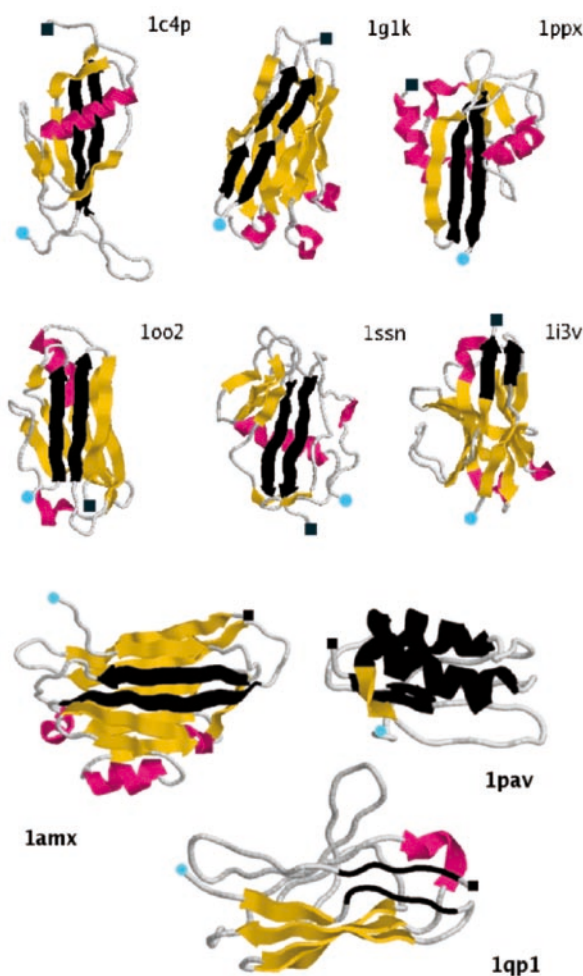


Rycina 3. Scenariusz rozwijania dla białka 1c4p (dla temperatury $T=0,3 \text{ } \epsilon/k_B$).

Symbole literowe wskazują, w jakich strukturach drugorzędowych zrywane są kontakty na odległości d_u . Symbole oznaczone gwiazdkami odpowiadają kontaktom spoza struktur drugorzędowych. Panel wewnątrz rysunku przedstawia wykresy siły w funkcji przemieszczenia. Linia ciągła dotyczy sytuacji, w której wszystkie kontakty są obecne. Pozostałe krzywe ilustrują wpływ usuwania kontaktów, np. gdy usunięte są kontakty między niemi β A i G.

1998, LI i MAKAROV 2004, WEST i współaut. 2006). Przykłady takich imadeł zaznaczone są na czarno na górnych sześciu panelach na Ryc. 4. Największe siły pojawiają się wtedy, gdy choć jedna z tych nici jest blisko punktu ciągnięcia. Imadło mechaniczne może być poddane ścinaniu na początku rozciągania. Jeśli jednak oba końce białka są w stanie natywnym po tej samej stronie białka, to najpierw musi nastąpić odwiniecie fragmentów otaczających. Powoduje to pojawienie się obrotu i mniejszych maksimum siły przed maksimum najwyższym. Siła imadła zależy przede wszystkim od liczby kontaktów (wiązań wodorowych) wewnątrz imadła a następnie od istnienia stabilizujących oddziaływań otaczających imadło.

Istnieją też imadła o innej strukturze. Trzy dolne panele na Ryc. 4 pokazują ich



Rycina 4. Wstążkowa reprezentacja struktury wybranych silnych białek.

Elementy strukturalne zaznaczone na czarno odpowiadają imadłu mechanicznemu, które generuje największy opór przeciwko rozciąganiu. Dla białka 1g1k nici A z imadła składają się z osobnych segmentów.

przykłady, jakie realizowane są w białkach 1amx, 1qp1, i 1pav. W pierwszym z nich (a także w 1ei5 i 1lm8) nici imadła są antyrównoległe do siebie. W drugim z nich (i również w 1tum i 1f86) nici nie mają struktury i nie tworzą kartki β . Wreszcie w trzecim z nich imadło jest utworzone przez „pudełko” zbudowane z dwóch antyrównoległych nici β umieszczonych tuż przy dwóch antyrównoległych helisach α . Wszystkie elementy tego pudełka trą względem siebie w sposób ściągający, gdy białko jest rozciągane.

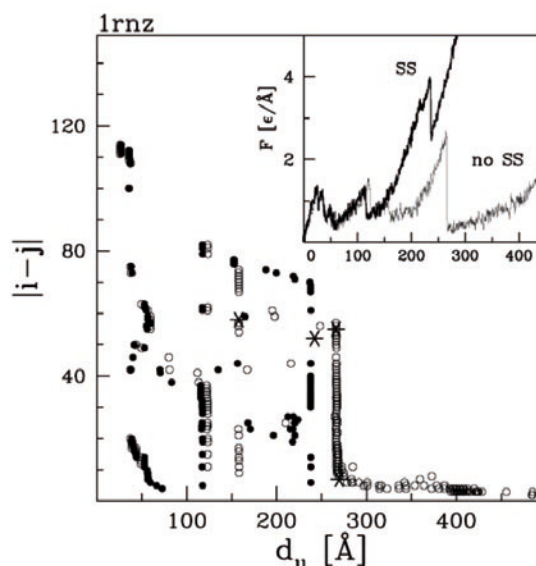
ROLA MOSTKÓW SIARCZKOWYCH

W modelu używanym do przeglądu wartości $F_{\max}$ dla białek z PDB nie są wyróżniane kontakty odpowiadające mostkom

siarczkowym (SS) pomiędzy cysteinami, choć takie wiązania, o naturze kowalencyjnej, nie podlegają zrywaniu w zakresie sił

dostępnych w mikroskopie siły atomowej. Problem ten ilustruje Ryc. 5 dla wołowej rybonukleazy A o kodzie struktury 1rnz. W Tabeli 1 jest to białko 127-me w rankingu siły. Prosty model dynamiczny prowadzi do $F_{\max}=2.9 \text{ } \epsilon/\text{\AA}$ przy czym maksimum to pojawia się w pobliżu $d=250 \text{ } \text{\AA}$. W białku tym są cztery kontakty SS, z których dwa pękają – w prostym modelu – przed $F_{\max}$, a dwa pozostałe dają wkład do $F_{\max}$. Można zapobiec temu niefizycznemu zrywaniu się tych kontaktów poprzez zwiększenie parametru energetycznego ϵ o np. czynnik 20. Prowadzi to do wcześniejszego pojawienia się największego maksimum siły oraz do podwyższenia $F_{\max}$ do wartości $4,0 \text{ } \epsilon/\text{\AA}$. Tym samym białko to awansuje do piątego miejsca w rankingu siły. Na podstawie diagramu scenariuszowego stwierdzamy, że pierwsza połowa poprawionego procesu przebiega tak, jak w prostym modelu, ale potem przebieg jest zupełnie odmienny i imadło mechaniczne jest skonstruowane z innych wiązań. Podobna sytuacja występuje dla homologicznie pokrewnego białka 1rbj, dla którego $F_{\max}$ skacze z 3,0 do $3,5 \text{ } \epsilon/\text{\AA}$, zmieniając setną pozycję rankingową na 20-tą.

Sama obecność wiązań SS nie musi jednak wpływać na wartość $F_{\max}$, zwłaszcza, gdy pęknięcie takiego wiązania odbywa się po osiągnięciu $F_{\max}$. W zbiorze S137 jest 14 białek, oprócz 1rnz i 1rbj, dla których prosty model przewiduje pęknięcie wiązania SS przed głównym maksimum siły. Dokonaliśmy ponownej analizy takich białek i stwierdziliśmy albo brak zmiany $F_{\max}$ (1ie5, 1i8k, 1kiq, 1lve), albo mały wzrost (1py9, 43c9), albo małe zmniejszenie (1h5b, 1c08, 1ivl, 1kir, 1kip, 1n4x, 1vfb). Jedyny przypadek dużej zmiany $F_{\max}$ z powodu obecności wiązań SS to białko 1lsl. Białko to ma sześć wiązań SS i jedno z nich efektywnie ogranicza rozciąganie



Rycina 5. Scenariusz rozwijania oraz wykres F w funkcji d , na panelu wewnętrznym, dla białka 1rnz w ramach dwóch modeli typu Go.

Pierwszy model jest standardowy i nie wyróżnia mostków siarczkowych. Odpowiadające mu krzywe są cieńsze, a symbole na scenariuszu zaznaczone są przez otwarte symbole i gwiazdy. Gwiazdy odpowiadają kontaktom między cysteinami. Drugi model nie pozwala na pęknięcie mostków siarczkowych. Odpowiadają mu grubiej wykreślone krzywe $F(d)$ i czarne kółka.

ganie do 37-aminokwasowego segmentu całego białka o $N=113$. W tym segmencie nie ma żadnego maksimum siły, co powoduje, że białko 1lsl powinno być usunięte z Tabeli 1. Porównywanie rozciągania z wiązaniami SS i bez nich ma sens fizyczny, gdyż wiązania SS można zamienić na znacznie słabsze wiązania SH (grupa hydrosulfidowa) poprzez zastosowanie czynnika redukującego DTT (1,4 ditiotretol). Tego typu badania doświadczalne przeprowadzono np. dla molekuly adhezji Mel-CAM (CARL i współaut. 2001).

WNIOSKI KOŃCOWE

Prosty model typu Go pozwala lepiej zrozumieć mechaniczne właściwości białek i pozwala przewidywać, które białka najbardziej opierają się rozciąganiu. Te silne białka należą do niewielkiej liczby klas topologicznych. Ich siła wynika z istnienia imadła mechanicznego, które w większości przypadków zbudowane jest z długich i równoległych nici β . Uwzględnienie w modelu grup bocznych i mostków siarczkowych powoduje przesunięcia

w rankingu białek względem siły $F_{\max}$, ale w zasadzie nie zmienia samej listy silnych białek. Sugeruje to, że opór białka na rozciąganie zawarty jest głównie w jego geometrii i w związku z tym doświadczenia i obliczenia pełnoatomowe powinny w większości potwierdzić te listę.

Ten projekt powstał w wyniku rozmowy z Julio M. Fernandezem z Columbia Univer-

sity. Cenne były uwagi uzyskane od Jayantha R. Banavara, Trinh Xuan Hoanga, Harald Ja-novjaka, i Piotra Szymczaka.

Ten projekt był finansowany przez Mini-sterstwo Nauki i Informatyzacji (numer gran-tu N202 021 31/0739).

STRETCHING OF MOLECULES OF PROTEINS – COMPARISON OF THEIR MECHANICAL PROPERTIES

Summary

Mechanical stretching of single proteins has been studied experimentally for about 55 proteins yielding a variety of force patterns and peak forces.

Here, we perform a theoretical survey of proteins of known native structure and map out the landscape of possible dynamical behaviors under stretching at a constant speed. We consider 7510 proteins each comprising not more than 150 amino acids. The model used is constructed based on the native geometry. It is solved by methods of molecular dynamics and validated by comparing the theoretical predictions to experimental results. We characterize the distribution of peak forces and cor-

relations with the system size and with the structure classification as characterized by the hierarchical classification of protein domain structures (CATH). Despite the presence of such correlations, proteins with the same CATH index may belong to different classes of the dynamical behavior. We identify proteins with the biggest forces and show that they belong to few topology classes. We determine which protein segments act as mechanical clamps and show that, in most cases, they correspond to long stretches of parallel β -strands, but other mechanisms are also possible.

LITERATURA

- ABE H., GO N., 1981. *Noninteracting local-structure model of folding and unfolding transition in globular proteins. II. Application to two-dimensional lattice proteins*. Biopolymers 20, 1013–1031.
- ALAM M.T., YAMADA T., CARLSSON U., IKAI A., 2002. *The importance of being knotted: effects of the C-terminal knot structure on enzymatic and mechanical properties of bovine carbonic anhydrase*. FEBS Letters 519, 3–40.
- BERMAN H. M., WESTBROOK J., FENG Z., GILLILAND G., BHAT T. N., WEISSIG H., SHINDYALOV I. N., BOURNE P. E., 2000. *The Protein Data Bank*. Nucleic Acids Res. 28 235–242.
- BEST R. B., LI B., STEWARD A., DAGGETT V., CLARKE J., 2001. *Can non-mechanical proteins withstand force? Stretching barnase by atomic force microscopy and molecular dynamics simulation*. Biophys. J. 81, 2344–2356.
- BROCKWELL D. J., PACI E., ZINOBER R. C., BEDDARD G. S., OLMSTED P., SMITH D. A., PERNHAM R. N., RADFORD S. E., 2003. *Pulling geometry defines mechanical resistance of β -sheet protein*. Nat. Struct. Biol. 10, 731–737.
- BROCKWELL D. J., GODFREY S., BEDDARD G. S., PACI E., WEST D. K., OLMSTED P. D., SMITH D. A., RADFORD S. E., 2005. *Mechanically unfolding small topologically simple protein L*. Biophys. J. 89, 506–519.
- CARL P., KWOK C. H., MANDERSON G., SPEICHER D. W., DISCHER D. E., 2001. *Force unfolding modulated by disulphide bonds in the Ig domains of a cell adhesion molecule*. Proc. Natl. Acad. Sci. USA 98, 1565–1570.
- CARRION-VAZQUEZ M., OBERHAUSER A. F., FISCHER T. E., MARSZALEK P. E., LI H., FERNANDEZ J. M., 2000. *Mechanical design of proteins studied by single-molecule force spectroscopy and protein engineering*. Prog. Biophys. Mol. Biol. 74 63–91.
- CARRION-VAZQUEZ M., LI H., LU H., MARSZALEK P. E., OBERHAUSER A. F., FERNANDEZ J. M., 2003. *The mechanical stability of ubiquitin is linkage dependent*. Nat. Struct. Biol. 10 738–743.
- CIEPLAK M., HOANG T. X., 2003. *Universality classes in folding times of proteins*. Biophys. J. 84 475–488.
- CIEPLAK M., HOANG T. X., ROBBINS M. O., 2004. *Thermal effects in stretching of Go-like models of titin and secondary structures*. Proteins: Struct. Funct. Bio. 56, 285–297.
- CIEPLAK M., PASTORE A., HOANG T. X., 2005. *Mechanical properties of the domains of titin in a Go-like model*. J. Chem. Phys. 122 054906.
- CIEPLAK M., FILIPEK S., JANOVJAK H., KRZYSKO K. A., 2006. *Pulling single bacteriorhodopsin out of a membrane: Comparison of simulation and experiment*. Bioch. Biophys. Acta 1758 537–544.
- FERNANDEZ J. M., LI H., 2004. *Force-clamp spectroscopy monitors the folding trajectory of a single protein*. Science 303 1674.
- JANOVJAK H., KESSLER M., OESTERHELT D., GAUB H., MULLER D. J., 2003. *Unfolding pathways of native bacteriorhodopsin depend on temperature*. EMBO J. 22, 5220–5229.
- KWIECINSKA J. I., CIEPLAK M., 2005. *Chirality and protein folding*. J. Phys.: Cond. Mat. 17, S1565–S1580.
- LAW R., CARL P., S. HARPER, DALHAIMER P., SPEICHER D. W., DISCHER D. E., 2003. *Cooperativity in force unfolding of tandem spectrin repeats*. Biophys. J. 84, 533–544.
- LEE G., ABDI K., JIANG Y., MICHAELY P., BENNETT V., MARSZALEK P., 2006. *Nanospring behavior of ankyrin repeats*. Nature 440, 246–249.
- LI H., LINKE W. A., OBERHAUSER A. F., CARRION-VAZQUEZ M., KERKVIET J. G., LU H., MARSZALEK P. E., FERNANDEZ J. M., 2002. *Reverse engineering of the giant muscle protein titin*. Nature 418, 998–1002.
- LI L. W., WETZEL S., PLUCKTHUN A., FERNANDEZ, J. M., 2006. *Stepwise unfolding of ankyrin repeats in a single protein revealed by atomic force microscopy*. Biophys. J. 90, L30–32.
- LI P. -C., MAKAROV D. E., 2004. *Simulation of the mechanical unfolding of ubiquitin: Probing different unfolding reaction coordinates by changing*

- the pulling geometry.* J. Chem. Phys. 121, 4826–4832.
- LU H., ISRALEWITZ B., KRAMMER A., VOGEL V., SCHULTEN K., 1998. *Unfolding of titin immunoglobulin domains by steered molecular dynamics simulation.* Biophys. J. 75, 662–671.
- ORENGO C. A., MICCHIE A. D., JONES S., JONES D. T., SWINDELS M. B., THORNTON J. M., 1997. *CATH – A hierarchical classification of protein domain structures.* Structure 5, 1093–1108.
- SCHLIERF M., LI H., FERNANDEZ J. M., 2004. *The unfolding kinetics of ubiquitin captured with single-molecule force-clamp techniques.* Proc. Natl. Acad. Sci. 101, 7299.
- SCHWAIGER I., KARDINAL A., SCHLEICHER M., NOEGEL A. A., RIEF M., 2004. *A mechanical unfolding intermediate in an actin-crosslinking protein.* Nat. Struct. Mol. Biol. 11, 81–85.
- SZYMCZAK P., CIEPLAK M., 2006. *Stretching of proteins in a uniform flow.* J. Chem. Phys. 125, 164903.
- TSAI J., TAYLOR R., CHOTHIA C., GERSTEIN M., 1999. *The packing density in proteins: Standard radii and volumes.* J. Mol. Biol. 290, 253–266.
- VEITSHANS T., KLIMOV D., THIRUMALAI D., 1997. *Protein folding kinetics: time scales, pathways and energy landscapes in terms of sequence-dependent properties.* Folding Des. 2, 1–22.
- WEST D. K., BROCKWELL D. J., OLMSTED P. D., RADFORD S. E., PACI E., 2006. *Mechanical resistance of proteins explained using simple molecular models.* Bioph. J. 90, 287–297.
- YANG G., CECCONI C., BAASE W. A., VETTER I. R., BREYER W. A., HAACK J. A., MATTHEWS B. W., DAHLQUIST F. W., BUSTAMANTE C., 2000. *Solid-state synthesis and mechanical unfolding of polymers of T4 lysozyme.* Proc. Nat. Acad. Sci. USA 97, 139–144.