

Konformacje białek wielodomenowych i częściowo nieustrukturyzowanych – symulacje i eksperymenty

Bartosz Różycki✉

Instytut Fizyki Polskiej Akademii Nauk, Laboratorium Fizyki Biologicznej, Warszawa

✉Instytut Fizyki Polskiej Akademii Nauk, Laboratorium Fizyki Biologicznej, Aleja Lotników 32/46, 02-668 Warszawa; e-mail: rozycki@ifpan.edu.pl

Artykuł otrzymano 8 lutego 2017 r.
Artykuł zaakceptowano 8 marca 2017 r.

Słowa kluczowe: biofizyka molekularna, biologia strukturalna, białka pozbawione struktury trzeciorzędowej, białka wielodomenowe, symulacje dynamiki molekularnej

Podziękowania: Badania prowadzone przez autora niniejszej pracy są finansowane ze środków przyznanych przez Narodowe Centrum Nauki na realizację w latach 2017-2020 projektu 2016/21/B/NZ1/00006.

STRESZCZENIE

Biologia strukturalna zajmuje się badaniem przestrzennych struktur makrocząsteczek takich jak białka, DNA, RNA oraz ich kompleksów w celu wyjaśnienia podstawowych mechanizmów ich działania. Wśród makrocząsteczek stanowiących obecnie największe wyzwania dla biologii strukturalnej są białka, które zbudowane są z kilku odrębnych domen połączonych długimi, nieustrukturyzowanymi odcinkami łańcucha polipeptydowego. Białka tego rodzaju pełnią wiele rozmaitych funkcji w organizmach żywych i jednocześnie są wyjątkowo trudne do scharakteryzowania za pomocą konwencjonalnych technik biologii strukturalnej. Ta luka w metodzie poznawczej doprowadziła niedawno do rozwoju tzw. metod hybrydowych, które wykorzystują nowatorskie metody obliczeniowe do łączenia danych z różnorodnych, wzajemnie uzupełniających się doświadczeń strukturalnych. Niniejsza praca poświęcona jest implementacji i wykorzystaniu metod hybrydowych. Zawiera między innymi szczegółowy opis sposobu, w jaki reprezentatywne struktury wielodomenowych białek uzyskiwane są metodą hybrydową o nazwie EROS (ang. *Ensemble Refinement of SAXS*).

WPROWADZENIE

Organizmy żywe, począwszy od bakterii a skończywszy na człowieku, zbudowane są z wysoko zorganizowanej i specyficznej materii, w większości nie-spotykanej w świecie nieożywionym. Materia ta obejmuje białka, kwasy nukleinowe, tłuszcze i węglowodany. Wśród nich białka stanowią arcydzieła Natury – są one prawdziwymi nanoskopijnymi maszynami w prawdziwych mikroskopijnych fabrykach, jakie stanowią organelle żywych komórek, takie jak np. mitochondria, chloroplasty lub siateczka śródplazmatyczna.

W białkach, podczas ich pracy, zachodzą różnego rodzaju procesy, m.in. zmiany konformacyjne, czyli zmiany ich przestrzennej struktury na skutek przemieszczania się pewnych grup atomów względem innych. Niektóre białka służą do transportowania ważnych substancji lub stanowią kanały, przez które określone cząsteczki lub jony mogą się przedostać do wnętrza lub na zewnątrz komórki. Jeszcze inne pośredniczą w przekazywaniu sygnałów, tzn. kontroli łańcuchów reakcji biochemicznych, które pozwalają komórce reagować na zmiany w swoim środowisku.

Biologia strukturalna zajmuje się badaniem przestrzennych struktur makrocząsteczek, a w szczególności białek, w celu szczegółowego zrozumienia zasad ich działania. Jej zasadniczym celem jest poznanie podstawowych procesów życiowych na poziomie molekularnym. Biologia strukturalna zrodziła się z rentgenografii strukturalnej. Obecnie wykorzystuje się dużo szersze spektrum metod, gdyż wielu ważnych zagadnień biologii strukturalnej nie można rozwiązać wykorzystując jedynie metodę rentgenografii. Przykładami godnymi uwagi są białka pozbawione struktury trzeciorzędowej, a w szczególności duże białka wielodomenowe, w których poszczególne domeny połączone są nieustrukturyzowanymi odcinkami łańcucha polipeptydowego.

Oprócz rentgenografii strukturalnej, współczesne badania strukturalne białek wykorzystują często takie metody jak magnetyczny rezonans jądrowy (NMR, ang. *Nuclear Magnetic Resonance*), kriogeniczna mikroskopia elektronowa (cryoEM, ang. *cryo-Electron Microscopy*), rozpraszanie promieniowania rentgenowskiego pod małymi kątami (SAXS, ang. *Small Angle X-ray Scattering*), Försterowski rezonansowy przekaz energii (FRET, ang. *Förster Resonance Energy Transfer*) oraz spektroskopia elektronowego rezonansu paramagnetycznego (EPR, ang. *Electron Paramagnetic Resonance*), w tym metoda pulsowego EPR o nazwie DEER (ang. *Double Electron-Electron Resonance*). Metody te, podobnie jak rentgenografia strukturalna, wywodzą się z fizyki materii skondensowanej i zostały stopniowo zaadaptowane na potrzeby biologii strukturalnej. Ich

Tabela 1. Podsumowanie kilku z najważniejszych metod współczesnej biologii strukturalnej. Przygotowano na podstawie pracy przeglądowej [2].

Metoda	Wymagania i ograniczenia	Zalety
Rentgenografia strukturalna	<ul style="list-style-type: none"> - źródło monochromatycznego promieniowania rentgenowskiego; - kryształy białek; - problem fazy; - nie znajduje zastosowania do badania białek typu IDP (białek natywnie nieuporządkowanych); 	<ul style="list-style-type: none"> - struktury makrocząsteczek wyznaczone z rozdzielczością atomową; - brak ograniczeń ze względu na rozmiar makrocząsteczek; - szybkie i łatwe tworzenie modeli strukturalnych białek (oprogramowanie przyjazne dla użytkownika); - powszechnie dostępne wyposażenie synchrotronów;
NMR	<ul style="list-style-type: none"> - drogie spektrometry NMR; - znakowanie białek izotopami, zwykle ^{13}C i ^{15}N; - ograniczenie do białek małych i średnich rozmiarów; - przyporządkowanie rezonansów i tworzenie modeli strukturalnych jest często trudne i czasochłonne; 	<ul style="list-style-type: none"> - struktury makrocząsteczek wyznaczone z rozdzielczością atomową; - nie wymaga hodowli kryształów; - ma zastosowanie do „elastycznych makrocząsteczek” takich jak np. IDP; - może dostarczać informacji o dynamice makrocząsteczek w środowisku wodnym;
Mikroskopia elektronowa (cryoEM)	<ul style="list-style-type: none"> - bardzo drogie mikroskopy elektronowe; - ograniczenie do bardzo dużych kompleksów makromolekularnych; - najnowocześniejsze algorytmy analizy obrazu są bardzo wymagające obliczeniowo; 	<ul style="list-style-type: none"> - nie wymaga dużych ilości białka; - nie wymaga ani znakowania białek ani hodowli kryształów; - wykorzystuje symetrię przestrzenną badanych układów, takich jak np. otoczki wirusów;
SAXS	<ul style="list-style-type: none"> - źródło monochromatycznego promieniowania rentgenowskiego; - niska rozdzielczość przestrzenna; - dostarcza informacji o rozmiarach i kształtach makrocząsteczek lecz nie o ich strukturach atomowych; 	<ul style="list-style-type: none"> - nie wymaga ani znakowania białek ani hodowli kryształów; - brak ograniczeń na rozmiary badanych makromolekuł; - wysoka przepustowość pomiarów i szybka analiza danych; - wiele synchrotronów oferuje obecnie linie badawcze dedykowane pomiarom SAXS;
FRET i DEER	<ul style="list-style-type: none"> - wprowadzenie odpowiednich mutacji aby cysteiny znajdowały się w określonych miejscach na powierzchni badanego białka; - wyznakowanie znacznikami fluorescencyjnymi lub paramagnetycznymi; - wyznaczenie struktur małych białek na podstawie pomiarów FRET lub DEER jest możliwe ale bardzo złożone i czasochłonne; 	<ul style="list-style-type: none"> - bezpośrednia obserwacja elastyczności i zmian konformacyjnych białek; - brak ograniczeń ze względu na rozmiary białek; - znajdują zastosowania w badaniach strukturalnych białek błonowych; - do pomiarów FRET na pojedynczych cząsteczkach wystarczają bardzo małe ilości białka;

szczegółowe omówienie znaleźć można w znakomitych podręcznikach [1]. Natomiast ich zwięzły, fragmentaryczny opis, głównie w kontekście mocnych i słabych stron, został przedstawiony w artykule przeglądowym [2]. Dlatego metody te z osobna nie będą omawiane w niniejszej pracy. Celem niniejszej pracy jest natomiast przedstawienie tzw. metod hybrydowych, które wykorzystywane są do pozyskiwania informacji o strukturach makrocząsteczek na podstawie połączenia danych z wielu różnorodnych, wzajemnie uzupełniających się doświadczeń, takich jak wymienione wcześniej NMR, cryoEM, SAXS, FRET, EPR lub innych [3].

Metody hybrydowe rozwijane są przede wszystkim w celu charakteryzacji strukturalnej dużych, dynamicznych kompleksów białkowych, które złożone są z wielu odrębnych domen, elastycznych łączników międzys domenowych oraz natywnie nieustrukturyzowanych odcinków łańcucha polipeptydowego mogących przyjąć unikalną strukturę tylko na skutek pewnych reakcji biochemicznych lub procesów biofizycznych, takich jak przyłączenie określonych cząsteczek. Tego typu architektura molekularna spotykana jest na przykład w kompleksach białkowych zaangażowanych w dynamiczne kształtowanie i rozwój organelli wewnątrzkomórkowych. Przykładami mogą być kompleksy białkowe ESCRT [4,5] i BLOC [6].

HYBRYDOWE METODY BIOLOGII STRUKTURALNEJ

Wszystkie metody biologii strukturalnej mają ograniczony zakres stosowalności, co zostało podsumowane w tabeli 1. Dlatego w pewnych przypadkach może okazać się, że żadna z tych metod nie jest wystarczająca do wyznaczenia struktur przestrzennych danych makrocząsteczek. W takich przypadkach warto spróbować strategii, która polega na wykorzystaniu kilku wzajemnie uzupełniających się metod. Przykładowo, krystalografia strukturalna jest coraz częściej uzupełniana przez SAXS w celu określenia struktur białek wielodomenowych oraz kompleksów białkowych [7,8]. Natomiast mapy gęstości wyznaczone przy użyciu mikroskopu elektronowego (metodą cryoEM) mogą być wykorzystane jako dane „wejściowe” do podstawienia molekularnego, które jest powszechnie używane w krystalografii strukturalnej. Takie połączenie krystalografii białek z metodą cryoEM zostało wykorzystane np. do wyznaczenia struktur otoczek wirusów z rozdzielczością kilku angstromów [9,10]. Z kolei metody łączące NMR i SAXS pozwalają wyznaczać atomowe struktury białek w środowisku wodnym [11,12]. Takie połączenie wydaje się szczególnie skuteczne, ponieważ NMR i SAXS dostarczają komplementarnych informacji strukturalnych. Mianowicie, pomiary NMR nakładają „lokalne” ograniczenia na odległości między atomami lub na kąty dwuścienne między płaszczyznami wiązań chemicz-

nych w badanym białku, natomiast dane z doświadczeń SAXS dostarczają „globalnych” informacji o rozmiarze i kształcie tego białka. Warto także wspomnieć, że połączenie metod SAXS, NMR i krystalografii strukturalnej zostało użyte do określenia zespołu struktur kompleksu białkowego p38 α :HePTP zawierającego elastyczne pętle i łączniki międzydomenowe [13]. W tym przypadku krystalografia została użyta do wyznaczenia struktur poszczególnych domen białkowych wchodzących w skład badanego kompleksu; zmiany przesunięcia chemicznego wyznaczone w doświadczeniach NMR dostarczyły informacji o aminokwasach tworzących kontakty międzydomenowe; natomiast dane rozproszeniowe z doświadczeń SAXS umożliwiły umiejscowienie składowych domen białkowych względem siebie.

Do wyznaczania modeli strukturalnych dużych, dynamicznych, wielodomenowych kompleksów białkowych szczególnie dobrze nadają się metody hybrydowe, takie jak EROS (ang. *Ensemble Refinement of SAXS*), które łączą rentgenografię strukturalną lub NMR (które to umożliwiają wyznaczenie struktur poszczególnych domen białkowych) z SAXS (która dostarcza informacji na temat rozmiaru i kształtu całego badanego kompleksu białkowego) oraz z DEER lub FRET (które natomiast nakładają lokalne ograniczenia na odległości pomiędzy wybranymi parami znaczników na powierzchni białek). Za pomocą takich metod hybrydowych jak EROS, wyniki doświadczeń rentgenografii strukturalnej, NMR, SAXS, DEER i FRET mogą zostać zintegrowane i łącznie wykorzystane do szczegółowej reprezentacji struktur i ruchów takich kompleksów białkowych jak ESCRT [4,5,14] lub kinaz białkowych tworzących dynamiczne kompleksy z fosfatazami [13].

W kolejnych akapitach opisany zostanie sposób, w jaki struktury hybrydowe są uzyskiwane przy użyciu metody EROS [14]. Modelowanie typu EROS przebiega w dwóch etapach. W pierwszym etapie wykonywane są symulacje komputerowe danego kompleksu białkowego, w wyniku których wygenerowany zostaje szereg modeli strukturalnych. W ten sposób otrzymywany jest zespół konformacyjny przeznaczony do „dalszej obróbki”. W drugim etapie ów zespół konformacyjny zostaje poprawiony w taki sposób, aby osiągnąć zgodność ze wszystkimi dostępnymi danymi doświadczalnymi. Te dwa kolejne etapy zostaną szczegółowo opisane poniżej.

EROS

Symulacje dynamiki molekularnej dużych kompleksów białkowych w skali atomowej są bardzo czasochłonne, szczególnie jeśli symulowane mają być znaczne zmiany konformacyjne. Dlatego na potrzeby metody EROS wykorzystywane są tzw. symulacje gruboziarniste, w których całe aminokwasy reprezentowane są przez pojedyncze „kulki” lub „ziarna” [15]. Co więcej, domeny białkowe o znanych strukturach (tzn. strukturach wyznaczonych przy użyciu metod krystalografii strukturalnej lub NMR) reprezentowane są przez bryły sztywne podlegające jedynie przesunięciom i obrotom w czasie symulacji. Natomiast nieustrukturyzowane pętle i elastyczne łączniki międzydomenowe reprezentowane są przez łańcuchy „ziaren” odpowiednich

aminokwasów. Łańcuchy te są lokalnie rozciągane, zgnane i skręcane w czasie symulacji, tak aby symulowane białka mogły przybierać różnorodne konformacje. Jednak wszystkie ruchy białek wykonywane są w czasie symulacji zgodnie z zasadami mechaniki i fizyki statystycznej. Oddziaływania fizyczne między domenami białkowymi opisane są na poziomie pojedynczych „ziaren” przez energię potencjalną, która uwzględnia statystyki występowania kontaktów między aminokwasami oraz potencjał elektrostatyczny w przybliżeniu Debye’a-Hückela. Zmiany długości Debye’a, która opisuje zasięg oddziaływań elektrostatycznych w roztworach wodnych, umożliwiają symulacje białek w roztworach o różnych siłach jonowych. Jak zostało wykazane w serii symulacji, w ten sposób dobrana energia potencjalna poprawnie przewiduje struktury i stałe równowagi wielu kompleksów białkowych [15]. Wynik ten wskazuje na to, że symulacje gruboziarniste typu EROS prawidłowo próbują rzeczywiste konformacyjne kompleksów białkowych.

Modele strukturalne wygenerowane w procesie symulacji są następnie segregowane na podstawie wzajemnego podobieństwa i grupowane w zbiory. Dany zbiór składa się z modeli strukturalnych (zwykle od kilku do kilkudziesięciu modeli), które są wzajemnie podobne. W celu podziału puli modeli na zbiory wykorzystuje się standardowe algorytmy grupowania, np. algorytm centroidów (ang. *k-means*) lub algorytm typu QT (ang. *Quality Threshold*), natomiast jako miary podobieństwa modeli używane są zwykle DRMS (ang. *Distance Root Mean Square*). Dla każdego ze zbiorów osobno oblicza się następnie wielkości, które zostały wyznaczone w doświadczeniach na badanych białkach, czyli np. krzywą rozproszeniową lub histogramy wydajności przekazu energii między parami znaczników. Obliczeń tych dokonuje się oczywiście na podstawie poszczególnych struktur przypisanych do zbiorów. Następnie zbiorom przypisuje się wagi statystyczne i przyjmuje się, że średnia ważona po zbiorach reprezentuje średnią po zespole statystycznym. Dlatego wielkości uśrednione po zespole statystycznym, które mają zostać porównane bezpośrednio z wielkościami wyznaczonymi w doświadczeniach (SAXS, FRET, DEER, etc.), zależą od wag statystycznych przypisanych zbiorom modeli strukturalnych. Wybierając zatem odpowiednie wagi statystyczne można zwykle osiągnąć zgodność wyników doświadczalnych z wynikami teoretycznymi (tzn. zgodność wielkości zmierzonych w doświadczeniach z wielkościami obliczonymi na podstawie symulowanych modeli strukturalnych po uśrednieniu z odpowiednimi wagami). Aby uniknąć problemu niejednoznaczności wyboru właściwych wag statystycznych, dopasowanie wielkości teoretycznych do doświadczalnych opiera się na metodzie maksymalnej entropii. Metoda ta sprowadza się do numerycznej minimalizacji różnicy dwóch funkcji. Pierwsza z tych funkcji opisuje stopień zgodności przewidywań teoretycznych z wynikami doświadczeń i zależy w skomplikowany sposób od wag statystycznych zbiorów symulowanych struktur. Druga z tych funkcji ma postać analogiczną do entropii Shannona i rośnie wraz ze zmianami wag statystycznych zbiorów. W wyniku minimalizacji różnicy tych dwóch funkcji otrzymuje się zespół konformacyjny (tzn. zestaw zbiorów modeli strukturalnych wraz z przyporządkowanymi wagami statystycznymi), który jest zgodny z wynikami doświadczeń, a jednocześnie jest możliwie podobny do zespołu konforma-

cyjnego otrzymanego bezpośrednio z symulacji (tzn. puli modeli strukturalnych bez przypisanych wag statystycznych).

Prostszy sposób konstrukcji zespołu konformacyjnego, który jest często używany w praktyce, polega na wyborze najmniejszego możliwego zestawu zbiorów struktur, który daje zgodność teorii z doświadczeniem [4,5,13]. Zaletą takiego podejścia jest niewątpliwie to, że otrzymany zespół konformacyjny jest zwykle reprezentowany tylko przez kilka struktur, które mogą być łatwo zwizualizowane i zinterpretowane. Jednak odrzucając znaczną część modeli strukturalnych otrzymanych z symulacji molekularnych, metoda „minimalnego zespołu” nie wykorzystuje w pełni prognostycznych możliwości symulacji, które oparte są na zasadach fizyki i chemii.

Współczesna biologia strukturalna korzysta obficie z metod modelowania molekularnego. Wyznaczanie struktur białek na podstawie wyników doświadczeń krystalograficznych lub NMR wymaga modelowania wszystkich atomów tworzących daną cząsteczkę białka. Natomiast do interpretacji wyników doświadczeń SAXS wystarcza modelowanie typu gruboziarnistego, przybliżające grupy atomów przez pojedyncze „ziarna”, gdyż rozdzielczość przestrzenna metody SAXS jest mniejsza od rozmiarów aminokwasów. Z kolei do analizy i interpretacji wyników doświadczeń FRET i DEER konieczne jest modelowanie znaczników fluorescencyjnych i paramagnetycznych z rozdzielczością atomową. Metoda EROS wykorzystuje symulacje gruboziarniste, gdyż została wprowadzona z myślą o badaniach dużych białek wielodomenowych, które mogą wykazywać znaczne zmiany konformacyjne. Natomiast atomowe struktury znaczników fluorescencyjnych i paramagnetycznych są wykorzystywane w metodzie EROS na etapie poprawiania zespołu konformacyjnego, gdy wyniki symulacji porównywane są z wynikami doświadczeń FRET i DEER [4,5].

Zespół konformacyjny może zostać wyznaczony na podstawie dopasowania wysymulowanych modeli strukturalnych do wyników pomiarów doświadczalnych (takich jak np. krzywa rozproszeniowa) lub do „przetworzonych” danych, które otrzymane zostały na podstawie danych doświadczalnych (takich jak np. rozkład odległości pomiędzy parami atomów w białku, który jest transformatą Fouriera krzywej rozproszeniowej). Aby jednak uniknąć potencjalnych błędów wynikających z wprowadzenia różnego rodzaju artefaktów przy przetwarzaniu danych, takich jak np. regularyzacja danych, zalecane jest dopasowanie zespołu konformacyjnego bezpośrednio do danych eksperymentalnych. Bardzo ważna jest także walidacja przeprowadzonej procedury obliczeniowej, która polega na sprawdzeniu, że zaproponowany zespół konformacyjny jest zgodny z wynikami doświadczeń, które nie zostały wykorzystane do konstrukcji tego zespołu.

PODSUMOWANIE

Wśród białek najtrudniejszych do scharakteryzowania strukturalnego są obecnie te, które zbudowane są z wielu domen połączonych długimi, nieustrukturyzowanymi, elastycznymi łącznikami. Mimo tego, że białka tego rodzaju

często występują w komórkach biologicznych i pełnią w nich wiele ważnych funkcji, to stanowią one białą plamę w dziedzinie biologii strukturalnej. Ich znaczny rozmiar oraz elastyczność strukturalna są powodem tego, że nie ma obecnie jednej, uniwersalnej metody badawczej, która mogłaby dostarczyć informacji o strukturach tego rodzaju białek. Jednak jeśli wyniki pomiarów przeprowadzonych kilkoma różnymi, uzupełniającymi się metodami, tj. krystalografia, NMR, SAXS, EPR, FRET, zostaną połączone przy pomocy zaawansowanych narzędzi obliczeniowych, to wyznaczenie reprezentatywnych konformacji tego typu białek staje się możliwe. Podejście takie zostało zastosowane np. do wyznaczenia zespołów konformacyjnych kompleksów białkowych ESCRT-I [4] i ESCRT-II [5]. W przyszłości analogiczne metody mogą być zastosowane do innych kompleksów białkowych zaangażowanych w transport wewnątrzkomórkowy, takich jak np. BLOC lub ATG.

Innym ważnym kierunkiem rozwoju biologii strukturalnej w nadchodzących latach może okazać się wprowadzenie nowych metod do badania struktur białek błonowych. Znaczące wyniki uzyskano niedawno przy użyciu techniki smFRET w połączeniu z symulacjami dynamiki molekularnej [16]. Ponadto rozpraszanie neutronów pod małymi kątami (SANS, ang. *Small Angle Neutron Scattering*) może stać się ważnym narzędziem do badania białek w środowisku błon lipidowych, podobnie jak SAXS jest obecnie ważną metodą o niskiej rozdzielczości do badania białek w środowisku wodnym. Ciekawą własnością SANS jest to, że w odpowiednio dobranych warunkach lipidy rozpraszają neutrony inaczej niż cząsteczki białek [17]. Ta własność może umożliwić np. porównanie struktur białka peryferyjnego przed i po związaniu z błoną lipidową lub nawet bezpośrednią obserwację zmian strukturalnych białek transbłonowych (np. transbłonowy receptor związany z ligandem i bez niego).

PIŚMIENNICTWO

1. Liljas A, Liljas L, Piskur J, Lindblom G, Nissen P, Kjeldgaard M. Text-book of structural biology. World Scientific Publishing, Singapore, 2009, pp. 580
2. Różycki B, Boura E (2014) Large, dynamic, multi-protein complexes: a challenge for structural biology. *J Phys Condens Matter* 26: 463103
3. Sali A, Berman HM, Schwede T, Trewhella J, Kleywegt G, Burley SK, Markley J, Nakamura H, Adams P, Bonvin AM, Chiu W, Peraro MD, Di Maio F, Ferrin TE, Grünewald K, Gutmanas A, Henderson R, Hummer G, Iwasaki K, Johnson G, Lawson CL, Meiler J, Marti-Renom MA, Montelione GT, Nilges M, Nussinov R, Patwardhan A, Rappsilber J, Read RJ, Saibil H, Schröder GF, Schwieters CD, Seidel CA, Svergun D, Topf M, Ulrich EL, Velankar S, Westbrook JD (2015) Outcome of the First wwPDB Hybrid/Integrative Methods Task Force Workshop. *Structure* 23: 1156-1167
4. Boura E, Różycki B, Herrick DZ, Chung HS, Vecer J, Eaton WA, Cafiso DS, Hummer G, Hurley JH (2011) Solution structure of the ESCRT-I complex by small-angle X-ray scattering, EPR, and FRET spectroscopy. *Proc Natl Acad Sci USA* 108: 9437-9442
5. Boura E, Różycki B, Chung HS, Herrick DZ, Canagarajah B, Cafiso DS, Eaton WA, Hummer G, Hurley JH (2012) Solution structure of the ESCRT-I and -II supercomplex: implications for membrane budding and scission. *Structure* 20: 874-886
6. Lee HH, Nemecek D, Schindler C, Smith WJ, Ghirlando R, Steven AC, Bonifacino JS, Hurley JH (2012) Assembly and architecture of biogenesis of lysosome-related organelles complex-1 (BLOC-1). *J Biol Chem* 287: 5882-5890

7. Ragusa MJ, Stanley RE, Hurley JH (2012) Architecture of the Atg17 complex as a scaffold for autophagosome biogenesis. *Cell* 151: 1501-1512
8. Rambo RP, Tainer JA (2013) Super-resolution in solution X-ray scattering and its applications to structural systems biology. *Annu Rev Biophys* 42: 415-441
9. Grigorieff N, Harrison SC (2011) Near-atomic resolution reconstructions of icosahedral viruses from electron cryo-microscopy. *Curr Opin Struct Biol* 21: 265-273
10. Namecek D, Plevka P, Boura E (2013) Using cryoEM reconstruction and phase extension to determine crystal structure of bacteriophage varphi6 major capsid protein. *Protein J* 32: 635-640
11. Evrard G, Mareuil F, Bontems F, Sizun C, Perez J (2011) DADIMODO: a program for refining the structure of multidomain proteins and complexes against small-angle scattering data and NMR-derived restraints. *J Appl Cryst* 44: 1264-1271
12. Grishaev A, Tugarinov V, Kay LE, Trewella J, Bax A (2008) Refined solution structure of the 82-kDa enzyme malate synthase G from joint NMR and synchrotron SAXS restraints. *J Biomol NMR* 40: 95-106
13. Francis DM, Różycki B, Koveal D, Hummer G, Page R, Peti W (2011) Structural basis of p38-alpha regulation by hematopoietic tyrosine phosphatase. *Nat Chem Biol* 7: 916-924
14. Różycki B, Kim YC, Hummer G (2011) SAXS ensemble refinement of ESCRT-III CHMP3 conformational transitions. *Structure* 19: 109-116
15. Kim YC, Hummer G (2008) Coarse-grained models for simulations of multiprotein complexes: application to ubiquitin binding. *J Mol Biol* 375: 1416-1433
16. Wang Y, Liu Y, Deberg HA, Nomura T, Hoffman MT, Rohde PR, Schulten K, Martinac B, Selvin PR (2014) Single molecule FRET reveals pore size and opening mechanism of a mechano-sensitive ion channel. *Elife* 3: e01834
17. Clifton LA, Neylon C, Lakey JH (2013) Examining protein-lipid complexes using neutron scattering. *Methods Mol Biol* 974: 119-150

Conformations of multi-domain and partially disordered proteins – simulations and experiments

Bartosz Różycki✉

Institute of Physics of the Polish Academy of Sciences, Laboratory of Biological Physics, 32/46 Aleja Lotników, 02-668 Warsaw, Poland

✉e-mail: rozycki@ifpan.edu.pl

Key words: molecular biophysics, structural biology, intrinsically disordered proteins, multi-domain proteins, molecular dynamics simulations

ABSTRACT

Structural biology unravels three-dimensional structures of macromolecules such as proteins, DNA, RNA, and their complexes in the attempt to explain the basic mechanisms of their functions. Among the proteins that are most difficult to characterize structurally are those which have several large domains connected by long, unstructured polypeptide segments. Such proteins perform diverse functions in living organisms and, at the same time, they are very difficult to characterize using conventional methods of structural biology. This gap in the market has recently led to the development of hybrid methods that use state-of-the-art computational tools to combine complementary data from various experiments. This review article is focused on the implementation and usage of such hybrid methods. It includes a detailed description of how representative structures of multi-domain proteins are obtained using the so-called EROS (Ensemble Refinement of SAXS) hybrid method.