

Uniwersytet im. Adama Mickiewicza w Poznaniu

Wydział Fizyki



mgr Michał Taube

**Badania strukturalne wybranych białek
uczestniczących w obronie przeciwko patogenom u
*Hordeum vulgare***

Rozprawa doktorska wykonana w Zakładzie Fizyki Makromolekularnej
pod kierunkiem prof. UAM dr hab. Macieja Kozaka

Poznań 2015

Streszczenie

Rośliny narażone są na atak szerokiego spektrum patogenów, takich jak bakterie, wirusy czy grzyby. W procesie ewolucji rośliny wykształciły skomplikowane mechanizmy obronne mające na celu zatrzymaniu infekcji wywołanej przez patogeny. Jednym z elementów tego systemu odporności na poziomie molekularnym są cytoplazmatyczne receptory odporność R (ang. resistance proteins), które rozpoznają bezpośrednio lub pośrednio cząsteczki pochodzenia patogennego i uruchamiają szlak sygnalizacyjny prowadzący do uruchomienia systemu obronnego rośliny. Do pełnienia swojej funkcji białka R potrzebują kompleksu białek HSP90-SGT1-RAR1, który jak się uważa może stabilizując ich strukturę i poziom w komórce.

W niniejszej rozprawie przedstawione zostały wyniki badań dotyczących struktury oraz dynamiki w roztworze poszczególnych białek roślinnego kompleksu HSP90-SGT1-RAR1 oraz kompleksu białek HSP90-SGT1 z ADP.

Białko SGT1 składa się z trzech domen: TPR, CS oraz SGS. Struktura domeny CS białka SGT1 jest znana, a także struktura domeny TPR z innych białek została dobrze scharakteryzowana. Jednak w przypadku białka SGT1 struktura pełnej długości białka nie jest znana. Co więcej, w roztworze roślinne białko SGT1 tworzy dimery, a proces dimeryzacji jest zależny od siły jonowej. Za pomocą małokątowego rozpraszania promieniowania rentgenowskiego i spektroskopii dichroizmu kołowego zbadano niskorozdzielczą strukturę pełnej długości białka SGT1 z jęczmienia (*Hordeum vulgare*) w roztworze. Mechanizm dimeryzacji białka SGT1 nie jest do końca poznany. Dimeryzacji ulega również białko SGT1 z drożdży ale nie ludzkie białko SGT1. W niniejszej pracy przedstawiona została także analiza procesu dimeryzacji domeny TPR białka SGT1 z jęczmienia, *Arabidopsis thaliana*, drożdży oraz w celach porównawczych domeny TPR z ludzkiego białka SGT1.

Podobnie jak białko SGT1, białko RAR1 ma modułową budowę i składa się z dwóch domen CHORD wiążących dwa jony cynku każda. Struktura domeny CHORD jest znana lecz brak jest informacji na temat struktury pełnej długości białka RAR1. Stosując to samo podejście eksperymentalne jak w przypadku białka SGT1 scharakteryzowana została struktura pełnej długości białka RAR1 w roztworze.

Charakterystykę poszczególnych białek kompleksu dopełnia analiza struktury i dynamiki w roztworze białka HSP90. Dynamika konformacyjna białka HSP90 zależy od wielu czynników takich jak pH czy związanie nukleotydów. W pracy scharakteryzowano także wpływ siły jonowej i temperatury na strukturę roślinnego białka HSP90 w roztworze.

W drugiej części pracy skupiono się na analizie niskorozdzielczej struktury oraz konformacji kompleksu białek HSP90 i SGT1 z ADP. Przeanalizowana została także stechiometria i niskorozdzielcza struktura w roztworze kompleksu pełnej długości białka HSP90 z pszenicy z białkiem SGT1 bez domeny SGS (SGT1 Δ SGS) z jęczmienia. Badania kompleksu białka HSP90 bez domeny C-terminalnej (HSP90 Δ C), która odpowiedzialna jest za dimeryzację, pozwoliły odpowiedzieć na pytanie czy dimeryzacja białka HSP90 wpływa na stechiometrię i konformację kompleksu z białkiem SGT1 Δ SGS. Jako dopełnienie tych rozważań scharakteryzowana została niskorozdzielcza struktura kompleksu białek HSP90 Δ C z domeną CS białka SGT1.

Abstract

Plants are constantly exposed to contacts with a variety of pathogens like bacteria, viruses and fungi. During the evolution plants developed complex mechanisms that provide resistance to many pathogenic agents. One of the most important elements of this system on the molecular level are resistance receptors (R proteins) that recognize pathogenic molecules directly or indirectly and trigger the signaling pathways that lead to the activation of immune response. For their function R proteins require the protein complex that consists of three major proteins: HSP90, SGT1 and RAR1 proteins. It is assumed that this complex stabilizes level of R proteins and their structure within the plant cell.

This PhD thesis presents results of a study concerning the structure and dynamic in solution of the HSP90-SGT1 complex with ADP and individual proteins of HSP90-SGT1-RAR1 complex.

SGT1 protein consists of three domains: TPR, CS and SGS ones. The structure of CS domain is known and TPR domain belongs to the structurally well characterized domain family. Unfortunately, the structure of the full length SGT1 protein has been poorly explored. What is more plant SGT1 protein forms dimers in solution and this process depends on ionic strength. Using small angle X-ray scattering and circular dichroism spectroscopy, the low resolution structure of the full length barley SGT1 protein was studied. The exact mechanism of the SGT1 protein dimerization is not fully understood. Dimerization of the SGT1 proteins was also discovered in fungi, but the human homolog does not undergo dimerization. In this study the dimerization process of the TPR domain from barley, *Arabidopsis thaliana*, yeasts and for comparison from the human homolog is analyzed. RAR1 protein is similar to SGT1 protein and has a modular structure consisting of two CHORD domains, each of them binding two zinc ions. The structure of the CHORD domain is known, but there is lack of information about the structure and conformation of the full length protein. Using the same approach as for SGT1 protein, the structure in solution of the RAR1 protein from barley is characterized.

The structure and dynamics of the wheat HSP90 protein is also presented. Conformational dynamics of HSP90 protein depends on many factors like pH and binding of nucleotides. The effect of the ionic strength and temperature on the HSP90 protein conformation is studied in details.

The second part of PhD thesis is devoted to a study of the low resolution structure and conformation of the HSP90-SGT1 complex with ADP. The stoichiometry and the structure in solution of the complex between the full length HSP90 protein from wheat and the barley SGT1 protein, with SGS domain deleted (SGT1 Δ SGS), is analyzed. Study of HSP90 protein with deleted C terminal dimerization domain (HSP90 Δ C) allows us to answer the question if HSP90 dimerization could affect the stoichiometry and conformation of the complex with SGT1 Δ SGS protein. To complement the discussion, the low resolution structure of the complex between HSP90 Δ C and SGT1 CS domain is also presented.

Podziękowania

Chciałbym serdecznie podziękować mojemu promotorowi Panu prof. Maciejowi Kozakowi za okazaną mi pomoc i wsparcie podczas studiów doktoranckich i przygotowywania pracy doktorskiej. Dziękuję również za obdarzenie mnie zaufaniem podczas pierwszych lat współpracy i umożliwienie mi pracy pod jego kierunkiem, za dzielenie się ze mną swoją wiedzą i doświadczeniem oraz za zarażenie mnie pasją do badań strukturalnych. Bez jego wsparcia niniejsza praca nie mogłaby powstać. Chciałbym również podziękować za stworzenie przemiłej i stymulującej atmosfery w laboratorium i poza nim, podczas licznych pomiarów synchrotronowych i konferencji.

Podziękowania kieruje także do wszystkich kolegów i koleżanek z Zakładu Fizyki Makromolekularnej, w szczególności Zuzanny i Magdy, za współpracę i pomoc podczas realizacji mojej pracy doktorskiej.

Szczególnie serdecznie chciałbym podziękować Pani prof. Zofii Szweykowskiej-Kulinskiej i Panu prof. Arturowi Jarmołowskiemu za wspieranie współpracy, inspirujące rozmowy i wprowadzenie w niezwykle interesujący świat biologii molekularnej, a przede wszystkim za umożliwienie mi prowadzenia badań w Zakładzie Ekspresji Genów. Praca w tym zespole pomogła mi zdobyć ogromną wiedzę i doświadczenie, bez którego powstanie tej pracy byłoby na pewno niemożliwe.

Chciałbym podziękować także wszystkim kolegom i koleżankom z Zakładu Ekspresji Genów, za wspieranie atmosfery, pasjonujące dyskusje i okazaną pomoc oraz spotkania z Wydziałem Matematyki. Szczególnie chciałbym podziękować Kasi, Agacie, Dorocie i Pawłowi za wiele interesujących rozmów i dyskusji nie tylko na tematy naukowe. Kasi i Agacie dziękuję również za motywację do wcześniejszego wstawania do pracy w laboratorium i poranną kawę.

Dziękuję również wszystkim moim znajomym, którzy wspierali mnie podczas trwania doktoratu. Szczególnie chciałbym podziękować Gosi, Jarkowi, Konradowi i Piotrkowi za wasze nieocenione wsparcie i wieloletnią przyjaźń.

Dziękuję również Agacie, Kasi, Kosmie i Sylwii za pomoc w korekcie pracy pod względem edytorskim.

Podziękować pragnę również moim Rodzicom za ich bezinteresowną miłość, wsparcie i pomoc. Już od najmłodszych lat zaszczylili we mnie chęć poznawania świata i szacunek dla zdobywania wiedzy co zaowocowało zajęciem się pracą naukową. Oddzielnie chciałbym podziękować mojemu bratu Andrzejowi, który był ze mną cały czas podczas przygotowywania pracy doktorskiej, za to że był dla mnie wsparciem, wzorem i inspiracją do ciężkiej pracy. Za to, że zawsze miał dla mnie czas i przeżywał ze mną zarówno chwile radości jak i zwątpienia. Dziękuję.

Praca częściowo zrealizowana w ramach grantu Preludium Narodowego Centrum Nauki „Badania struktury kompleksów Hsp90:Sgt1:Rar1 oraz dimeru domeny TPR białka Sgt1 przy pomocy małokątowego rozpraszania promieniowania synchrotronowego oraz mikroskopii TEM” (2012/05/N/ST3/03087), której kierownikiem jest autor niniejszej dysertacji.



Prace dedykuję moim Rodzicom i bratu Andrzejowi.



Spis treści

Streszczenie	iii
Abstract	iv
Podziękowania	v
1 Cel pracy	1
2 Wstęp	3
2.1 Oddziaływanie roślina-patogen	3
2.1.1 Odpowiedź uruchamiana przez wzorce molekularne związane z patogenami . . .	4
2.1.2 Odpowiedź uruchamiana przez efektory	5
2.1.3 Receptory odporności R oraz ich udział w odpowiedzi uruchamianej przez efek- tory	7
2.1.4 Struktura domenowa białek R	9
2.2 Rola białek HSP90, SGT1, RAR1 w odpowiedzi obronnej skierowanej przeciwko pa- togenom	10
2.2.1 Budowa i funkcja białka SGT1	12
2.2.2 Budowa i funkcja białka RAR1	15
2.3 Funkcja, struktura i dynamika białek z rodziny HSP90	16
2.3.1 Funkcja białek HSP90	16
2.3.2 Struktura białka HSP90	18
2.3.3 Dynamika konformacyjna białka HSP90	20
2.4 Struktura i funkcja domen z rodziny TPR	21
2.4.1 Struktura domeny TPR	21
2.4.2 Funkcje białek zawierających powtórzenia TPR	24
2.4.3 Oligomeryzacja domen TPR	25
2.4.4 Zastosowanie domen TPR w inżynierii białkowej	25
2.5 Wybrane metody badania struktury przestrzennej makromolekuł - podstawy teoretyczne	27
2.5.1 Oddziaływanie promieniowania z materią	28
2.5.2 Rozpraszanie promieniowania rentgenowskiego przez atom	29
2.6 Małokątowe rozpraszanie promieniowania rentgenowskiego	31
2.6.1 Odwrotna transformata Fouriera $I(s)$ oraz krzywa rozkładu odległości wewnątrz- cząsteczkowych	33
2.6.2 Asymptotyczne zachowanie krzywej $I(s)$	35
2.6.2.1 Wykres Guiniera, promień żyracji	35
2.6.2.2 Prawo Poroda – asymptotyczne zachowanie się krzywej $I(s)$ w granicy $s \rightarrow \infty$	35
2.6.3 Czynniki kształtu	36
2.6.4 Metody <i>ab-initio</i> w modelowaniu niskorozdzielczych struktur białek i kwasów nukleinowych na podstawie danych SAXS	38

2.6.5	Obliczanie krzywej rozpraszania dla makrocząsteczek w roztworze	39
2.6.5.1	Metoda harmonijek sferycznych	39
2.6.5.2	Metoda wielomianów Zernike'a	40
2.6.6	Metody <i>ab-initio</i> do modelowanie niskorozdzielczej struktury makromolekuł . .	42
2.6.6.1	Algorytm genetyczny	42
2.6.6.2	Algorytm symulowanego wyżarzania	44
2.6.6.3	Modelowanie struktury domenowej białek za pomocą algorytmu symulowanego wyżarzania	46
2.6.6.4	Modelowanie bryły sztywnej - wykorzystanie dostępnych struktur wysokorozdzielczych	48
2.6.6.5	Modelowanie homooligomerów o określonej symetrii	49
2.6.6.6	Modelowania hetero-oligomerów	50
2.6.6.7	Inne metody modelowania na podstawie danych z małokątowego rozpraszania promieniowania rentgenowskiego	51
2.6.7	Metoda SVD	56
2.7	Dichroizm kołowy	60
2.7.1	Zjawisko dichroizmu kołowego	60
2.7.2	Wykorzystanie spektroskopii dichroizm kołowego do wyznaczania struktury drugorzędowej białek	64
2.8	Źródła promieniowania rentgenowskiego	68
2.8.1	Promieniowanie synchrotronowe	68
3	Materiały i metody	73
3.1	Lista wektorów używanych do nadekspresji białek	73
3.2	Transformacja bakterii <i>E.coli</i> metodą szoku cieplnego	73
3.3	Przygotowanie konstruktów plazmidowych do nadekspresji białek w systemie prokariotycznym	73
3.3.1	Przygotowanie wektora do klonowania	73
3.3.2	Cięcie wektora za pomocą enzymów restrykcyjnych	74
3.3.3	Przygotowanie insertu DNA do klonowania do wektora plazmidowego	74
3.3.4	Ligacja insertu DNA do wektora plazmidowego	75
3.3.5	Reakcja PCR "kolonijna"	75
3.3.6	Namnażanie pozytywnych klonów	76
3.4	Nadekspresja białek na małą skalę w komórkach <i>E.coli</i>	76
3.4.1	Nadekspresja białek na małą skalę metodą szoku cieplnego i osmotycznego . . .	76
3.5	Analiza poziomu ekspresji białek za pomocą elektroforezy w żelu poliakrylamidowym .	77
3.6	Nadekspresja na dużą skalę rekombinowanych białek w bakteriach <i>E.coli</i>	78
3.7	Oczyszczanie białek z użyciem systemu do chromatografii cieczowej Äkta Explorer 10	78
3.7.1	Oczyszczanie za pomocą chromatografii powinowactwa	78
3.7.2	Oczyszczanie białek za pomocą sączenia molekularnego	79
3.7.3	Oczyszczanie białek za pomocą chromatografii jonowymiennej	80
3.7.4	Oczyszczanie białka SGT1 z komórek <i>E.coli</i>	80

3.7.5	Skład buforów używanych podczas oczyszczania białek	81
3.7.6	Oczyszczanie wariantów białek RAR1, SGT1 i HSP90	81
3.8	Pomiary dichroizmu kołowego białka SGT1 oraz konstruktów zawierających domeny TPR-CS, CS-SGS, CS	82
3.9	Pomiary małokątowego rozpraszania promieniowania synchrotronowego	82
3.9.1	Pomiary SAXS białka HvSGT1	82
3.9.2	Pomiary SAXS białka HvRAR1	82
3.9.3	Pomiary SAXS białek HvTPR, AtTPR, ScTPR, HsTPR	83
3.9.4	Pomiary SAXS białka TaHSP90	83
3.9.5	Pomiary SAXS kompleksów białek TaHSP90-HvSGT1 Δ SGS	83
3.9.6	Pomiary SAXS kompleksów białek HvHSP90 Δ C- HvSGT1 Δ SGS i HvHSP90 Δ - HvSGT1-CS	83
3.9.7	Lista programów używanych w analizie danych SAXS	84
4	Wyniki	85
4.1	Otrzymywanie pełnej długości białka SGT1	85
4.2	Pomiary spektroskopii dichroizmu kołowego pełnej długości białka SGT1 oraz kon- struktów zawierających domeny: TPR-CS, CS-SGS oraz CS	86
4.2.1	Oczyszczanie białek do pomiarów dichroizmu kołowego	86
4.2.2	Pomiary dichroizmu kołowego	87
4.3	Badania małokątowego rozpraszania promieniowania rentgenowskiego białka SGT1 z jęczmienia	90
4.3.1	Analiza SVD danych z małokątowego rozpraszania promieniowania synchro- tronowego	90
4.3.2	Analiza krzywych rozpraszania dla monomeru i dimeru białka SGT1 za pomocą procedury MCR-ALS	92
4.3.3	Modelowanie <i>ab-initio</i> monomeru białka SGT1	93
4.4	Badanie struktury dimeru białka SGT1 w roztworze za pomocą małokątowego rozpra- szania promieniowania synchrotronowego	97
4.4.1	Badanie dimeryzacji domeny TPR białka SGT1 z jęczmienia	98
4.4.1.1	Oczyszczanie domeny TPR białka HvSGT1	98
4.4.1.2	Badania małokątowego rozpraszania promieniowania rentgenowskiego białka HvTPR	99
4.4.2	Modelowanie dimeru białka SGT1 pełnej długości	103
4.5	Badania struktury i konformacji w roztworze białka RAR1 z jęczmienia	106
4.5.1	Oczyszczanie białka RAR1 z bakterii <i>E.coli</i>	106
4.5.2	Badania małokątowego rozpraszania promieniowania synchrotronowego białka RAR1	106
4.5.3	Modelowanie <i>ab initio</i> struktury białka RAR1 w roztworze	107
4.6	Badanie dimeryzacji domeny TPR białka SGT1a z <i>Arabidopsis thaliana</i>	110
4.6.1	Oczyszczanie białka AtTPR z bakterii <i>E.coli</i>	110

4.6.2	Badania małokątowego rozpraszania promieniowania synchrotronowego białka AtTPR	111
4.6.3	Modelowanie niskorozdzielczej struktury monomeru i dimeru domeny TPR białka AtSGT1a	114
4.7	Badania struktury domeny TPR białka SGT1 z drożdży piekarskich w roztworze	117
4.7.1	Oczyszczanie białka ScTPR z bakterii <i>E.coli</i>	117
4.7.2	Badania małokątowego rozpraszania promieniowania rentgenowskiego białka ScTPR	117
4.7.3	Modelowanie struktury monomeru białka ScTPR	121
4.7.4	Modelowanie struktury dimeru białka ScTPR	122
4.7.5	Modelowanie struktury tetrameru domeny TPR białka ScSGT1	124
4.8	Badanie struktury domeny TPR ludzkiego białka SGT1	126
4.8.1	Oczyszczanie białka HsTPR z bakterii <i>E.coli</i>	126
4.8.2	Badania małokątowego rozpraszania promieniowania rentgenowskiego białka HsTPR	126
4.9	Wpływ siły jonowej i temperatury na konformację białka HSP90 w roztworze	129
4.9.1	Oczyszczanie białka TaHSP90 z bakterii <i>E.coli</i>	129
4.9.2	Badanie wpływu temperatury na dynamikę konformacyjną białka HSP90 z pszenicy w roztworze	130
4.9.3	Modelowanie <i>ab-initio</i> struktury w roztworze białka HSP90 z pszenicy	133
4.9.4	Wpływ siły jonowej na dynamikę konformacyjną białko HSP90 z pszenicy	135
4.10	Badanie kompleksu białek HSP90 i SGT1 z ADP	139
4.10.1	Struktura kompleksu TaHSP90-HvSGT1 Δ SGS z ADP	139
4.11	Modelowanie <i>ab-initio</i> struktury białek TaHSP90, HvSGT1 Δ SGS oraz kompleksu TaHSP90-HvSGT1 Δ SGS	142
4.11.1	Modelowanie białka HvSGT1 Δ SGS na podstawie danych SAXS	142
4.11.2	Modelowanie struktury białka TaHSP90 oraz kompleksu białek TaHSP90 - HvSGT1 Δ SGS	143
4.12	Badanie struktury kompleksu białek HvHSP90 Δ C i HvSGT1 Δ SGS	145
4.12.1	Oczyszczanie białka HvHSP90 Δ C z bakterii <i>E.coli</i>	145
4.12.2	Badania SAXS kompleksu białek HvHSP90 Δ C i HvSGT1 Δ SGS	145
4.12.3	Modelowanie struktury białka HvHSP90 Δ C oraz kompleksu białek HvHSP90 Δ C -HvSGT1 Δ SGS z ADP	148
4.13	Badania struktury kompleksu białek HvHSP90 Δ C i HvSGT1-CS	151
4.13.1	Oczyszczanie białka HvSGT1-CS do pomiarów SAXS	151
4.13.2	Badania SAXS kompleksu białek HvHSP90 Δ C i HvSGT1-CS	151
4.13.3	Modelowanie struktury białka HvSGT1-CS	154
4.13.4	Modelowanie struktury kompleksu białek HvHSP90 Δ C-HvSGT1-CS z ADP	155
5	Dyskusja	157
5.1	Badanie struktury pełnej długości białka SGT1 z jęczmienia	157
5.2	Badanie niskorozdzielczej struktury białka RAR1 z jęczmienia	160

5.3	Zastosowanie metody SVD i MCR-ALS w analizie danych SAXS	161
5.4	Wpływ siły jonowej i temperatury na dynamikę konformacyjną białka HSP90 z pszenicy	162
5.5	Stechiometria i struktura kompleksu białek HSP90 i SGT1 z ADP	165
5.6	Analiza dimeryzacji domeny TPR białka SGT1 z różnych organizmów	168
5.6.1	Modelowanie dimeru domeny TPR białka SGT1	169
6	Podsumowanie	173
7	Literatura	177



1 Cel pracy

Białka HSP90 (ang. heat shock protein 90 kDa), SGT1 (ang. suppressor of G2 allele of *skp1*), RAR1 (ang. required for MLA12 resistance) u roślin są zaangażowane w procesy związane z uruchamianiem odpowiedzi obronnej przeciwko patogenom. Ten sam kompleks jest także zaangażowany w podobne procesy u ludzi. Jak dotąd struktura białek RAR1 i SGT1 z roślin jest bardzo słabo poznana. Pomimo dobrze określonej struktury przestrzennej dla białka HSP90 z bakterii i drożdży wpływ innych białek oddziałujących z tym białkiem na jego strukturę i konformację jest słabo poznany.

Głównym celem niniejszej pracy było zbadanie konformacji i niskorozdzielczej struktury roślinnych białek HSP90, SGT1 i RAR1 w roztworze za pomocą małokątowego rozpraszania promieniowania rentgenowskiego (SAXS) i dichroizmu kołowego. Do tej pory określono strukturę przestrzenną domeny CS białka SGT1 z człowieka i *Arabidopsis thaliana* oraz domeny CHORD II białka RAR1. Ponieważ nie ma informacji na temat jaka jest struktura przestrzenna pełnej długości wyżej wymienionych białek w pracy spróbuje odpowiedzieć na pytanie jaka jest struktura przestrzenna białka RAR1 oraz monomeru i dimeru białka SGT1? Roślinne oraz drożdżowe białko SGT1 ulega dimeryzacji w roztworach o niskiej sile jonowej natomiast ludzkie białko nie ulega dimeryzacji. W pracy chciałbym także scharakteryzować za pomocą małokątowego rozpraszania promieniowania rentgenowskiego proces dimeryzacji i strukturę domeny TPR białka SGT1 z jęczmienia oraz do celów porównawczych struktury tych domen z *Arabidopsis thaliana* i drożdży piekarskich jak również strukturę domeny TPR ludzkiego białka SGT1. Czynniki takie jak pH i osmolity wpływają na konformację białka HSP90. Nie jest wiadomo jednak jaki wpływ mają temperatura i siła jonowa na dynamikę konformacyjną białka HSP90. W trakcie pracy spróbuje także scharakteryzować niskorozdzielczą strukturę roślinnego białka HSP90, określić wpływ temperatury, siły jonowej oraz wiązania nukleotydów na dynamikę konformacyjną białka HSP90 w roztworze i porównać otrzymane wyniki z wynikami badań dla homologów białka HSP90 z innych organizmów. Wiele białek oddziałujących z białkiem HSP90 zmienia jego konformację i modulują jego aktywność hydrolizy ATP. Jak dotąd określono strukturę przestrzenną kompleksu N-terminalnej domeny białka HSP90 w kompleksie z domeną CS białka SGT1 i domeną CHORD II białka RAR1, jednak nie wiadomo jaką strukturę i konformację przyjmuje kompleks pełnej długości białek. W pracy postaram się odpowiedzieć na pytanie jak wygląda niskorozdzielcza struktura kompleksu białek HSP90-SGT1 z ADP w roztworze i czy białko SGT1 wpływa na konformację białka HSP90 oraz jakie może mieć to znaczenie w pełnieniu funkcji przez kompleksu roślinnych białek HSP90-SGT1.



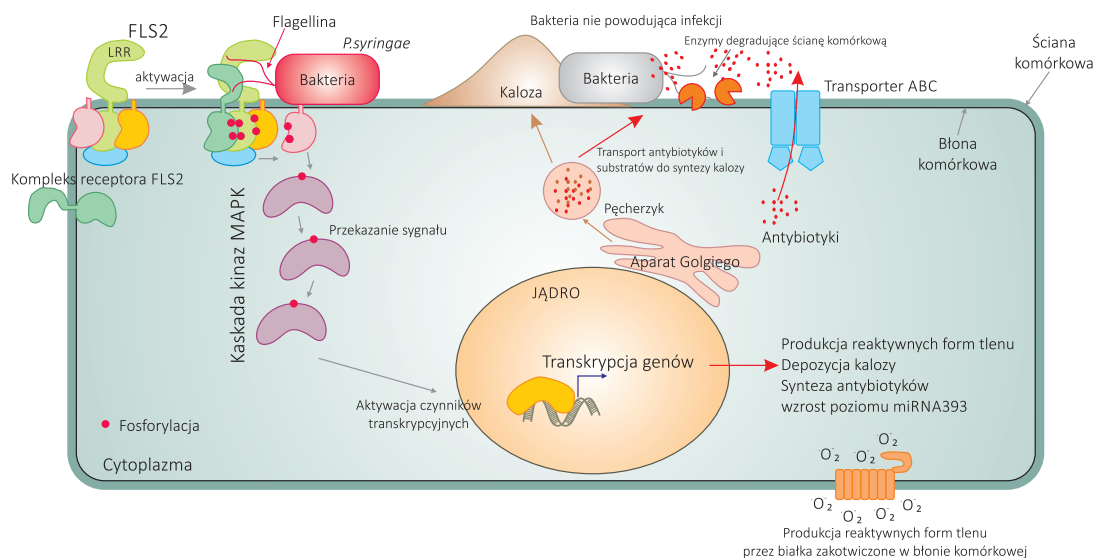
2 Wstęp

2.1 Oddziaływanie roślina-patogen

Rośliny podobnie jak zwierzęta nieustannie narażone są na kontakt z organizmami i czynnikami patogennymi takimi jak wiroidy, wirusy, bakterie, grzyby czy nicienie, które mogą powodować choroby a w ostateczność śmierć rośliny. Jednak w odróżnieniu od wyższych organizmów eukariotycznych nie posiadają one rozwiniętego układu immunologicznego opartego na wyspecjalizowanych komórkach, które mogą zatrzymywać i eliminować potencjalnie szkodliwe organizmy jak również substancje toksyczne. W przypadku roślin szczególnie daje się zauważyć brak swoistej odpowiedzi obronnej opartej na szerokiej gamie przeciwciał wytwarzanych przez limfocyty B, które neutralizują czynnik patogenny. Jednakże, w naturze rzadko spotyka się rośliny porażone chorobą. W toku ewolucji organizmy roślinne wykształciły swoiste mechanizmy obronne, które pozwalają na skuteczną ochronę przed szerokim spektrum patogenów (Dangl i wsp., 2001, Jones i wsp., 2006, Macho i wsp., 2014).

Różne organizmy patogenne, wykorzystują odmienne sposoby na wniknięcie do komórki gospodarza. Bakterie wykorzystują mechaniczne zranienia rośliny jak również aparaty szparkowe umiejscowione w epidermie i służące do wymiany gazowej przez roślinę, do wtargnięcia do przestrzeni między komórkowej – apoplastu. Nicienie i mszyce posługują się kłujką, którą wprowadzają wprost do komórki roślinnej w celu pozyskania substancji odżywczych. Grzyby natomiast mogą wprowadzać wydłużone części strzępek zwane haustoriami pomiędzy lub nawet wprost do komórki roślinnej tworząc w jej wnętrzu system błon. Powierzchnia rośliny jak również ściana komórkowa są pierwszymi punktami styku pomiędzy rośliną a patogenem. Jest to także pierwsza linia obronna rośliny przed patogenami. Powierzchnia epidermy – wierzchniej warstwy komórek np. liścia, pokryta jest mieszaniną wosków będących estrami kwasów tłuszczowych o około 80 atomów węgla w łańcuchu węglowodorowym i wyższych alkoholi monowodorotlenowych oraz kutykulą, składającą się z kutyny (mieszanina spolimeryzowanych kwasów tłuszczowych) oraz kutanu (polimeru węglowodorowego). Kutykula stanowi barierę mechaniczną chroniącą przed wtargnięciem patogenów do komórek epidermy a także, zabezpiecza roślinę przed utratą wody. Zauważono również, że roślina potrafi wytwarzać kalozę (β -glukan) w miejscu oddziaływania z patogenem (Luna i wsp., 2011). Zaobserwowano także, że w pierwszej linii obronnej biorą udział także białka transportujące toksyny – transportery ABC (ang. ATP-binding cassette transporters) (Stein i wsp., 2006). W obronie rośliny przed patogenami wykazano również udział białek zaangażowany w transport pęcherzyków transportujących metabolity we wnętrzu komórki (Collins i wsp., 2003).

Rośliny wytwarzają także szereg złożonych związków chemicznych zwanych fitoantycypinami, często należących do glikoalkaloidów i izoflawonoidów, które są produkowane konstytutywnie i chronią roślinę przed zagrożeniem ze strony patogenów, głównie mikroorganizmów (Van Etten i wsp., 1994).



Rysunek 2.1 Schemat odpowiedzi roślinnej PTI na organizmy patogenne na przykładzie aktywacji receptora FLS2 przez bakterię *Pseudomonas syringae* oraz odpowiedź na niepatogenne szczepy bakterii.

2.1.1 Odpowiedź uruchamiana przez wzorce molekularne związane z patogenami

Organizmy patogenne po pokonaniu wcześniej opisanych barier mogą skolonizować roślinę, lecz dużo częściej ich rozpoznanie przez zainfekowany organizm powoduje wystąpienie odpowiedzi uruchamianej przez wzorce molekularne związane z patogenami - PTI (ang. pathogen associated molecular patterns triggered immunity) (Dangl i wsp., 2001, Zipfel 2008). Schemat odpowiedzi PTI jest przedstawiony na rysunku 2.1 (dokładny opis w tekście). Wzorce molekularne związane z patogenami - PAMPs (ang. pathogen associated molecular patterns) są grupą ewolucyjnie zachowanych cząstek, które występują niezmiennie w szerokim spektrum mikroorganizmów. Typowym „wzorcem molekularnym” jest peptyd flg22 stanowiący N-terminalny fragment białka bakteryjnego flageliny, które jest głównym składnikiem wici służącej do poruszania się bakterii w środowisku wodnym (Zipfel i wsp., 2004, Chinchilla i wsp., 2006). Do zachowanych ewolucyjnie wzorców molekularnych należą również: ksylanaza (Ron i wsp., 2004) – enzym hydrolizujący wiązania β -1,4 glikozydowe w polisacharydzie – ksylanie, chityna u grzybów (Shibuya i wsp., 2001, Kaku i wsp., 2006) oraz heptaglukan (Umemoto i wsp., 1997). Zachowane ewolucyjnie wzorce molekularne rozpoznawane są przez receptory PRR (ang. pattern recognition receptors). Receptory te u roślin są wyłącznie białkami transbłonowymi, posiadającymi pozakomórkową domenę rozpoznającą dany PAMP oraz w większości przypadków cytoplazmatyczną domenę służącą do przekazywania sygnału uruchamiającego odpowiedź PTI, która jest najczęściej domeną o aktywności kinazy (Zipfel 2008). Jednym z najlepiej scharakteryzowanych receptorów PRR jest białko FLS2 rozpoznające peptyd flg22 u *Arabidopsis thaliana* (Zipfel i wsp., 2004). Białko FLS2 zawiera pozakomórkową domenę składającą się z 28 powtórzeń aminokwasowych bogatych w leucynę, która odpowiedzialna jest za rozpoznanie PAMPs oraz cytoplazmatycznej domeny o aktywności kinazy serynowo-treoninowej (Dangl i wsp., 2001). Aktywacja receptorów

proceedzi do uruchomienia PTI charakteryzującej się wytwarzaniem reaktywnych form tlenu, zwiększonym przepływem jonów, aktywacją kaskady kinaz MAP regulujących czynniki transkrypcyjne, biosyntezą hormonów roślinnych, depozycją kalozy na powierzchni epidermy oraz zmianą w ekspresji ponad 900 genów (u *Arabidopsis*) (Macho i wsp., 2014).

Rośliny produkują małowcząsteczkowe hormony takie jak etylen, kwas salicylowy czy kwas jasmonowy, które między innymi regulują odpowiedź obronną. Kwas salicylowy jest zaangażowany w wytwarzanie systemowej odpowiedzi nabytej SAR (ang. systemic acquired resistance), która indukuje odporność na czynnik patogenny w tkankach oddalonych od miejsca jego percepcji (Durrant i wsp., 2004, Fu i wsp., 2013). Rozpoznanie peptydu flg22 czy LPS (lipopolisacharydu) powoduje indukcję SAR. Do kolejnej grupy hormonów roślinnych zaliczamy auksyny. Auksyny zaangażowane są w wiele procesów związanych ze wzrostem i rozwojem rośliny (Teale i wsp., 2006). Wiele organizmów patogennych takich jak bakterie *Agrobacterium tumefaciens* wykorzystują auksyny lub ich analogii do przeprogramowania rośliny w celu wytworzenia warunków jak najbardziej sprzyjających ich kolonizacji (Thomashow i wsp., 1984).

W odpowiedzi PTI zaangażowane są również cząsteczki mikro RNA (miRNA). Cząsteczki miRNA są cząsteczkami RNA o długości 21-24 nukleotydów, które powstają poprzez wycięcie ze struktury spinki do włosów występującej w transkryptach genów kodujących miRNA (Sobkowiak i wsp., 2008). Biorą one udział w post-transkrypcyjnym wyciszaniu i regulacji ekspresji genów. Cząsteczki te są komplementarne do mRNA kodujących funkcjonalne białka lub cząsteczki RNA. Enzymy należące do rodziny AGO (ang. Argonaute) w kompleksie z miRNA przyłączają się do docelowego mRNA i katalizują jego rozcięcie w miejscu komplementarnym lub hamują ich translację. Aktywacja FLS2 powoduje wzrost poziomu cząsteczki miRNA393, która jest zaangażowana w degradację receptora auksyn TIR-1 (Navarro i wsp., 2006). Aktywacja miRNA393 prowadzi więc do obniżenia poziomu auksyn i wzrostu odporności na infekcję bakterią *Pseudomonas syringae* u *Arabidopsis*.

2.1.2 Odpowiedź uruchamiana przez efekторы

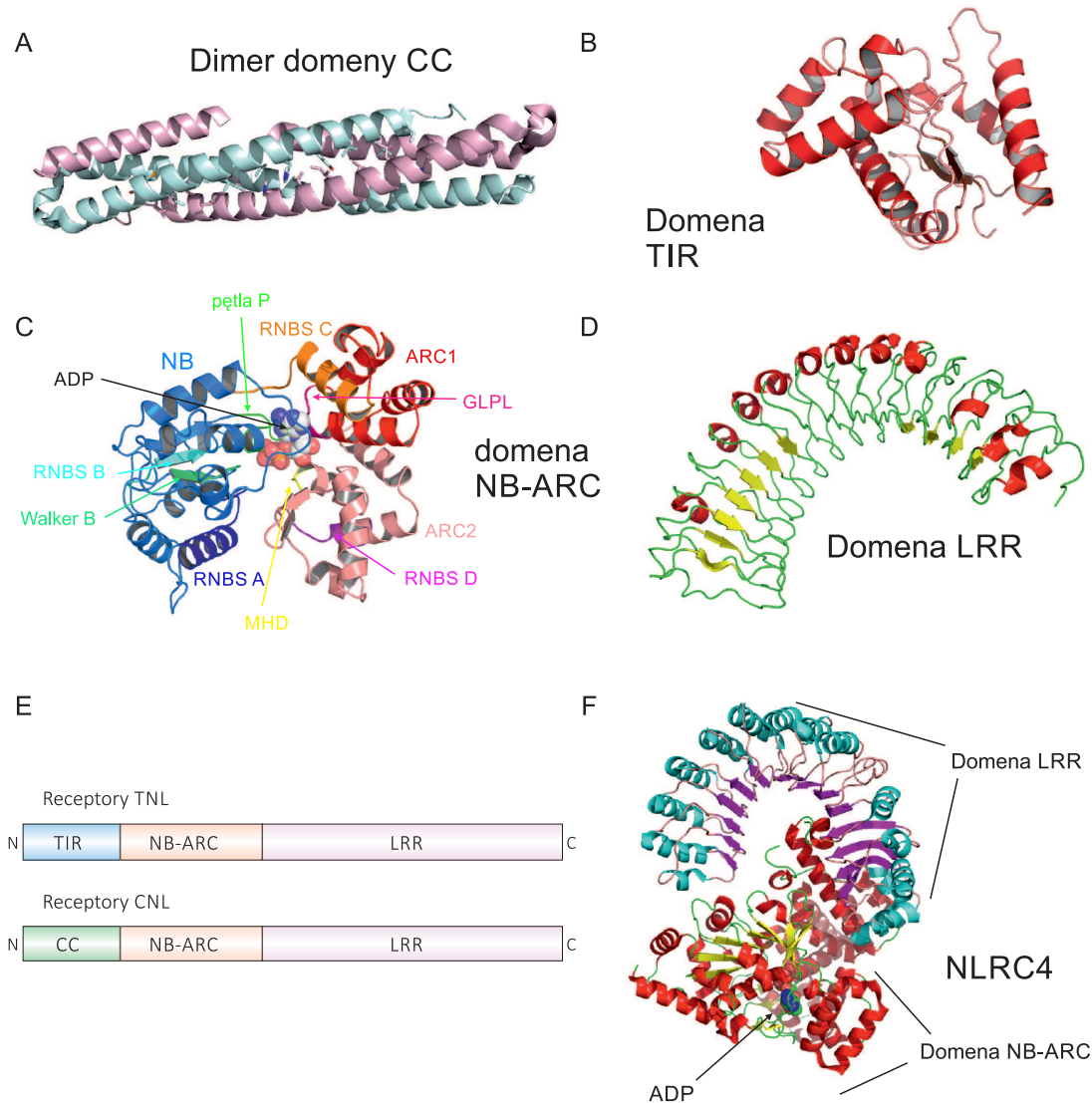
W większości przypadków odpowiedź PTI i naturalne bariery ochronne rośliny są w stanie efektywnie powstrzymać organizmy patogene przed kolonizacją rośliny. W toku ewolucji organizmy patogene nauczyły się jednak osłabiać odpowiedź PTI poprzez manipulację odpowiedzią PTI na poziomie molekularnym (Jones i wsp., 2006, Göhre i wsp., 2008, Deslandes i wsp., 2012). W tym celu wprowadzają one do komórki gospodarza cząsteczki zwane efektorami, które modyfikują komponenty odpowiedzi PTI. Schemat podsumowujący działanie wybranych efektorów przedstawiony jest na rysunku 2.2 (opis w dalszej części tekstu). Bakteryjne cząsteczki efektorów są wprowadzane do komórki gospodarza przez skomplikowane struktury białkowe zwane aparatem sekrecyjnym. Aparat sekrecyjny tworzy kanał, który łączy bakterię z komórką roślinną, który umożliwia transport przez błonę komórkową cząsteczek efektorów (Enninga i wsp., 2008). Składa się on z części zakotwiczonej w błonie komórkowej bakterii oraz struktury o charakterze „igły”. Aparat sekrecyjny wielu bakterii kolonizujących komórki roślinne należy do grupy aparatów sekrecyjnych typu trzeciego TTSS lub T3SS (ang. type three/3 secretion system), który wykazuje wysokie podobieństwo do bakteryjnej wici służącej do poruszania się w środowisku wodnym. W przypadku grzybów efekторы są dostarczane przez system błon do przestrzeni międzykomórkowej. Głównym zadaniem efektorów jest zablokowanie

nie odpowiedzi PTI. Bakterie i grzyby dostarczają efektorów, które funkcjonalnie reprezentują wiele rodzin białkowych np. proteaz czy ligaz ubikwityny E3 (białka biorące w procesie degradacji innych białek). Najlepiej opisane efekторы pochodzą z bakterii *Pseudomonas syringae* a oddziaływania z roślinami takimi jak *Arabidopsis thaliana*, pomidor czy tytoń są modelowym układem w badaniu interakcji patogenów z gospodarzem. Efekторы pozwalają na pokonanie barier obronnych ograniczających wzrost patogenu. Poniżej zostaną opisane przykłady w jaki sposób cząsteczki efektorów przezwyciężają barierę obronną rośliny.

Bakteryjna toksyna koronatylna – analog kwasu jasmonowego, hormonu roślinnego regulującego m.in. otwarcie aparatów szparkowych, powoduje aktywację szlaku sygnalizacyjnego związanego z kwasem jasmonowym prowadząc do otwarcia aparatu szparkowego (Melotto i wsp., 2006). Poprzez otwarte aparaty szparkowe bakterie mogą łatwo dostać się do przestrzeni międzykomórkowej. Jednak zamknięcie aparatów szparkowych może być również korzystne dla bakterii szczególnie w okresie suszy, czy w sytuacji kiedy kilka szczepów bakterii musi konkurować ze sobą, ponieważ różne gatunki bakterii preferują różne warunki do rozwoju. Innym roślinnym hormonem, biorącym udział w regulacji aparatów szparkowych jest kwas abscysynowy ABA (ang. abscisic acid). ABA powoduje zamknięcie aparatów szparkowych, szczególnie w stresie suszy (Finkelstein 2013). Efektor HopAM1 wzmacnia sygnał kwasu abscysynowego powodując zamknięcie aparatów szparkowych i deregulację ekspresji genów odpowiedzialnych za odpowiedź obronną rośliny (Goel i wsp., 2008) powodując, że rośliny są bardziej podatne na kolonizację patogenem.

Inne efekторы potrzebne są do degradacji ściany komórkowej w celu umieszczenia aparatu TTSS. Bakteria *Xanthomonas oryzae pv. oryzae* wydziela enzymy degradujące ścianę komórkową do apoplastu (Goel i wsp., 2008). Poza enzymami degradującymi ścianę komórkową wydzielane są również inhibitory proteaz, które blokują występujące w apoplastie roślinne enzymy proteolityczne (Göhre i wsp., 2008). Bakterie nauczyły się także blokować sygnał uruchamiający PTI już na wczesnym jego etapie. Dwa efekторы z bakterii *Pseudomonas syringae* atakującej pomidora: AvrPto (Avr ang. avirulence, Pto ang. *Pseudomonas syringae* tomato) i AvrPtoB mają zdolność do wiązania się i blokowania aktywności kinaz zaangażowanych w PTI np. FLS2 (Xiang i wsp., 2008). AvrPto i AvrPtoB oddziałują z kinazami Pto i Fen, które wiążą się z białkiem Prf (ang. Pto resistance and fenthion sensitivity), białkiem NB-LRR uruchamiającym odpowiedź obronną ETI. AvrPtoB może kierować kinazę Fen do degradacji (Kim i wsp., 2002). Liczne efekторы wiążą się i/lub modyfikują białko RIN4 (ang. RPM1 interacting protein 4) z *Arabidopsis thaliana*, które jest negatywnym regulatorem odpowiedzi PTI (Göhre i wsp., 2008). Innym ciekawym efektozem jest HopF2 również z bakterii *Pseudomonas syringae*. HopF2 jest mono-ADP rybozylazą, która dodaje ADP-rybozę do kinazy ze szlaku kinaz MAPK przenoszących sygnał uruchamiający PTI we wnętrzu komórki i blokuje jej aktywność (Wang i wsp., 2010a). Efektorami są również czynniki transkrypcyjne, które pozwalają na bezpośrednie przeprogramowanie komórki roślinnej. Do tej klasy efektorów należą aktywatory transkrypcji TAL (ang. transcription activator-like) z bakterii *Xanthomonas* np. białko Xo1. Efekторы TAL składają się z 34 aminokwasowych powtórzeń, każde rozpoznające jeden nukleotyd w podwójnej helisie DNA (Deng i wsp., 2012, Mak i wsp., 2012). Efekторы TAL zostały także niedawno użyte jako narzędzia mogące służyć do edycji genomu w inżynierii genetycznej i terapii genowej (Miller i wsp., 2011).

wiedzi HR. Wszystkie białka R uruchamiają ten sam rodzaj odpowiedzi HR jednakże elementy szlaku sygnalizacyjnego wydają się być różne dla różnych receptorów. Dwie klasy receptorów R różniących się domeną N-terminalną różnią się zasadniczo elementami szlaku sygnalizacyjnego (Meyers i wsp., 2002, McHale i wsp., 2006, Qi i wsp., 2013).



Rysunek 2.3. Struktura przestrzenna receptorów R. **A.** Struktura przestrzenna dimeru domeny CC białka MLA10 z jęczmienia (Maekawa i wsp., 2011, kod pdb: 3QFL). **B.** Struktura przestrzenna domeny TIR białka N z tytoniu (otrzymana w programie I-Tasser). **C.** Struktura przestrzenna domeny NB-ARC białka (otrzymana w programie I-Tasser). **D.** Struktura przestrzenna domeny LRR białka N z tytoniu (otrzymana w programie I-Tasser). **E.** Struktura domenowa dwóch klas białek R: TNL i CNL. **F.** Struktura przestrzenna mysiego białka NLRC4 w kompleksie z ADP (Hu i wsp., 2013, kod pdb: 4KXF).

Pierwsza klasa białek TNL (białka składające się z domen: TIR-NBS-LRR, TIR - Toll/interleukin-1 receptor homology domain, NBS - nucleotide binding site, LRR - leucine rich repeats) wymaga białka EDS (ang. enhanced disease susceptibility 1) i aktywuje czynnik transkrypcyjny SPL6 (ang. Squamosa promoter-binding protein-like 6) (Jacob i wsp., 2013, Padmanabhan i wsp., 2013). Druga klasa białek CNL (zawierająca domenę typu coiled-coil na N-końcu zamiast domeny TIR) potrzebuje białka NPR1 (ang. nonexpresser of PR genes) i aktywuje czynniki transkrypcyjne z rodziny MYB (ang. MYB (myeloblastosis) domain containing proteins) oraz inaktywuje represory transkrypcji WRKY (Jacob i wsp., 2013). Dla obu klas białek NB-LRR zaobserwowano również zmianę lokalizacji komórkowej podczas infekcji patogenami. Wykazano dla kilku białek R, że zmiana ta jest niezbędna dla pełnej aktywacji odpowiedzi HR.

Prawidłowe funkcjonowanie białek NB-LRR zależy od wielu czynników, które regulują pozytywnie bądź negatywnie odpowiedź ETI. Do tej klasy białek należy kompleks białek opiekuńczych: **białka szoku cieplnego HSP90 (ang. heat shock proein 90 kDa), RAR1 (ang. required for MLA12 resistance 1) oraz SGT1 (ang. suppressor of the G2 allele of SKP1)**, które stanowią obiekty badań niniejszej pracy (Kadota i wsp., 2010). Białka HSP90 oraz RAR1 są pozytywnymi regulatorami natomiast SGT1 zarówno pozytywnym jak i negatywnym regulatorem poziomu białek NB-LRR. Ponadto białko SGT1 oddziałuje z komponentami kompleksu ligazy ubiquityny zaangażowanymi w procesy degradacji białek poprzez kompleks proteasomu. Aktywacja receptorów NB-LRR prowadzi do odpowiedzi HR. W przypadku białek TNL wydaje się, że aktywacja odpowiedzi zachodzi poprzez oddziaływanie z represorami transkrypcji. Jednocześnie stwierdzono, że niektóre receptory posiadają na N-końcu dodatkową domenę WRKY np. RRS1-R (ang. resistance to *Ralstonia solanacearum* 1) z *Arabidopsis*, charakterystyczną dla czynników transkrypcyjnych biorących udział w odpowiedzi przeciwko patogenom (Caplan i wsp., 2008).

2.1.4 Struktura domenowa białek R

Najważniejszą cechą białek R jest ich ewolucyjnie zachowana, modułowa budowa domenowa. W sekwencji receptorów R możemy wyodrębnić trzy domeny: domenę N-terminalną, środkową domenę NB-ARC (ang. nucleotide-binding adaptor shared by Apaf-1, resistance proteins and CED-4) oraz C-terminalną domenę zawierającą powtórzenia bogate w leucynę LRR (ang. leucine rich repeats) (Meyers i wsp., 2003, Jones i wsp., 2006, Caplan i wsp., 2008). Schemat struktury domenowej białek R oraz struktury przestrzenne domen białek R przedstawione są na rysunku 2.3. Ze względu na rodzaj domeny N-terminalnej możemy wyróżnić dwie klasy białek R: TNL oraz CNL. Schemat struktury domenowej białek R przedstawiony jest na rysunku 2.3 E.

Białka TNL posiadają na N-końcu domenę TIR. Domena ta występuje u ssaków w receptorach TLR (ang. Toll-like receptors), które stanowią ważną część wrodzonego systemu immunologicznego oraz w receptorze Interleukiny 1 i stanowi platformę do oddziaływań białko-białko. Wykazano również, że domena ta odpowiedzialna jest za dimeryzację białek TNL (Mestre i wsp., 2006, Bernoux i wsp., 2011). Struktura domeny TIR z białka L6 jest bardzo podobna do struktury białek zawierających domenę TIR np. ludzkiego białka MyD88 i TLR4 (Gay i wsp., 2014). Domena TIR należy do białek typu α/β gdzie wewnątrz białka stanowi równoległy arkusz beta o skręconej powierzchni, otoczony przez helisy alfa (rysunek 2.3 B). W odróżnieniu jednak od białek zaangażowanych w od-

powieź immunologiczną u ssaków, roślinne domeny TIR posiadają dodatkowy motyw helisa-skret-helisa (Chan i wsp., 2010).

W białkach CNL, N-terminalną domeną jest domena o strukturze typu super-helisy, CC (ang. coiled-coil) (rysunek 2.3 A). Jak dotąd udało się jedynie rozwiązać strukturę dla jednej domeny typu CC pochodzącej z receptora MLA10 z jęczmienia (Maekawa i wsp., 2011). Podobnie jak w przypadku domeny TIR, w strukturze krystalograficznej domena CC wstępuje w formie dimeru. Każdy z monomerów posiada strukturę typu helisa-zwrot-helisa, gdzie druga helisa składa się z dwóch mniejszych przedzielonych krótkim, niepofałdowanym fragmentem. Izolowana domeny CC i TIR, wystarczają do aktywacji odpowiedzi obronnej i są odpowiedzialne za oddziaływanie z czynnikami transkrypcyjnymi (Shen i wsp., 2007, Padmanabhan i wsp., 2013).

Środkowa domena NB-ARC posiada podobieństwo sekwencji aminokwasowej do białek apoptosomu tj. ludzkiego i mysiego białka Apaf-1 (ang. apoptotic protease activating factor 1) oraz CED4 (ang. cell death protein 4) u nicienia *Caenorhabditis elegans* (van der Biezen i wsp., 1998). Apoptosom jest kompleksem białkowym, który uruchamia programowaną śmierć komórki (Riedl i wsp., 2007). Struktura przestrzenna domeny NB-ARC wraz z zaznaczonymi zachowanymi motywami w jej obrębie przedstawiona jest na rysunku 2.3 C. Mutacje w tych motywach w wielu białka R negatywnie wpływają na ich funkcjonowanie.

Najbardziej polimorficznym fragmentem białek R jest domena LRR (Takken i wsp., 2012). Domena LRR składa się z kilku 20-30 aminokwasowych powtórzeń bogatych w leucynę. Każde powtórzenie zawiera fragment wstęgi beta, która tworzy rozległy arkusz beta z sąsiednimi powtórzeniami oraz najczęściej helisę alfa (Kobe i wsp., 1993). Powtórzenie składa się z dwóch zakonserwowanych fragmentów: części hydrofobowej nadającej strukturę oraz części zmiennej, która nadaje specyficzność w oddziaływaniach białko-białko. Kilka powtórzeń tworzy razem strukturę typu podkowy.

W obecnie przyjętym modelu działania pod nieobecność cząsteczek efektorów lub zmienionego białka roślinnego, białka NB-LRR związane są z ADP i posiadają zamkniętą konformację podobną do struktury ludzkiego białka NLRC4 (Hu i wsp., 2013). Białko NLRC4 w kompleksie z ADP posiada zamkniętą konformację, w której to C-terminalna domena LRR oddziałuje z domeną NB-ARC (rysunek 2.3 F). Po rozpoznaniu cząsteczki efektoru następują zmiany konformacyjne prowadzące do aktywacji receptora, uwolnienia domeny NB-ARC od oddziaływania z domeną LRR i związania ATP. Równocześnie domena LRR jest w stanie oddziaływać z innymi partnerami białkowymi. Tak aktywowany kompleks przyjmuje ostatecznie konformację, która umożliwia oddziaływanie z innymi białkami i przekazanie sygnału do uruchomienia ETI.

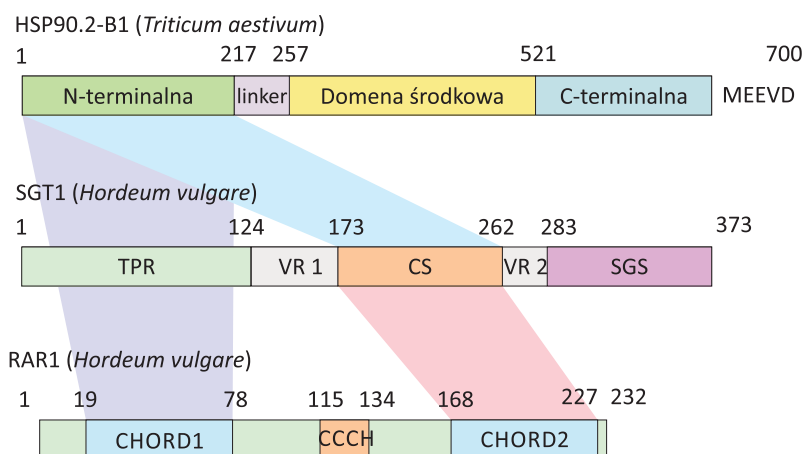
2.2 Rola białek HSP90, SGT1, RAR1 w odpowiedzi obronnej skierowanej przeciwko patogenom

Jako pierwsze z kompleksu białek HSP90/SGT1/RAR1 zostało odkryte i scharakteryzowane białko RAR1 (ang. required for Mla12 resistance) (Shirasu i wsp., 1999). Gen *RAR1* zidentyfikowano po raz pierwszy u jęczmienia w eksperymencie mającym na celu znalezienie czynników genetycznych zaangażowanych w odpowiedź immunologiczną skierowaną przeciwko grzybowi *Blumeria graminis f.sp. hordei* powodującego mączniaka prawdziwego atakującego rośliny jęczmienia. Jednocześnie autorzy tej pracy zidentyfikowali homologi białka RAR1 w licznych gatunkach niższych i wyższych

eukariotów z wyjątkiem drożdży, które nie posiadają genu kodującego białko RAR1. U nicienia *Caenorhabditis elegans* wyciszenie genu kodującego homolog białka RAR1, CHP1 (ang. CHORD containing protein 1) powoduje wady rozwojowe oraz letalność embrionów (Shirasu i wsp., 1999). Niedługo po zidentyfikowaniu białka RAR1 w jęczmieniu wykazano, że białko RAR1 jest zaangażowane m.in. w uruchamianie odpowiedzi przeciwko bakteriom *Pseudomonas syringae pv tomat* DC3000 oraz grzybom *Peronospora parasitica* u *Arabidopsis thaliana* (Muskett i wsp., 2002). W tytoniu eksperymenty z wyciszeniem genu *RAR1* poprzez technikę VIGS (ang. virus induced gene silencing) wykazały, że białko RAR1 jest zaangażowane w odpowiedź odpornościową skierowaną przeciwko wirusowi mozaiki tytoniu TMV (ang. tobacco mosaic virus) zależną od receptora NB-LRR N (Liu i wsp., 2002b). Rola białka RAR1 na poziomie molekularnym nie jest do końca poznana aczkolwiek wydaje się, że białko RAR1 kontroluje poziom niektórych białek R np. MLA1 i MLA6 z jęczmienia (Bieri i wsp., 2004). Poprzez oddziaływanie z małą GTP-azą RAC1, białko RAR1 może regulować proces syntezy reaktywnych form tlenu, ponieważ białko RAC1 jest pozytywnym regulatorem tego procesu (Thao i wsp., 2007).

Białko SGT1 zostało zidentyfikowane w eksperymencie poszukiwania białek oddziałujących z białkiem RAR1 z *Arabidopsis* oraz z jęczmienia (Azevedo i wsp., 2002, Austin i wsp., 2002). Rola białka SGT1 jest synergistyczna z rolą białka RAR1, ponieważ podwójne wyciszenie ekspresji genów białek SGT1 i RAR1 podwyższa wrażliwość na infekcje w stosunku wyciszenia ekspresji tylko jednego z tych białek. Wykazano, że podobnie jak białko RAR1, białko SGT1 jest potrzebne do poprawnego funkcjonowania odpowiedzi immunologicznej zależnej od receptorów NB-LRR. Pokazano, że białko SGT1 jest niezbędne do aktywacji kinaz zależnych od wapnia CDPK (Ludwig i wsp., 2005). Wykazano również, że białko SGT1 oddziałuje z białkami HSP70 oraz HSP90 i podobnie jak białko HOP może łączyć te dwa kompleksy białek opiekuńczych (Takahashi i wsp., 2003, Noël i wsp., 2007).

Jako trzecie białko z kompleksu HSP90/SGT1/RAR1 zostało scharakteryzowane białko szoku cieplnego 90 kDa HSP90 (ang. heat shock protein 90 kDa) przez grupę prof. Shirasu (Takahashi i wsp., 2003). Wykazali oni, że białko HSP90 oddziałuje bezpośrednio z białkiem RAR1 jak również z białkiem SGT1 z jęczmienia. Ponadto, geldanamycyna – małowcząsteczkowy inhibitor aktywności ATPazy białka HSP90 blokuje uruchamianie odpowiedzi HR co dodatkowo potwierdza udział białka HSP90 w procesie uruchamiania odpowiedzi HR. Pokazano również, że białko HSP90 oddziałuje z białkami RAR1 i SGT1 u innych gatunków roślin np. w dzikim tytoniu (Liu i wsp., 2004, Shirasu 2009). Uważa się, że białko HSP90 wraz z białkami SGT1 i RAR1 bierze udział w stabilizacji białek NB-LRR w nieaktywnej konformacji podczas nieobecności patogenu. Wykazano, że za oddziaływanie białek NB-LRR z kompleksem białek HSP90-SGT1-RAR1 odpowiedzialna jest domena LRR (Beri i wsp., 2004, Liu i wsp., 2004, Leister i wsp., 2005). Ponieważ białko SGT1 oddziałuje z komponentami ligazy ubikwityny SCF (Kitagawa i wsp., 1999, Liu i wsp., 2002b), uważa się, że białko SGT1 może rekrutować maszynę degradacji białek do kompleksów NB-LRR związanych z kompleksem białek HSP90-RAR1-SGT1. Inna hipoteza zakłada, że białko SGT1 w kompleksie z białkami HSP90 i RAR1 promuje oddziaływanie domen NB i LRR receptorów R, utrzymując je w zamkniętej, nieaktywnej konformacji. Większość białek R wymaga białka HSP90 do pełnienia swojej funkcji jednakże część jest niezależna od białek RAR1 lub SGT1, a niekiedy od obu białek (Shirasu 2009, Seo i wsp., 2008).



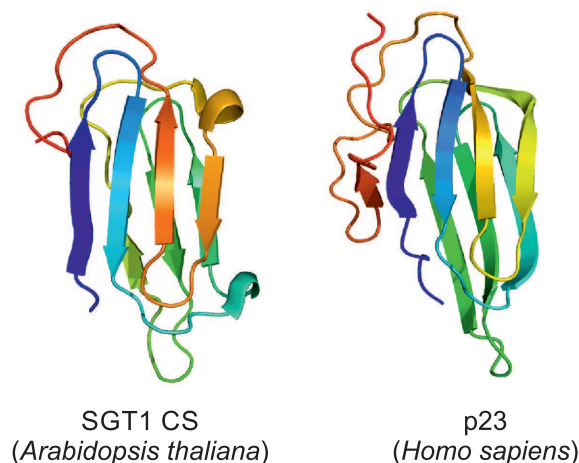
Rysunek 2.4. Schemat budowy domenowej białek HSP90 (numeracja aminokwasów białka HSP90 z pszenicy), SGT1 (schemat numeracji aminokwasów białka z jęczmienia), RAR1 (schemat numeracji aminokwasów białka z jęczmienia) oraz ich oddziaływań (dla białek z jęczmienia, Takahashi i wsp., 2003).

2.2.1 Budowa i funkcja białka SGT1

Białko SGT1 posiada zakonserwowaną budowę domenową i składa się z trzech domen: N - terminalnej domeny zawierającej trzy powtórzenia TPR (ang. tetratricopeptide repeat), domeny CS (ang. CHORD and SGT1 proteins domain) oraz C-terminalnej domeny SGS (ang. SGT1 specific domain) (Azevedo i wsp., 2002) (rysunek 2.4). Pomiędzy domenami znajdują się fragmenty dużo mniej zachowane ewolucyjnie nazywane w literaturze regionami zmiennymi (ang. variable regions) - VR1 (pomiędzy domenami TPR i CS) oraz VR2 (pomiędzy domenami CS i SGS). Domena TPR występuje w wielu białkach i odpowiedzialna jest za oddziaływania typu białko-białko. Jak na razie zidentyfikowano jedynie dwa białka, które oddziałują z domeną TPR białka SGT1. Pierwszym białkiem jest białko Skp1p u drożdży, które jest zaangażowane w degradację białek (Kitagawa i wsp., 1999). Roślinny oraz ludzki homolog Skp1 także oddziałują z białkiem SGT1 (Liu i wsp., 2002b, Zhang i wsp., 2008). W roślinie *Arabidopsis thaliana* dodatkowo zidentyfikowano białko SRFR1 (ang. suppressor of rps4-RLD1), które zawiera powtórzenia TPR i jest odpowiedzialne za wiązanie się z domeną TPR białka SGT1 (Li i wsp., 2010). SRFR1 negatywnie reguluje poziom białek NB-LRR w celu zabezpieczenia rośliny przed autoagresją. Domena TPR jest również charakterystyczna dla różnych białek oddziałujących z motywem MEEVD na C-końcu białka HSP90. Domena TPR białka SGT1 jednakże nie oddziałuje z białkiem HSP90 (Takahashi i wsp., 2003). Dodatkowo jest ona odpowiedzialna za akumulację białek SGT1a i SGT1b u *Arabidopsis thaliana* (Azvedo i wsp., 2006). Poziom białka SGT1b w komórce jest wyższy niż poziom białka SGT1a. Wykazano również, że domena TPR nie jest niezbędna w procesie odpowiedzi immunologicznej w roślinach *Arabidopsis* (Azvedo i wsp., 2006). Nyarko i współpracownicy wykazali, że białko SGT1 z jęczmienia, *Arabidopsis* i drożdży piekarskich (ale nie ludzkie białko) tworzą dimery w roztworze (Nyarko i wsp., 2007). Wykazali oni, że za dimeryzację odpowiedzialna jest domena TPR. Dimeryzacja białka SGT1

z jęczmienia jest zależna od jego stężenia oraz od siły jonowej, której wzrost powoduje przesunięcie równowagi monomer-dimer w stronę monomeru. Dimeryzacja domeny TPR jest także zależna od stanu redoks. Oksydacja zakonserwowanych cystein powoduje zahamowanie dimeryzacji. W modelu domeny TPR dwie cysteiny 84 i 117 są w bliskiej odległości od siebie i potencjalnie mogą tworzyć mostek dwusiarczkowy, który zmienia konformację domeny TPR i uniemożliwia dimeryzację. W odróżnieniu od drożdżowych i roślinnych białek SGT1 ludzki homolog nie ulega dimeryzacji (Nyarko i wsp., 2007). Rzucającą się w oczy różnicą jest brak naładowanych aminokwasów na C-końcu domeny TPR. Eksperyment chemicznego sieciowania czynnikiem sieciującym grupy boczne lizyny wraz z analizą spektrometrii mas pokazał, że przypuszczalne miejsce dimeryzacji znajdują się właśnie na C-końcu domeny TPR. Funkcja dimeryzacji u roślin nie została dotychczas wykazana. Jedynie u drożdży piekarskich wykazano, że dimeryzacja domeny TPR jest potrzebna do tworzenia się kinetochoru i podziału komórki (Bansal i wsp., 2009a). Intrygującą obserwacją jest blokowanie dimeryzacji SGT1 poprzez fosforylację seryny 361 w obrębie domeny SGS (Bansal i wsp., 2009b). Fosforylacja ta powoduje osłabienie dimeryzacji i zaburza funkcjonowanie białka SGT1. Wykazano, że domena TPR białka CHIP wiąże ufosforylowany fragment białka Smad1 (Wang i wsp., 2011). Podobnie, fosforylacja domeny SGS mogłaby powodować wiązanie jej przez domenę TPR i zahamować dimeryzację. Nyarko i współpracownicy, zaobserwowali również, że białka SGT1 z drożdży i *Arabidopsis*, w sączeniu molekularnym wykazują masę molekularną większą niż białko SGT1 z jęczmienia. Nie wiadomo czy związane jest to z odmiennym stanem oligomerycznym czy też rozbieżności wynikają z właściwości fizyko-chemicznych i różnic sekwencyjnych lub konformacyjnych.

Domena CS wykazuje podobieństwo sekwencyjne i strukturalne do białka p23. Białko p23 u ludzi i drożdży oddziałuje z N-terminalną domeną ATPazy białka HSP90 i dodatkowo powoduje znaczące obniżenie jej aktywności enzymatycznej (Dittmar i wsp., 1998, McLaughlin i wsp., 2006). Białko p23 utrzymuje białko HSP90 w zamkniętej konformacji ze związanym ATP, umożliwiając zatrzymanie związanego białka-klienta i umożliwienie mu przyjęcia poprawnej konformacji. Strukturalnie białko p23 i domena CS białka SGT1 składają się z dwóch naprzeciwległych, antyrównoległych arkuszy beta tworzących strukturę typu beta-kanapki (Weaver i wsp., 2000, Botër i wsp., 2007) (rysunek 2.5). Porównanie rozkładu ładunku na powierzchni obu białek wykazuje duże różnice, które mogą tłumaczyć różnice funkcjonalne pomiędzy tymi białkami. Domena CS białka SGT1 jest odpowiedzialna za wiązanie się z domeną N-terminalną białka HSP90 u roślin, drożdży i człowieka (Kadota i wsp., 2010). W odróżnieniu jednak od białka p23, domena CS nie blokuje aktywności ATPazy i wiąże się w innym miejscu na powierzchni białka HSP90 (Kadota i wsp., 2008, Zhang i wsp., 2008). Pełnej długości białka, zapewne przez zawady steryczne, nie mogą jednocześnie wiązać się do dimeru HSP90 (Kadota i wsp., 2008). Ponadto, domena CS jest zaangażowana, w przeciwieństwie do domeny TPR, w odpowiedź immunologiczną u roślin (Botër i wsp., 2007).



Rysunek 2.5 Porównanie struktury przestrzennej domeny CS białka SGT1b z *Arabidopsis thaliana* (Zhang i wsp., 2008, kod pdb: 2JKI) oraz ludzkiego białka p23 (Weaver i wsp., 2000, kod pdb: 1EJF).

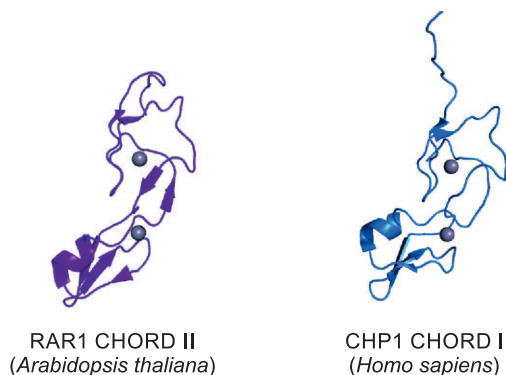
Botër wraz ze współpracownikami zidentyfikowali w białku SGT1 z *Arabidopsis* aminokwasy zaangażowane w wiązanie się do białka HSP90 oraz wymagane do akumulacji białka Rx, receptora NB-LRR zaangażowanego w odporność na wirusa PVX w ziemniaku (Botër i wsp., 2007). Zidentyfikowane mutacje częściowo pokrywają się z miejscem wiązania się z białkiem HSP90. Mutanty domeny CS białka SGT1 pozbawione aktywności wiązania do białka HSP90 negatywnie wpływają na poziom białka Rx jak również osłabiają odpowiedź obronną. U ludzi domena CS jest zaangażowane w tworzenie kompleksu z białkiem HSP90 oraz receptorami NB-LRR Nod1 (ang. nucleotide-binding oligomerization domain-containing protein 1) i NALP3 (ang. NACHT, LRR and PYD domains-containing protein 3) (Mayor i wsp., 2007, da Silva Correia i wsp., 2007). Domena CS białka SGT1 oddziałuje z domenami CHORD (ang. cysteine and histidine rich domain), które występują w białku RAR1 oraz jego homologach u ssaków Chp1 (ang. CHORD containing protein 1) i meluzynie (Takahashi i wsp., 2003, Sbroggiò i wsp., 2008, Hong i wsp., 2013). Oddziaływanie to jest zakonserwowane zarówno wśród roślin jak i ssaków. Domena CS występuje także w białkach Chp1 oraz meluzynie i wykazuje aktywność antyagregacyjną (Sbroggiò i wsp., 2008, Michowski i wsp., 2010). Białko SGT1 też wykazuje tę samą aktywność, a ludzki gen kodujący białko SGT1 posiada w swoim promotorze elementy odpowiedzi na szok cieplny (Zabka i wsp., 2008). Nie jest wiadomo jednak czy domena CS białka SGT1 jest także odpowiedzialna za właściwości chaperonowe.

Trzecią domeną białka SGT1 jest domena SGS. Nie posiada ona podobieństwa sekwencyjnego do innych białek poza ludzkim białkiem wiążącym kalcyklinę CacyBP (ang. Calcyclin binding protein) należącym do rodziny białek wiążących wapń (Nowotny i wsp., 2000). Strukturalnie domena ta wykazuje cechy białka natywnie nieuporządkowanego. Doskonale jest to widoczne w eksperymentach spektroskopii magnetycznego rezonansu jądrowego oraz dichroizmu kołowego (Nowotny i wsp., 2000). Jednakże, C-terminalny fragment białka CacyBP, posiadający podobieństwo sekwencyjne do

domeny SGS białka SGT1 w kompleksie z kalcykliną posiada strukturę alfa helikalną. Fragment ten składa się z dwóch helis przedzielonych krótką pętlą. Obie helisy leżą niemal w jednej płaszczyźnie i są ułożone pod kątem 90° . Istnieje zatem możliwość porządkowania się domeny SGS pod wpływem wiązania z innymi białkami lub w zatłoczonym komórkowym środowisku. Sekwencja domeny SGS w białku SGT1 z różnych organizmów wykazuje największe zachowanie ze wszystkich jego domen (Azevedo i wsp., 2002). Botër i współautorzy (2007) w pracy dotyczącej struktury i funkcji domeny CS w odpowiedzi obronnej na wirusa PVX w tytoniu zidentyfikowali wiele mutacji w obrębie domeny SGS, które powodują obniżenie poziomu białka Rx i utratę odporności. Wiele z nich była także negatywnie-dominująca. Żadna z tych mutacji nie powodowała zahamowania oddziaływania z białkiem HSP90. Podobnie jak domena CS domena SGS jest zaangażowana w wiązanie białka SGT1 do receptorów NB-LRR Nod1 i NALP3 u ludzi (Mayor i wsp., 2007, da Silva Correia i wsp., 2007). Co ciekawe, zidentyfikowano również bezpośrednie oddziaływania domeny SGS z domenami LRR białek nie związanymi w procesy immunologiczne. U drożdży domena SGS jest zaangażowana w aktywację cykazy adenylationowej Cdc35p (ang. cell division control 35) (Dubacq i wsp., 2002) a także w regulację dimeryzacji domeny TPR SGT1 poprzez ufosforylowaną serynę 361 (Bansal i wsp., 2009b). U roślin dodatkowo, poza licznymi przykładami bezpośredniego zaangażowania tej domeny w procesy odporności, wykazano udział białka SGT1, poprzez domenę SGS, w sygnalizacji związanej z hormonem auksyną (Gray i wsp., 2003). Grupa Prof. Henniga z Instytutu Biochemii i Biofizyki PAN w Warszawie wykazała, że białko SGT1 w tytoniu jest zaangażowane w proces relokalizacji receptora NB-LRR N podczas infekcji wirusem mozaiki tytoniu (Hoser i wsp., 2013). Fosforylacja seryny 358 (numeracja aminokwasów dla białka z tytoniu dzikiego) w obrębie domeny SGS jest odpowiedzialna za tę funkcję.

2.2.2 Budowa i funkcja białka RAR1

Jak już wspomniano, białko RAR1 jest zakonserwowanym białkiem, występującym u wszystkich eukariotów z wyjątkiem drożdży (Shirasu i wsp., 1999). Składa się z dwóch domen: CHORD I (ang. cysteine and histidine rich domain) oraz CHORD II. Obie domeny wykazują duże podobieństwo sekwencyjne do siebie. Domena CHORD posiada około 60 aminokwasów a całe białko ma masę około 25 kDa. Struktura domeny CHORD II białka RAR1 oraz domeny CHORD I białka Chp1 przedstawiona jest na rysunku 2.6. Każda domena składa się z dwóch motywów wiążących cynk. Wiązanie cynku pełni rolę strukturalną i umożliwia przyjęcie przez domenę odpowiedniej konformacji (Heise i wsp., 2007, Zhang i wsp., 2010a). W obrębie obu domen aminokwasy koordynujące cynk: sześć cystein i trzy histydyny są zakonserwowane u wszystkich białek z rodziny RAR1. U roślin dodatkowo pomiędzy domenami CHORD wyodrębniony jest motyw CCCH, który koordynuje jeszcze jeden atom cynku. Jego rola biologiczna jest jak dotąd nie poznana. U ssaków i innych eukariotów np. nicienia *Caenorhabditis elegans* homologi białka RAR1 posiadają dodatkowo na C-końcu domenę CS tak jak białko SGT1 (Brancaccio i wsp., 2003). Jednak stopień wzajemnego pokrycia sekwencji domeny CS z białek Chp1 i meluzyny w stosunku do białka SGT1 jest niski.



Rysunek 2.6 Porównanie struktury przestrzennej domeny CHORD II białka RAR1 z *Arabidopsis thaliana* (Zhang i wsp., 2010a, pdb: 2XCM) oraz domeny CHORD I ludzkiego białka Chp1 (kod pdb: 2YRT).

Białko meluzyna oprócz domeny CS posiada jeszcze dodatkowo potencjalny motyw wiążący wapń na C-końcu (Brancaccio i wsp., 2003). U roślin wydaje się, że białko RAR1 bierze udział jedynie w procesach obrony przeciwko patogenom. Mutant *Arabidopsis thaliana* - *rar1* z nieaktywnym genem kodującym białko RAR1 nie wykazuje żadnego widocznego fenotypu, który wskazywał by na jeszcze inne funkcje jakie mogłoby pełnić to białko (Muskett i wsp., 2002). Białka zawierające domeny CHORD są odpowiedzialne za oddziaływania typu białko-białko. Dla białka RAR1 u roślin stwierdzono oddziaływania z białkiem HSP90 jak również z białkiem SGT1 (Takahashi i wsp., 2003). Obie domeny CHORD są zdolne do wiązania białka HSP90 w obrębie domeny N-terminalnej. Natomiast domena CHORD II dodatkowo oddziałuje z domeną CS białka SGT1. Białka CHP1 oraz meluzyna również oddziałują z białkami HSP90 i SGT1 (Sbroggiò i wsp., 2008, Wu i wsp., 2005). Jednak w oddziaływaniu zaangażowana jest także domena CS. Nie jest wiadomo do końca czy wiązanie białek CHP1 i meluzyny przebiega w sposób podobny jak w przypadku białek RAR1 i HSP90 u roślin, gdzie znana jest struktura przestrzenna rdzenia kompleksu białek HSP90 (domena N-terminalna), SGT1 (domena CS) oraz RAR1 (domena CHORD II).

2.3 Funkcja, struktura i dynamika białek z rodziny HSP90

2.3.1 Funkcja białek HSP90

Białko HSP90 należy do klasy białek zwanych białkami opiekuńczymi (Söti i wsp., 2005). Jak wskazuje nazwa, białka z rodziny HSP90 „opiekują” się innymi białkami pomagając im w poprawnym fałdowaniu się (Taipale i wsp., 2010). Jednak w odróżnieniu od innych białek opiekuńczych, które biorą udział w fałdowaniu się łańcucha polipeptydowego w trakcie lub niezwłocznie po translacji, białka HSP90 są raczej wymagane na późniejszych etapach tego procesu. W normalnych warunkach fizjologicznych białko HSP90 stanowi 1-2 % wszystkich białek w komórce (Borkovich i wsp., 1989). W czasie stresów ten stosunek może jeszcze wzrosnąć. Dlatego też białka HSP90 są kluczowe dla funkcjonowania komórki i zaangażowane są w wiele procesów komórkowych. Z niektórych ekspery-

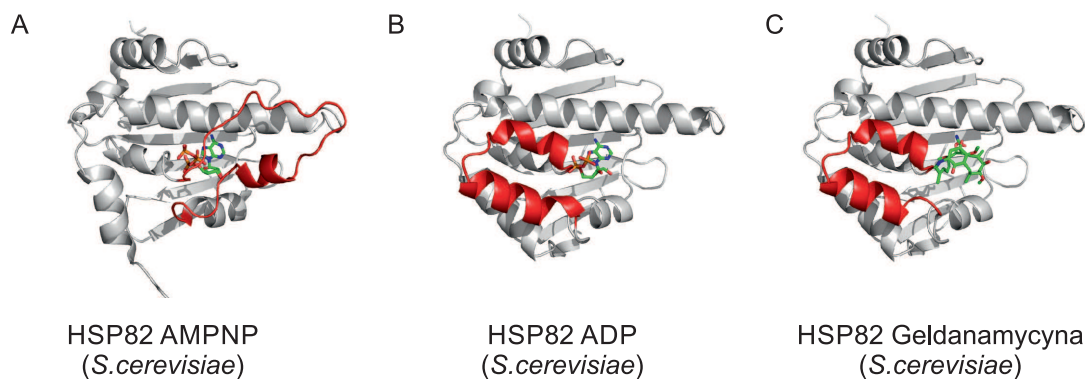
mentów wynika, że ok. 20 % wszystkich białek w drożdżach piekarskich zależy do białka HSP90 (<http://thebiogrid.org/35923/summary/saccharomyces-cerevisiae/hsp82.html>). Białko HSP90 kontroluje wiele procesów i oddziałuje z wieloma białkami, które mogą modulować jego funkcje. Na stronie internetowej <http://www.picard.ch/downloads> znajduje się aktualna lista białek oddziałujących z białkiem HSP90 oraz lista klientów białka HSP90. Do procesów, w które zaangażowane jest białko HSP90 należą między innymi kontrakcja mięśni i ruch komórkowych poprzez „opiekowanie się” białkiem miozyny (motora molekularnego odpowiedzialnego za kontrakcję mięśni), wrodzona odpowiedź immunologiczna: poprzez stabilizację receptorów NB-LRR u ludzi i roślin, odbudowa telomerów (fragmentów na końcach chromosomów odpowiedzialnych za skończoną liczbę podziałów komórki) poprzez stabilizację podjednostki odwrotnej transkryptazy kompleksu telomerazy, odpowiedzi na szok cieplny poprzez oddziaływanie z głównym czynnikiem transkrypcyjnym odpowiedzialnym za transkrypcję genów związanych ze szokiem cieplnym – HSF1 (ang. heat shock factor 1), infekcję wirusową poprzez stabilizację np. proteazy NS3 z wirusa zapalenia wątroby typu C, przekazywanie sygnałów wewnątrz komórkowych: poprzez stabilizację kinaz białkowych takich jak: ERBB2/HER2 (ang. erythroblastosis oncogene B/human epidermal growth factor receptor 2) czy CDK4 (ang. cell division control protein 4) oraz onkogennych mutantów kinaz: BRAF (ang. v-Raf murine sarcoma viral oncogene homolog B) (występuje w ok. 80 % przypadków czerniaka), BCR-ABL (ang. breakpoint cluster region - abelson) (występuje w 95 % przewlekłych białaczek szpikowych), EGFR (ang. epidermal growth factor receptor) i v-Src (ang. viral sarcoma) (Taipale i wsp., 2010). W związku ze stabilizacją zmutowanych kinaz białkowych oraz telomerazy HSP90 pełni istotną rolę w procesie onkogenezy (Whitesell i wsp., 2005). W ostatnich latach zaprojektowano wiele specyficznych inhibitorów białka HSP90, które mogą w przyszłości okazać się lekiem na raka. Jednym z nich jest geldanamycyna wyizolowana z bakterii *Streptomyces hygroscopicus* oraz jej syntetyczne pochodne (Franke i wsp., 2013). Ponieważ HSP90 pomaga w fałdowaniu się/stabilizacji dużej grupy kinaz, białko HSP90 wydaje się być interesującym celem w terapii wielu a nie jednego typu nowotworów.

Z mechanistycznego punktu widzenia proces fałdowania białka przez białko HSP90 można opisać w postaci cyklu (Röhl i wsp., 2013). W każdym z etapów cyklu biorą udział ko-chaperony białka, które będą modulowały aktywność białka HSP90. Przed związaniem z substratem białko HSP90 występuje w konformacji otwartej, w której domeny N-terminalne nie oddziałują za sobą. W pierwszym etapie kompleks białek HSP70 (ang. heat shock protein 70 kDa) i HOP (ang. HSP70/HSP90 organizing protein) ze związanym substratem oraz ATP wiążą się do białka HSP90. Białko HOP hamuje aktywność ATPazy białka HSP90. Poprzez domenę TPR, HOP wiąże się z C-terminalnym fragmentem MEEVD białka HSP90. Zatrzymana hydroliza ATP wydaje się być niezbędna do wiązania się kolejnego białka do kompleksu - białka z rodziny izomeraz petydylo-proliny (enzym katalizujący reakcję isomeryzacji cis-trans wiązania peptydowego zawierającego prolinę). W kolejnym etapie następuje oddysocjowanie białek HSP70, HOP a przyłączenie białka AHA1 (ang. activator of HSP90 ATPase). Białko AHA1 remodeluje konformację białka HSP90 tak aby możliwa była hydroliza ATP. Samo białko AHA1 stymuluje hydrolizę ATP najsilniej ze wszystkich poznanych ko-chaperonów. Po oddysocjowaniu białka AHA1 przyłącza się białko p23, które jest inhibitorem ATPazy HSP90. Białko p23 utrzymuje konformację HSP90, która jest zdolna do hydrolizy. Uważa się, że w tym czasie następuje proces dojrzewania związanego substratu, który po hydrolizie ATP wraz z pozostałymi białkami

kompleksu jest uwalniany. Nie jest jednak do końca wiadomo jak białka SGT1 i RAR1 mogą wpływać na cykl fałdowania białek przez białko HSP90.

2.3.2 Struktura białka HSP90

Białko HSP90 tworzy zarówno *in-vitro* jak i *in-vivo* dimery. Wykazano również, że w wyższych temperaturach białko HSP90 tworzy wyższe formy oligomeryczne (Nemoto i wsp., 1998, Chadli i wsp., 1999). Każda podjednostka dimeru HSP90 składa się z trzech domen: N-terminalnej domeny ATPazy (NTD), środkowej domeny (MD) oraz odpowiedzialnej za dimeryzację domeny C-końcowej (CD) (Pearl i wsp., 2000, Taipale i wsp., 2010, Li i wsp., 2012, Li i wsp., 2013a). N-terminalna domena o aktywności ATPazy należy do rodziny GHKL (ang. Gyrase, HSP90, histidine kinase and MutL). Rodzina ATPaz GHKL charakteryzują się specyficzną budową domeny ATPazy która posiada strukturę typu kanapki α/β (Dutta i wsp., 2000). Centralną część domeny zajmuje antyrównoległy arkusz beta składający się z co najmniej czterech wstęg beta otoczony przez trzy helisy alfa. Charakterystyczną cechą tej domeny jest fragment pokrywy (ang. ATP lid), który zamyka się na związanym nukleotydzie ATP. Ze związanym ADP lub przy braku nukleotydów fragment ten posiada otwartą konformację i nie oddziałuje z ADP lub inną częścią domeny ATPazy. Struktura N-terminalnej domeny białka HSP82 z drożdży związana z różnymi związkami przedstawiona jest na rysunku 2.7. W przypadku HSP90, związane ATP przyjmuje szczególną, nietypową, zgiętą konformację odmiennie niż w przypadku innych enzymów wykorzystujących ATP do reakcji enzymatycznych (Prodromou i wsp., 1997). Jest to szczególnie ważne, gdyż umożliwia to zaprojektowanie leków, które wiążą się specyficznym do kieszeni wiążącej ATP białek z rodziny HSP90. Jak już wspomniano, do takich inhibitorów należy geldanamycyna oraz jej pochodne (Franke i wsp., 2013). Blokują one kieszeń wiążącą ATP ale nie zmieniają konformacji segmentu pokrywy.



Rysunek 2.7 Struktura domeny N-terminalnej białka HSP82 z drożdży ze związanymi ligandami. **A.** Struktura domeny N-terminalnej w kompleksie z AMPNP (Ali i wsp., 2006, kod pdb: 2CG9). **B.** Struktura domeny N-terminalnej w kompleksie z ADP (Prodromou i wsp., 1997, kod pdb: 1AM1). **C.** Struktura domeny N-terminalnej w kompleksie z geldanamycyną (Prodromou i wsp., 1997, kod pdb: 1AH4). Fragment pokrywy zaznaczony jest kolorem czerwonym a związana cząsteczka kolorem

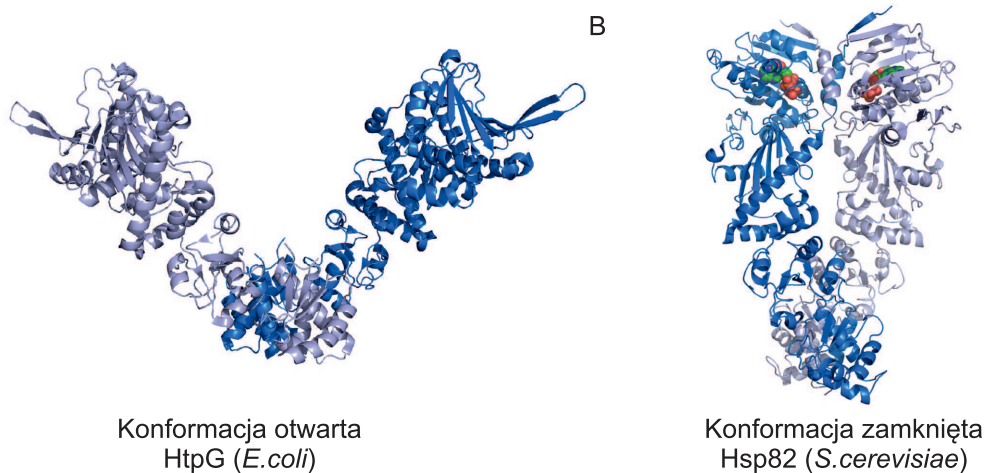
zielonym (atomy węgla). Wykonano na podstawie (Krukenber i wsp., 2011).

Domena NTD po związaniu ATP tworzy dimer z domeną NTD drugiej podjednostki białka HSP90 (Prodromou i wsp., 1997). W strukturze krystalicznej białka HSP90 z drożdży w kompleksie z nie-hydrolizowalnym analogiem ATP (AMP-PNP) i białkiem Sba1 (ang. increased sensitivity to benzoquinone ansamycins), dimeryzacja domeny N-terminalnej zachodzi poprzez krótki fragment beta wstęgi (aminokwasy 1-9), która oddziałuje z krawędzią arkusza beta drugiej domeny N-terminalnej licznymi wiązaniami wodorowymi oraz pierwszą helisę alfa (aminokwasy 13-22) (Ali i wsp., 2006). Po związaniu z AMPNP następuje również przemieszczenie się fragmentu pokrywy, która zakrywa kieszeń wiążącą nukleotydy. Dimeryzacja domeny C-terminalnej zachodzi poprzez utworzenie przez dwie ostatnie helisy w obu podjednostkach HSP90 wiązki czterech helis. C-terminalna domena zawiera na samym końcu sześćoaminokwasową sekwencję MEEVD. Do tej sekwencji wiążą się ko-chaperony białka HSP90 poprzez ich domenę TPR. Domena C-terminalna wiąże się także z antybiotykiem nowobiocyną (produktem naturalnym wyizolowanym z bakterii *Streptomyces niveus*) oraz jej pochodnymi oraz z cisplatyną (szeroko stosowanym lekiem przeciwnowotworowym) (Marcu i wsp., 2000, Itoh i wsp., 1999). Domena C-terminalna ma także dodatkowe miejsce wiązania ATP jednak jego funkcja jest jeszcze niewyjaśniona (Söti i wsp., 2002). Końcowy fragment, zawierający negatywnie naładowane aminokwasy bierze też udział w wiązaniu substratów przez białko HSP90 np. domeny wiążącej DNA białka p53 (Hagn i wsp., 2011).

Środkowa domena białka HSP90 składa się z dwóch mniejszych fragmentów: pierwszy, większy fragment jest domeną typu $\alpha - \beta - \alpha$. Połączona jest ona poprzez, krótki alfa helikalny fragment z mniejszym fragmentem również o strukturze typu $\alpha - \beta - \alpha$. Domena środkowa jest uwikłana w wiązanie substratów przez białko HSP90 (Hagn i wsp., 2011, Street i wsp., 2011). Negatywnie naładowane aminokwasy w obrębie domeny środkowej są odpowiedzialne za wiązanie się z domeną wiążącą DNA białka p53 (Hagn i wsp., 2011). Środkowa domena jest również odpowiedzialna za wiązanie się także z fragmentem $\Delta 131\Delta$ nukleazy z bakterii *Staphylococcus* (Street i wsp., 2011). Jest to modelowe białko do badania oddziaływania substratu z białkiem HSP90. Wiązanie fragmentu $\Delta 131\Delta$ powoduje zmiany konformacyjne polegające na zbliżeniu do siebie obu podjednostek białka HSP90. Również w przypadku wiązania domeny wiążącej ligand receptora glikosteroidu zaangażowana jest domena środkowa wraz z domeną N-terminalną (Lorenz i wsp., 2014). Dzięki naładowanym aminokwasom domena środkowa neutralizuje ładunek na powierzchni substratu zapobiegając jego agregacji. Środkowa oraz N-terminalna domena są także odpowiedzialne za wiązanie białka Tau (Karagöz i wsp., 2014). Białko Tau jest białkiem natywnie nieuporządkowanym biorącym udział w transporcie żelaza oraz pełniącym rolę w determinowaniu struktury filamentów mikrotubularnych w komórkach nerwowych. Nadmierna fosforylacja białka Tau prowadzi do jego nadmiernej agregacji, co może być jedną z przyczyn powstawania choroby Alzheimera (Ballatore i wsp., 2007). Białko HSP90 oddziałuje z fragmentem białka Tau, który jest najbardziej podatny na agregację i najbardziej hydrofobowy w całym białku Tau. O ile białka natywnie nieuporządkowane nie są substratami dla chaperonów, wydają się, że to hydrofobowy charakter powtórzeń decyduje o zdolności oddziaływania białka Tau z białkiem HSP90.

2.3.3 Dynamika konformacyjna białka HSP90

Z badań krystalograficznych wynika, że białko HSP90 w zależności od związanego nukleotydu lub jego braku może przyjmować wiele konformacji (Krukenberg i wsp., 2011). Na rysunku 2.8 przedstawione są struktury bakteryjnego białka HtpG z *E.coli* (konformacja otwarta) oraz drożdżowego białka HSP82 w kompleksie z AMPPNP i białkiem Sba1 (konformacja zamknięta). W konformacji zamkniętej domeny N-terminalne oddziałują ze sobą (Ali i wsp., 2006). Podobną konformację przyjmuje ludzkie mitochondrialne białko HSP90-TRAP1 w kompleksie z analogami ATP (Lavery i wsp., 2014). Jednak w odróżnieniu od struktury drożdżowego białka HSP90, wydaje się, że obie podjednostki mają inną konformację świadczącą o złożonym procesie hydrolizy ATP przez białko HSP90. Jest to zgodne z wynikami eksperymentów, w których wykazano, że jedna cząsteczka białka AHA1 wystarczy do aktywacji dimeru HSP90 (Retzlaff i wsp., 2010). Izoforma białka HSP90 z retikulum endoplazmatycznego, białko GRP94 w kompleksie z AMPPNP przyjmują konformację zamkniętą (Dollins i wsp., 2007). Jednak w odróżnieniu od drożdżowego białka HSP90 N-terminalne domeny nie oddziałują ze sobą i aby przyjąć konformację zdolną do hydrolizy ATP musi dojść do obrotu N-terminalnej domeny o 90° oraz obrotu domeny środkowej względem C-terminalnej. W przypadku struktury białka HtpG z *E.coli* (analoga HSP90) otrzymanej za pomocą krystalografii rentgenowskiej, białko ma konformację otwartą, w której domeny N-terminalne nie oddziałują ze sobą a całe białko ma kształt litery V (Shiau i wsp., 2006).



Rysunek 2.8 Struktura białka HSP90. **A.** Struktura krystalograficzna białka HtpG z *E.coli* bez związanych nukleotydów (Shiau i wsp., 2006, kod pdb: 2IOQ). **B.** Struktura przestrzenna białka HSP82 z drożdży w kompleksie z AMPPNP i Sba1 (Ali i wsp., 2006, kod pdb: 2CG9). Na rysunku widoczny jest tylko dimer białka HSP82.

Jeszcze inną konformację ma białko HtpG w kompleksie z ADP (Shiau i wsp., 2006). Białko HtpG ma konformację podobną do litery U a domeny N-terminalne nie oddziałują ze sobą. W eksperymentach mikroskopii elektronowej zaobserwowano alternatywną konformację białka HtpG w kompleksie z ADP. Konformacja ta różni się zasadniczo od tej zaobserwowanej w kryształach. W

tej konformacji domena N-terminalna jest obrócona o kąt 90° w stosunku do pozycji tej domeny w strukturze krystalograficznej tak, że cała podjednostka ma kształt litery L. Z drugiej strony, w kompleksie z ADP, białko GRP94 ma taką samą konformację jak w kompleksie z AMPPNP (Dollins i wsp., 2007).

W roztworze białko HSP90 występuje w równowadze konformacyjnej pomiędzy konformacją zamkniętą. Z eksperymentów SAXS wynika, że białko HtpG ma konformację otwartą, nieco inną niż tą zaobserwowaną w kryształach, w której kąt utworzony przez podjednostki w dimerze jest dużo większy (Krukenberg i wsp., 2008). Jednocześnie w kompleksie z AMPPNP, białko HtpG w roztworze występuje w równowadze, pomiędzy otwartą konformacją (zaobserwowaną w roztworze) a konformacją zamkniętą analogiczną jak dla białka HSP90 w kompleksie z AMPPNP i białkiem Sba1. Co ciekawe w przypadku homologów HSP90 z drożdży i człowieka, równowaga jest zależna od gatunku (Southworth i wsp., 2008). Dla ludzkiego białka HSP90 α , niezależnie od związanego nukleotydu, równowaga ta jest mocno przesunięta w stronę konformacji otwartej. W przypadku drożdżowego białka wiązanie AMPPNP, powoduje przesunięcie równowagi w stronę konformacji zamkniętej, podobnie jak dla białka HtpG. Natomiast kompaktowa konformacja ze związanym ADP występuje dla drożdżowego i ludzkiego homologa na bardzo niskim poziomie.

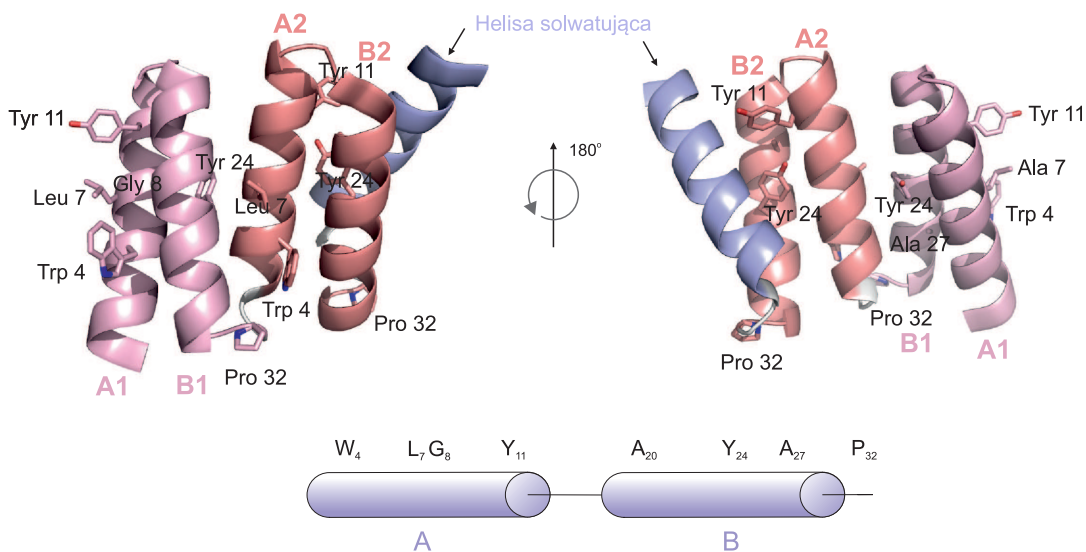
Na równowagę konformacyjną mają wpływ czynniki zewnętrzne takie jak pH czy też stężenie osmotów (Krukenberg i wsp., 2009a, Street i wsp., 2010). Związanie klientów, również powoduje przyjęcie konformacji zamkniętej (Street i wsp., 2011). Równowaga konformacyjna, odzwierciedla małe różnice energetyczne pomiędzy kolejnymi punktami w cyklu konformacyjnym białka HSP90. Pozwala to na kontrolowanie poszczególnych etapów cyklu za pomocą ko-chaperonów by zapewnić możliwie jak najbardziej efektywne działanie kompleksu białka HSP90. Podsumowując, niezależnie od gatunku białko HSP90 w roztworze występuje w równowadze pomiędzy konformacjami, a sama równowaga zależy od gatunku jak również środowiska. Różnice mogą wynikać z przede wszystkim z różnych warunków życia wymienionych wyżej organizmów i szybkości, z jaką muszą reagować na zmiany środowiska. Drożdże i bakterie *E. coli* muszą szybciej reagować na zmienne środowisko. Wydaje się także uzasadnione stwierdzenie, że różnice w skomplikowaniu proteomu pomiędzy prostym organizmem jak *E. coli* a skomplikowanym ludzkim organizmem mogą dyktować różne zachowanie się białka HSP90. W toku ewolucji białko HSP90, stało się bardziej białkiem regulatorowym wpływającym na metastabilną konformację białek-klientów niż białkiem odpowiedzi na stres.

2.4 Struktura i funkcja domen z rodziny TPR

2.4.1 Struktura domeny TPR

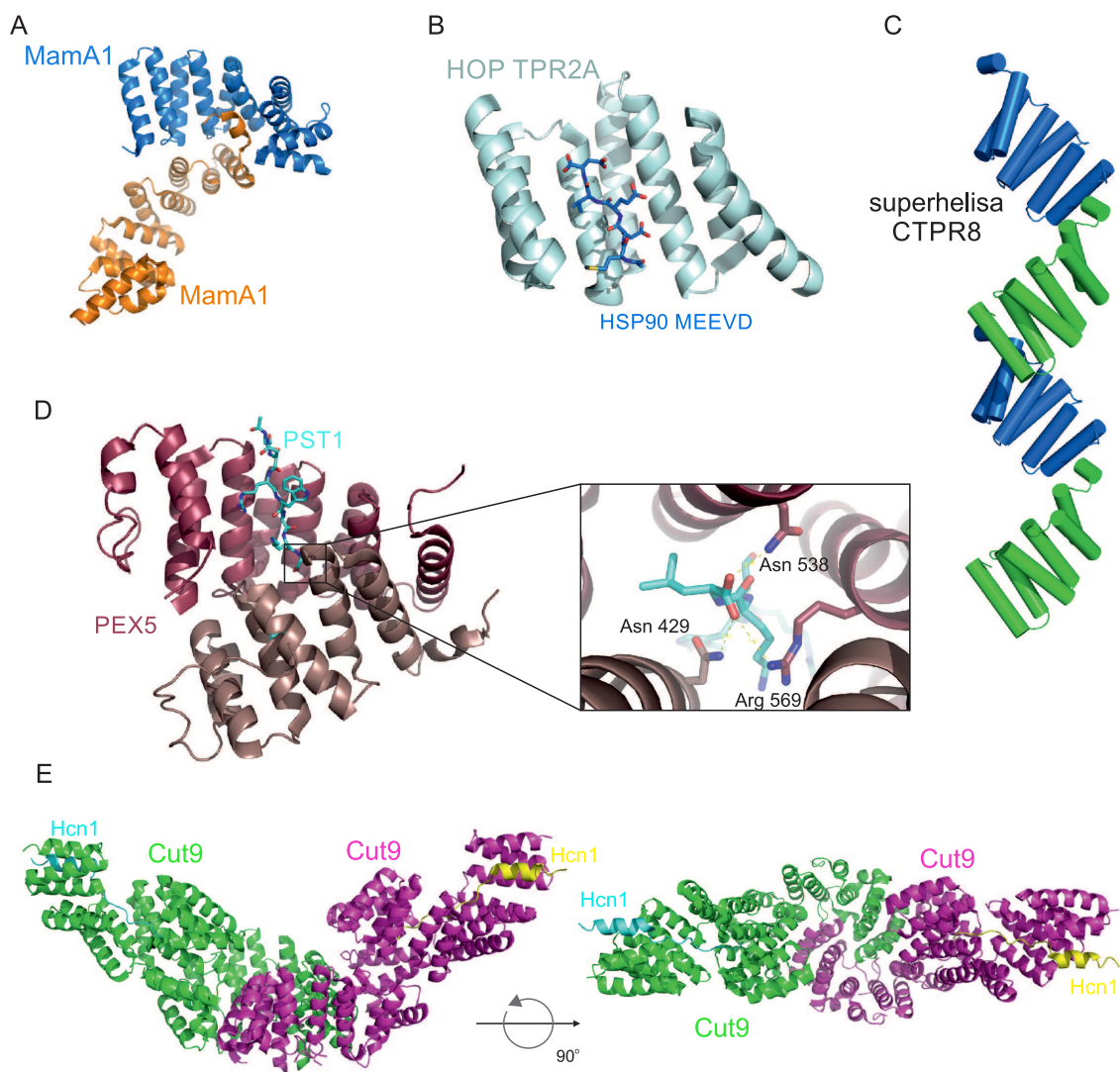
Powtórzenia typu TPR (ang. tetratricopeptide repeats) należą do klasy domen białkowych charakteryzujących się modułowym występowaniem, podobieństwem sekwencyjnym oraz strukturalnym i pośredniczącym w oddziaływaniach typu białko-białko. Powtórzenia typu TPR charakteryzują się 34 aminokwasową długością pojedynczego segmentu i występują w liczbie 3-16 segmentów (D'Andrea i wsp., 2003, Zeytuni i wsp., 2012). Każde powtórzenie składa się z dwóch helis: A i B przedzielonych skrętem. Powtórzenia TPR charakteryzują się ograniczonym podobieństwem sekwencyjnym. Jedynie sekwencja kilku aminokwasów jest zakonserwowana w 100 % a dodatkowo kilka pozycji w sekwencji wykazuje tendencję do zakonserwowania charakteru fizykochemicznego aminokwasu. W

obrębie pojedynczego segmentu TPR aminokwasy 4, 7, 8, 11, 20, 24, 27, 32 wykazują wspomniane zakonserwowanie. W pozycji 8 występuje alanina lub glicyna a w 20 i 27 alanina i są to najbardziej zakonserwowane aminokwasy w sekwencji pojedynczego powtórzenia TPR. W pozycjach 4, 7, 11 oraz 24 najczęściej występują duże hydrofobowe aminokwasy, a w pozycji 32 (w obrębie skrętu łączącego dwa powtórzenia) często znajdują się aminokwas destabilizujący strukturę helisy alfa np. prolina. Na rysunku 2.9 przedstawiona jest struktura białka CTPR2 składająca się z dwóch powtórzeń TPR o konsensusowej sekwencji oraz helisy solwatuującej (Main i wsp., 2003).



Rysunek 2.9 Struktura powtórzenia TPR. Struktura białka CTPR2 (Main i wsp., 2003, kod pdb: 1NA3) składającego się z dwóch powtórzeń TPR oraz helisy solwatuującej. Na schemacie i strukturze zaznaczone są miejsca występowania najbardziej zakonserwowanych aminokwasów. Rysunek wykonano na podstawie (D'Andrea i wsp., 2003).

Dwie helisy A i B w obrębie pojedynczego segmentu TPR tworzą motyw typu helisa-zwrot-helisa. Kąt pomiędzy tymi dwoma helisami wynosi około 24° . Takie wzajemne ułożenie helis powoduje, że każde kolejne powtórzenie TPR jest obrócone względem poprzedniego również o 24° , tworząc prawoskrętną superhelisę (Das i wsp., 1998, Kajander i wsp., 2007). W strukturze przestrzennej domeny TPR składającej się z kilku powtórzeń można zidentyfikować dwie powierzchnie zdolne do tworzenia oddziaływań z innymi makromolekułami. Jedna z nich jest utworzona z helis A pochodzących od kilku powtórzeń TPR i stanowi wklęsłą stronę domeny TPR. Druga powierzchnia znajduje się po przeciwnej stronie i posiada przeciwną, wypukłą topologię, utworzona jest zarówno przez helisy A i B z kilku powtórzeń TPR. Obecnie w bazie danych Protein Data Bank PDB znajduje się ponad 100 zdeponowanych struktur zawierających powtórzenia TPR. Charakterystyczną cechą wielu z nich okazało się występowanie dodatkowej helisy na C-końcu domen TPR, która jak się uważa ma za zadanie stabilizować całą strukturę i zwiększać rozpuszczalność całego białka.



Rysunek 2.10 Przykłady struktur białek składających się z powtórzeń TPR. **A.** Dimeryzacja białka MamA (Zeytuni i wsp., 2011, kod pdb: 3AS5) **B.** Wiązanie peptydu MEEVD białka HSP90 przez białko domenę TPR2A białka HOP (Scheufler i wsp., 2000, kod pdb: 1ELR) **C.** Superhelisa utworzona przez białko CTPR8 (Kajander i wsp., 2007, kod pdb: 2HYZ) **D.** Struktura kompleksu białka PEX5 z peptydem PST1. Na mniejszym rysunku przedstawiony jest mechanizm kłamy wiążącej grupę karboksylową peptydu PST1 (Sampathkumar i wsp., 2008, pdb: 3CVP) **E.** Struktura dimeru białka Cut9 w kompleksie z fragmentem białka Hcn1 (Zhang i wsp., 2010b, pdb: 2XPI).

Wraz z zidentyfikowaniem coraz większej liczby białek zawierających powtórzenia TPR oraz poznanych struktur przestrzennych wielu domen TPR stało się możliwe zaprojektowanie konsensusowego powtórzenia TPR, które mogłoby zostać użyte jako rusztowanie do projektowania specyficznych struktur białkowych i modulowania oddziaływań pomiędzy nimi. Pierwsze zaprojektowane w ten sposób białko nazwano CTPR (ang. consensus TPR domain) (Main i wsp., 2003). Zaprojektowane białko CTPR3 ma dużo większą stabilność termiczna niż białko PP5. Z badań konformacyjnych tych białek wykonanych metodą spektroskopii dichroizmu kołowego wynika, że temperatura topnienia dla domeny TPR białka PP5 wynosi 47°C w porównaniu do 83°C dla białka CTPR3. Kajander wraz ze współpracownikami (2007) skonstruowali białka zawierające 8 (CTPR8) oraz 20 (CTPR20) powtórzeń TPR bazujących na konsensusowej sekwencji zaproponowanej przez Main i wsp., (2003) oraz określili ich strukturę przestrzenną w kryształach za pomocą dyfrakcji promieniowania rentgenowskiego. Białka CTPR8 i CTPR20 posiadają strukturę superhelisy.

2.4.2 Funkcje białek zawierających powtórzenia TPR

Jak już było wspomniane wyżej domeny białkowe złożone z kilku powtórzeń aminokwasowych biorą udział przede wszystkim w oddziaływaniach białko-białko (Zeytuni i wsp., 2012). Do najbardziej znanych przykładów należą białka ko-chaperony wiążące się z C-terminalną domeną białek z rodziny białek szoku cieplnego: HSP70 i HSP90. Wiążą one zakonserwowany motyw aminokwasowy: MEEVD w przypadku białka HSP90 i IEEVD w przypadku białka HSP70. Przykładem białka wiążącego oba białka HSP jest białko HOP (ang. HSP organizing protein) (Scheufler i wsp., 2000). Białko HOP zawiera trzy odrębne domeny TPR: TPR1, TPR2A i TPR2B. Domena TPR1 wiąże białko HSP70, a domena TPR2A białko HSP90. Charakterystyczną cechą dla obu domen jest występowanie zakonserwowanych aminokwasów tworzących rodzaj klamry na grupach karboksylowych ostatniego kwasu asparaginowego w motywie EEVD i oddziałujących poprzez oddziaływania elektrostatyczne. Do ko-chaperonów białka HSP90, zawierających powtórzenia TPR, należą również takie białka jak CHIP (ang. C terminus of HSP70 interacting protein) czy UNC-45 (ang. uncoordinated 45) (Tapiale i wsp., 2010, Murata i wsp., 2001). Ko-chaperony z rodziny UCS (ang. UNC-45/CRO1/She4p proteins) biorą udział w poprawnym fałdowaniu się domeny ATPazy motoru molekularnego miozyny (Hutagalung i wsp., 2002). Białka TOM70 (ang. translocase of outer membrane 70 kDa) oraz TOC64 (ang. translocon at the outer envelope of chloroplast 64 kDa) wchodzi w skład kompleksów biorących udział w transporcie białek do mitochondriów oraz chloroplastów. W proces ten zaangażowane są także białka z rodziny HSP90. Poprzez oddziaływania z domeną TPR białek translokonu, białka HSP90 kierują białko celowe (mające być skierowane do mitochondrium lub chloroplastu) do kanału transportującego białko przez zewnętrzną błonę organellum (Schleiff i wsp., 2011). Ważną rodziną białek charakteryzujących się posiadaniem domeny TPR i oddziałujące z białkiem HSP90 są immunofiliny- białka wiążące leki immunosupresyjne takie jak: cyklosporyna A-wiązana przez białko CyP40 (ang. cyclophilin D 40 kDa), czy FK506 wiązane przez białka FKBP51 i FKBP52 (ang. FK506 binding protein 51k Da/52 kDa) (Ratajczak i wsp., 2003). Immunofiliny funkcjonalnie należą do izomeraz peptydyloprolinowych czyli enzymów, które katalizują reakcje izomeryzacji cis-trans proliny w łańcuchu polipeptydowym.

Innym ciekawym przykładem wiązania ligandu przez domeny TPR jest białko PEX5 (ang. Pe-

roxin 5) (Stanley i wsp., 2006, Sampathkumar i wsp., 2008). Białko PEX5 jest zaangażowane w transport białek do peroksisomu poprzez wiązanie do peptydu sygnałnego PST1. Peptyd sygnałny PST1 zawiera zakonserwowany fragment trój-aminokwasowy o sekwencji SKL (lub mniej zakonserwowanej sekwencji SHL lub AKL). Peptyd PST1 jest wiązany w charakterystyczny sposób. C-końcowy łańcuch boczny tworzy wiązania hydrofobowe a C-terminalna grupa karboksylowa tworzy wiązania elektrostatyczne w podobny sposób jak w przypadku białka HOP i peptydów z białek HSP poprzez mechanizm klamry. Oprócz specyficznych oddziaływań z grupami bocznymi kilka reszt aminokwasowych asparaginy tworzy wiązanie z łańcuchem głównym peptydu PST1. Opisane powyżej oddziaływania zilustrowane zostały na rysunku 2.10.

2.4.3 Oligomeryzacja domen TPR

W przypadku kilku białek zawierających domeny TPR zaobserwowano ich oligomeryzację zachodzącą dzięki oddziaływaniom domen TPR. Pierwszym białkiem, dla którego zaobserwowano ten efekt było białko Cut9 (ang. cell untimely torn protein 9) z drożdży *Schizosaccharomyces pombe* wchodzące w skład kompleksu APC (ang. anaphase promoting complex) - kompleksu zaangażowanego w przejście komórki z metafazy do anafazy podczas podziału komórkowego (Zhang i wsp., 2010b). Kompleks APC jest ligazą ubikwityny E3, która odpowiedzialna jest za degradację białek zaangażowanych w cykl komórkowy (Peters 2006). Białko Cut9 praktycznie w całości składa się z 14 powtórzeń TPR i można je strukturalnie podzielić na dwie części, każdą po 7 powtórzeń. N-terminalna część odpowiedzialna jest za homodimeryzację, natomiast C-terminalna bierze udział w wiązaniu białka Hcn1 (ang. high copy suppressor of cut9 protein 1). Białko Cut9 tworzy strukturę typu „V” gdzie dimer domen N-terminalnych stanowi centrum kompleksu. Innym ciekawym białkiem, które ulega homo-oligomeryzacji jest białko MamA (ang. magnetosome protein A) występujące w bakteriach magnetotaktycznych takich jak: *Magnetospirillum magneticum* AMB-1 oraz *Magnetospirillum gryphiswaldense* MSR-1 (Matsunaga i wsp., 2003, Taoka i wsp., 2006). Bakterie magnetotaktyczne charakteryzują się występowaniem specjalnego organellum zwanego magnetosomem, składającego się z nanokrystalicznego tlenku żelaza Fe_3O_4 w formie magnetytu lub siarczku żelaza Fe_3S_4 w formie greigitu, otoczonego błoną komórkową z wbudowanymi białkami oraz innymi białkami wchodzącymi w skład macierzy białkowej (rusztowania) magnetosomu (Blakemore 1975, Matsunaga i wsp., 2003). Magnetosomy w bakterii ułożone są liniowo tworząc dipol magnetyczny pozwalający orientować się bakterii względem pola magnetycznego. Białko MamA jest 22-24 kDa białkiem składającym się z pięciu powtórzeń TPR oraz jednego domniemanego powtórzenia TPR na N-końcu. Pięć powtórzeń TPR tworzy prawoskrętną superhelisę zniekształconą w stosunku do superhelisy białka CTPR8 (Zeytuni i wsp., 2011). Struktury białek: Cut9, MamA oraz CTPR8 przedstawione są na rysunku 2.10.

2.4.4 Zastosowanie domen TPR w inżynierii białkowej

Powtórzenia TPR z racji swojej struktury przestrzennej, modularnej budowy, wiązania ligandów oraz stosunkowo łatwych metod ich ekspresji i oczyszczania stanowią doskonały obiekt dla inżynierii białkowej. Jednym z pierwszych zastosowań domen TPR w inżynierii białek było zaprojektowanie „inteligentnego” hydrożelu. Grupa prof. Lynne Regan z Yale University zaprojektowała hydrożel

składający się z trzech powtórzeń TPR wiążących peptyd o sekwencji DESVD (przeznaczonych domeną TPR również o długości trzech powtórzeń która nie wiązała ligandu (Grove i wsp., 2010, Grove i wsp., 2012). Tak powstałe białko składające się w sumie z 6 powtórzonych domen TPR (trzy wiążące i trzy nie wiążące, łącznie 18 powtórzeń motywu TPR) tworzyło hydrożel, przy stężeniu wagowo-objętościowym 1 %, po dodaniu modyfikowanego 4 ramiennego polimeru na bazie glikolu polietylenowego zawierającego peptydowy ligand dla domeny TPR w stosunku 1:2 powtórzeń wiążących peptyd do ligandu. Zaletą takiego podejścia było ograniczenie się jedynie do zmieszania dwóch składników żelu bez udziału żadnych modyfikacji chemicznych, pH, stanu redoks. Proces sieciowania zachodził spontanicznie w temperaturze pokojowej i nie był zakłócony poprzez dodanie serum z 7 % surowiczą albuminą wołu. Hydrożel TPR wykazywał dobre właściwości lepko-sprężyste typowe dla biologicznych materiałów. Zmierzony reometrycznie moduł magazynowania G' wyniósł 270 Pa co odpowiada właściwościom elastycznym 5 % żelu poliakrylamidowego oraz 15 % żelatynie. Siła wiązania peptydu do domeny TPR zależna jest od siły jonowej. Dzięki zmianie stężenia soli w roztworze można kontrolować erozję hydrożelu. Hydrożel TPR następnie wykorzystano do przygotowania nośnika leku – doksorubicyny oraz eksperymentalnego leku C9 (pochodna 7-azapterydyny), inhibitora białka HSP90. Tempo uwalniania leku jest zależne od stężenia żelu jak również siły jonowej. Wraz ze wzrostem siły jonowej, czas uwalniania leku staje się krótszy co odzwierciedla stabilność żelu w zależności od siły jonowej. Żel TPR skutecznie uwalnia C9 do hodowli komórkowej i zabija HER-2 pozytywne komórki raka piersi BT-474 (Grove i wsp., 2012).

2.5 Wybrane metody badania struktury przestrzennej makromolekuł - podstawy teoretyczne

Wszystkie organizmy żywe, począwszy od prostych bakterii i archeonów a kończąc na człowieku opierają swoje funkcjonowanie na skomplikowanych i ściśle ze sobą powiązanych procesach biochemicznych zachodzących w podstawowej jednostce organizacyjnej jaką jest komórka. Za początki życia na Ziemi uważa się pojawienie samoreplikujących się cząsteczek kwasu rybonukleinowego – tak jak głosi teoria świata RNA (ang. RNA world) (Alberts i wsp., 2002, Robertson i wsp., 2012). Również podstawowe reakcje biochemiczne najpierw miały być przeprowadzane przez cząsteczki RNA z czasem jednak pojawiły się bardziej wydajne i sprawniejsze makromolekuły składające się z aminokwasów – białka. Stopniowo wypierały one z pełnienia różnych procesów biologicznych cząsteczki RNA, choć w obecnych organizmach żywych wiele bardzo ważnych procesów biologicznych takich jak choćby synteza białka czy splicing – czyli składanie cząsteczek informacyjnego RNA jest nadal katalizowane przez RNA (Alberts i wsp., 2002). Również jako nośnik informacji genetycznej cząsteczki RNA zostały zastąpione przez cząsteczki kwasu deoksyrybonukleinowego. O życiu decydują więc trzy rodzaje makromolekuł: kwas rybonukleinowy – RNA, deoksyrybonukleinowy - DNA oraz białka. Poznanie ich funkcji ma więc fundamentalne znaczenie na zrozumienie podstawowych procesów zachodzących we wszystkich organizmach żywych. Wymienione wyżej cząsteczki posiadają strukturę trójwymiarową, która definiuje ich funkcje. Poznanie ich struktury oraz dynamiki jest więc bardzo ważne nie tylko z punktu widzenia nauk podstawowych, ale również z powodu możliwości wykorzystania ich w wielu dziedzinach jak np. medycyna, bioinżynieria czy inżynieria materiałowa.

Na początku XX wieku naukowcy wciąż byli podzieleni co do poglądu na temat atomowej struktury materii. Wyjaśnienie na gruncie nowej teorii kwantów efektu fotoelektrycznego przez Alberta Einsteina czy promieniowania ciała doskonale czarnego przez Plancka potwierdziły korpuskularną naturę światła. Dopiero kilka lat po tych odkryciach, doświadczenia wykonane przez uczniów Arnolda Sommerfelda - Paula Knippinga i Waltera Friedricha a zaproponowane przez Maxa von-Lauego potwierdziły, że kryształy składają się z periodycznie ułożonych atomów lub cząsteczek (Friedrich i wsp., 1913). Max von Laue posłużył się odkrytym w 1895 roku przez Roentgena promieniowaniem X. Równocześnie Sir William Henry Bragg wraz z synem Sir Williamem Lawrencem sformułowali słynne prawo dyfrakcji promieniowania na periodycznie ułożonych płaszczyznach nazwane od nazwiska ich odkrywców prawem Bragga (Bragg 1913):

$$n\lambda = 2d \sin(\theta) \quad (2.1)$$

gdzie: n – rząd ugięcia, liczba całkowita ograniczona od góry zależnością $\sin(\theta) \leq 1$, λ – długość fali promieniowania, d – odległość pomiędzy płaszczyznami w sieci krystalicznej.

Wszystkie te odkrycia pozwoliły na poznanie najpierw struktury przestrzennej prostych związków nieorganicznych takich jak NaCl, KCl, KI czy też niektórych metali. Następnie określono strukturę bardziej złożonych związków nieorganicznych takich jak $MgAl_2O_4$ czy $CaCO_3$. Pierwszym związkiem organicznym, którego strukturę przestrzenną określono za pomocą dyfrakcji promieniowania rentgenowskiego była urotropina. Jej strukturę rozszyfrowano w 1923 roku (Dickinson i wsp., 1923). Dorothy Crowfoot Hodgkin pracowała nad strukturą ważnych cząsteczek biologicznych takich jak

cholesterol, penicylina czy witamina B12 (Carlisle i wsp., 1945, Crowfoot i wsp., 1949, Hodgkin i wsp., 1955). Pierwszym białkiem, którego struktura została rozwiązana była mioglobina z mięśni kaszalota otrzymana w 1958 roku przez Sir Johnego Kendrew (Kendrew i wsp., 1958). W 1960 roku Max Perutz rozwiązał strukturę cząsteczki hemoglobina z krwi konia (Perutz i wsp., 1960). Obaj byli pracownikami laboratorium Medical Research Council przy laboratorium Cavendisha na Uniwersytecie w Cambridge w Wielkiej Brytanii. Za swoje odkrycia zostali uhonorowani Nagrodą Nobla w dziedzinie chemii w 1962 roku. W tym samym roku Maurice Wilkins, Francis Crick i James Watson otrzymali Nagrodę Nobla w dziedzinie medycyny za odkrycie struktury cząsteczki kwasu deoksyrybonukleinowego (Watson i wsp., 1953). Wilkins, Crick i Watson posłużyli się również dyfrakcją promieniowania rentgenowskiego, jednakże nie na pojedynczych kryształach a na włóknach w celu rozszyfrowania struktury podwójnej helisy DNA. Z tego samego laboratorium co Perutz, Kendrew, Crick i Watson wywodzi się również jeden z twórców metody transmisyjnej mikroskopii elektronicznej, laureat Nagrody Nobla z dziedziny chemii w 1982 roku Aaron Klug. Klug jako pierwszy ustalił strukturę przestrzenną wirusa mozaiki tytoniu, był jednym z twórców krystalografii elektronicznej oraz jako pierwszy określił strukturę nukleosomu (Klug i wsp., 1960, Richmond i wsp., 1984). Większość ze wspomnianych wyżej naukowców była z wykształcenia fizykami, a ich osiągnięcia wskazują na istotne powiązanie fizyki i biologii nie tylko jeśli chodzi o opracowanie metod badawczych ale również na unikatowe spojrzenie dzięki któremu udało im się rozwiązać jedne z najważniejszych pytań biologii molekularnej.

W poniższych rozdziałach zostaną opisane podstawy teoretyczne stosowanych przeze mnie metod eksperymentalnych ze szczególnym uwzględnieniem metod obliczeniowych stosowanych w analizie wyników.

2.5.1 Oddziaływanie promieniowania z materią

Promieniowanie rentgenowskie (nazywane też promieniowaniem X) jest promieniowaniem elektromagnetycznym o długości fali λ od 100 Å (miękkie promieniowanie rentgenowskie) do 0.1 Å (twarde promieniowanie rentgenowskie) i mieści się w zakresie pomiędzy promieniowaniem UV a promieniowaniem γ , powstającym podczas reakcji jądrowych. Tak jak każde promieniowanie elektromagnetyczne, fotony promieniowania rentgenowskiego poruszają się z prędkością 299 792 km/s w próżni, a związek pomiędzy energią wyrażoną w keV a długością fali w Å definiuje następujący wzór:

$$\lambda \left[\text{Å} \right] = \frac{hc}{E} = \frac{12.398}{E[\text{keV}]} \quad (2.2)$$

gdzie: h – stała Plancka $6.62606957(29) \cdot 10^{-34}$ J·s, c – prędkość światła w próżni, E – energia fotonu.

Charakterystyczną cechą promieniowania rentgenowskiego jest współczynnik załamania równy 1 z dokładnością do 10^{-5} dla znanych nam materiałów. Jest to jednoznaczne z tym, iż nie jesteśmy w stanie zbudować odpowiednich soczewek w celu skupiania lub rozpraszania promieniowania rentgenowskiego, tak jak jest to możliwe dla widzialnego spektrum promieniowania elektromagnetycznego czy też elektronów i neutronów. Nie istnieje więc mikroskop mogący powiększać obiekty o rozmiarach atomowych za pomocą promieniowania rentgenowskiego. Nie mniej jednak naukowcy wynaleźli

szereg innych metod badawczych za pomocą, których jesteśmy w stanie zobaczyć materię w skali atomowej.

Promieniowanie rentgenowskie może oddziaływać z materią w dwojaki sposób: poprzez rozpraszanie elastyczne (bez zmian długości fali) i nieelastyczne (następuje zmiana długości fali) oraz może zostać zaabsorbowane przez atom, wzbudzając go do wyższego stanu energetycznego. Po absorpcji, w wyniku powrotu do stanu podstawowego następuje emisja promieniowania, charakterystyczna dla danego atomu i energii zaabsorbowanego fotonu. Do badania struktury przestrzennej biomolekuł znalazło zastosowanie głównie rozpraszanie promieniowania rentgenowskiego. Również w swoich badaniach posługiwałem się głównie promieniowaniem w zakresie promieniowania rentgena, dlatego w dalszej części pracy skupię się na opisie teoretycznym tego procesu.

2.5.2 Rozpraszanie promieniowania rentgenowskiego przez atom

Ponieważ długość promieniowania rentgenowskiego jest porównywalna z rozmiarami atomu wiązka promieniowania będzie rozpraszana przez elektrony atomu i ulegała interferencji. Na rysunku 2.11 pokazany jest schemat układu odniesienia. Oznaczenie w tekście odnoszą się do elementów na rysunku 2.11.

Fala padająca jest falą płaską o wektorze falowym o długości $k_p = 2\pi/\lambda$. Fala rozproszona pod kątem 2θ scharakteryzowana jest przez wektor falowy o długości $k_r = 2\pi/\lambda$. Różnica dróg dwóch fal rozproszonych przez elementy atomu: jeden w początku układu współrzędnych O oraz drugi w punkcie r_j , jest równa sumie dwóch odcinków AO oraz OB. Różnica dróg pomnożona przez stosunek $2\pi/\lambda$ jest równa różnicy faz. Z rysunku 2.11 widać, że tak zdefiniowana różnica faz między dwoma rozproszonymi falami jest równa iloczynowi skalarnemu:

$$\Delta\phi(r_j) = \vec{r}_j \cdot (\vec{k}_r - \vec{k}_p) = |OA| + |OB| \quad (2.3)$$

Różnica faz powoduje, że fale rozproszone przez elektrony, atomu będą ulegały interferencji. Amplitudę rozpraszania przez N centrów rozpraszania możemy zapisać w postaci następującego wzoru:

$$\vec{A} = \sum_{i=1}^N \vec{A}_j \exp(i\vec{r}_j \cdot \vec{s}_j) \quad (2.4)$$

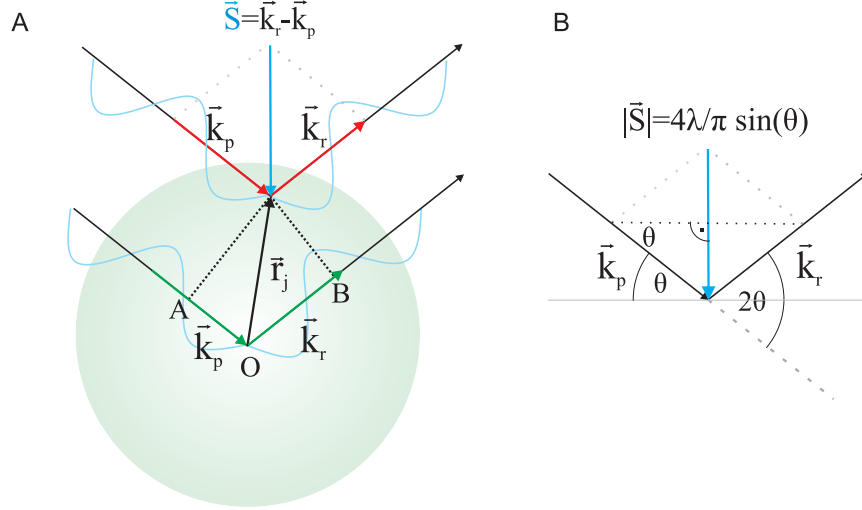
gdzie: \vec{A}_j jest amplitudą rozpraszania przez j-te centrum rozpraszania, $\vec{s}_j = \vec{k}_r - \vec{k}_p$ jest wektorem rozpraszania o długości $|\vec{s}_j| = \frac{4\pi}{\lambda} \sin(\theta)$, 2θ - jest kątem rozpraszania.

Aby otrzymać wzór na fale rozproszone przez dany obiekt należy rozwiązać równanie falowe (Landau i Lifszyc 1985, Feigin i Svergun 1987):

$$[\Delta + k_p^2 + \nu\phi(\vec{r})] \psi(\vec{r}) = 0 \quad (2.5)$$

gdzie: Δ - operator Laplace'a, $k_p = 2\pi/\lambda$ wektor falowy fali padającej, ν - jest parametrem opisującym siłę oddziaływania fali padającej z polem skalarnym $\phi(\vec{r})$. Pole $\phi(\vec{r})$ w zależności od natury fali może być gęstością elektronową, gdy rozpraszaniu ulega promieniowanie rentgenowskie czy też gęstością jądrową gdy rozpraszaniu ulegają neutrony i opisuje ciągły rozkład centrów rozpraszania w przestrzeni. Rozwiązaniem powyższego równania jest superpozycja fal rozproszonych $\psi'(\vec{r})$ będących

falami sferycznymi oraz fali padającej $\psi^p(\vec{r})$, która nie uległa rozproszeniu, zgodnie z zasadą Huygensa–Fresnela, gdy badamy interferencje fal w odległości $\vec{r} \gg \vec{r}'$ oraz pominiemy ponowne ugięcie fali rozproszonej.



Rysunek 2.11 Schemat rozpraszania promieniowania rentgenowskiego przez atom.

Rozwiązanie przedstawione jest w postaci wzoru:

$$\psi^p(\vec{r}) + \psi^r(\vec{r}) = A_p \exp(i\vec{k}_p \vec{r}) + \frac{A_p \nu}{4\pi} \frac{\exp(i\vec{k}_r \vec{r})}{r} \int \phi(\vec{r}') \exp(i\vec{s} \vec{r}') d\vec{r}' \quad (2.6)$$

W powyższym równaniu wyrażenie:

$$f(\vec{s}) = \frac{\nu}{4\pi} \int \phi(\vec{r}') \exp(i\vec{s} \vec{r}') d\vec{r}' \quad (2.7)$$

opisuje amplitudę rozpraszania przez pole skalare $\phi(\vec{r})$ i nazywane jest czynnikiem rozpraszania. Ponadto amplituda ta jest transformatą Fouriera pola skalarnego w związku z tym, znając $f(\vec{s})$ jesteśmy w stanie otrzymać $\phi(\vec{r})$ poprzez obliczenie odwrotnej transformaty Fouriera:

$$\phi(\vec{r}) = \frac{1}{2\pi^2 \nu} \int f(\vec{s}) \exp(i\vec{s} \vec{r}) d\vec{s} \quad (2.8)$$

W teorii znając dokładnie amplitudę rozpraszania tj. funkcję $f(\vec{s})$ jesteśmy w stanie otrzymać pole skalare $\phi(\vec{r})$ czyli kształt, strukturę obiektu rozpraszającego. Niestety eksperymentalnie możemy jedynie zarejestrować liczbę rozpraszanych cząstek (fotonów, neutronów, elektronów itp.) lub przepływ energii, które są proporcjonalne do kwadratu amplitudy rozpraszania:

$$\frac{dW}{d\Omega} = \frac{A_p^2}{r^2} |f(\vec{s})|^2 = \frac{A_p^2}{r^2} I(\vec{s}) \quad (2.9)$$

gdzie: $I(\vec{s})$ jest intensywnością rozpraszania, Ω – jest kątem bryłowym. Znacząc jedynie eksperymentalnie wyznaczoną funkcję $I(\vec{s})$, a więc kwadrat modułu zespolonej funkcji nie jesteśmy w stanie

otrzymać fazy α , potrzebnej do rekonstrukcji pola za pomocą odwrotnej transformaty Fouriera tak jak w równaniu (2.8). Problem ten w krystalografii znany jest pod terminem problemu fazowego.

W przypadku rozpraszania przez atom funkcja $f(\vec{s})$ nazywana jest atomowym czynnikiem rozpraszania i można ją zapisać w postaci następującego wzoru:

$$f(\vec{s}) = \int \rho_a(\vec{r}) \exp(i\vec{r}\vec{s}) d\vec{r} \quad (2.10)$$

gdzie: $\rho_a(\vec{r})$ jest gęstością elektronową atomu w punkcie \vec{r} . Intensywność rozpraszania otrzymamy poprzez przemnożenie powyższego wyrażenia przez amplitudę rozpraszania przez elektron.

Przy obliczaniu atomowych czynników rozpraszania np. w rentgenowskiej analizie strukturalnej możemy rozdzielić wyrażenie na czynnik struktury dla atomu o liczbie atomowej Z , na wkład od poszczególnych elektronów:

$$f(\vec{s}) = \sum_{j=1}^Z \int \rho_e(\vec{r}_j) \exp(i\vec{r}_j\vec{s}) d\vec{r}_j \quad (2.11)$$

gdzie: $\rho_e(\vec{r}_j) = e|\psi_j(\vec{r})|^2$ jest gęstością elektronową określoną przez prawdopodobieństwo znalezienia elektronu j opisanego za pomocą funkcji falowej $\psi_j(\vec{r})$.

Natomiast dla cząsteczki składającej się z N atomów wyrażenie na czynnik rozpraszania przyjmie następującą postać:

$$F_M(\vec{s}) = \sum_{j=1}^N f_a^j(\vec{s}) \exp(i\vec{r}_j\vec{s}) \quad (2.12)$$

2.6 Małokątowe rozpraszanie promieniowania rentgenowskiego

Znając podstawy fizyczne procesu rozpraszania można przejść do rozważania zagadnienia małokątowego rozpraszania promieniowania rentgenowskiego (ang. SAXS – small angle X-ray scattering) przez układ składający się z identycznych nie oddziałujących ze sobą cząstek rozpuszczonych w matrycy jaką jest np. rozpuszczalnik w przypadku układów koloidalnych. Małokątowe rozpraszanie promieniowania rentgenowskiego jest techniką za pomocą, której jesteśmy w stanie badać strukturę materii w zakresie wielkości od około 1 nm do kilkudziesięciu a nawet kilkuset nanometrów ponieważ długość fali promieniowania jakie jest używane w typowym eksperymencie SAXS mieści się w zakresie 0,1-0,2 nm. Jako pierwsi technikę SAXS wykorzystali Guinier i Fournet (Guinier i Fournet 1955) do badania stopów metali. Wykazali oni, że rozpraszanie pod małym kątem θ blisko wiązki pierwotnej niesie informację na temat kształtu i wielkości cząsteczki rozpraszającej ale także o różnego rodzaju fluktuacjach struktury wewnętrznej badanych materiałów. Na początku analiza danych otrzymanych z eksperymentów SAXS ograniczała się do analizy prostych parametrów mówiących jedynie o wielkości cząstek. Wraz z rozwojem nowych technik eksperymentalnych i postępach w teoretycznym opisie rozpraszania przez cząsteczki o określonym kształcie i strukturze wewnętrznej sprawiły, że z eksperymentów SAXS można było wyciągnąć bardzo dużo informacji na temat struktury wewnętrznej badanych obiektów. Krokiem milowym w badaniach z wykorzystaniem promieniowania rentgenowskiego a także rozpraszania neutronów było wykorzystanie techniki zwanej wariacją kon-

trastu. Technika ta polegała na pomiarach krzywych rozpraszania w kilku różnych rozpuszczalnikach (np. mieszanina H₂O-D₂O), którego długość rozpraszania była taka sama jak dla części badanych cząsteczek. W ten sposób zmieniając kontrast rozpuszczalnika można było selektywnie „wygasić” rozpraszanie tych cząsteczek, które miały taki sam kontrast jak rozpuszczalnik. Pierwszym sukcesem tej techniki było potwierdzenie teorii Florego zachowania się łańcucha polimerowego w stopach polimerów. Innym osiągnięciem było zmapowanie położenia RNA i białek w bakteryjnym rybosomie przed opublikowaniem struktur krystalograficznych (Svergun i wsp., 2000). Na początku lat 90-tych pojawiły się pierwsze metody *ab-initio* oraz metody pozwalające na otrzymanie niskorozdzielczych modeli makromolekuł. Pierwsza metoda polegała na zaprezentowaniu kształtu cząsteczki za pomocą funkcji $F(\omega)$ zdefiniowaną następująco:

$$\rho(\vec{r}) = \begin{cases} 1 & 0 \leq \vec{r} \leq F(\omega) \\ 0 & \vec{r} \geq F(\omega) \end{cases} \quad (2.13)$$

Jest to więc obwiednia kształtu cząsteczki. Metoda ta zakładała rozkład funkcji za pomocą harmonijek sferycznych:

$$F(\omega) = \sum_{l=0}^L \sum_{m=-l}^l f_{lm} Y_{lm}(\omega) \quad (2.14)$$

gdzie: f_{lm} są współczynnikami multipolowymi zdefiniowanymi poprzez relację: $f_{lm} = \int_{\omega} F(\omega) Y_{lm}^*(\omega) d\omega$, funkcje $Y_{lm}(\omega)$ są harmonijkami sferycznymi. Współczynniki f_{lm} są następnie iteracyjnie obliczane, tak aby minimalizowały funkcję:

$$R_I^2(I_{exp}(s_i), f_{lm}) = \frac{\sum_{i=1}^N (I_{exp}(s_i) - I_{calc}(s_i))^2 W(s_i)}{\sum_{i=1}^N I_{exp}(s_i)^2 W(s_i)} \quad (2.15)$$

gdzie: $I_{exp}(s_i)$ jest eksperymentalną funkcją rozpraszania o N punktach o współrzędnych s_i , $W(s_i)$ jest funkcją wagi np. może być to odchylenie standardowe. Za pomocą współczynników f_{lm} obliczana jest funkcja rozpraszania dla modelu $I_{calc}(s_i)$. Metoda ta została zaimplementowana w programie SASHA i była historycznie pierwszą metodą, która pozwalała na uzyskanie modeli niskorozdzielczych białek na zasadzie nieliniowej optymalizacji, a nie metodą prób i błędów (Svergun i wsp., 1996). Następnie nastąpił rozwój modelowania oparty na modelu kulkowym gdzie cząsteczka reprezentowana jest przez zbiór pseudo-atomów. Pierwsze programy: DALAI (grupa Pablo Chacóna z Instytutu Chemii Fizycznej CSIC w Madrycie) oraz DAMMIN (grupa Dimitra Sverguna z EMBL w Hamburgu) zostały oparte odpowiednio na algorytmach genetycznym i Monte Carlo (Chacón i wsp., 1998, Svergun 1999). Innym podejściem było wykorzystanie dostępnych modeli atomowych do modelowania cząsteczek za pomocą modelowania bryły sztywnej poprzez obroty i translacje kilku fragmentów cząsteczki (np. domen czy podjednostek) bądź kompleksu. Metody te zostaną dokładnie opisane w następnych podrozdziałach. Jednocześnie postęp technologiczny w dziedzinie wytwarzania promieniowania synchrotronowego (wprowadzenie synchrotronów trzeciej generacji) pozwolił na skrócenie czasu pomiaru i korzystanie z mniejszej ilości próbki. Również detektory promieniowania rentgenowskiego na przestrzeni lat stały się coraz szybsze i dokładniejsze, co pozwoliło na pomiary czasowo-rozdzielcze. Przykładowo, najnowsze półprzewodnikowe detektory firmy Dectris mają moż-

liwości rejestracji pojedynczych fotonów oraz posiadają dynamiczny zakres rejestracji $1 - 10^6$ fotonów (nawet 10^3 fotonów różnicy pomiędzy sąsiednimi pikselami). Funkcja przenoszenia (ang. point spread function) dla detektorów Pilatus wynosi jeden piksel. Również ilość próbki potrzeba do przeprowadzenia pojedynczego pomiaru zmniejszyła się tak, że niekiedy wystarczy 5-10 μl próbki białka o stężeniu 1 mg/ml. Jednocześnie na wielu liniach pomiarowych wprowadzono automatyczne stacje pipetujące, pozwalające na całkowitą automatyzację pomiarów a nawet na zdalne ich przeprowadzanie za pośrednictwem internetu z innego miejsca na świecie.

2.6.1 Odwrotna transformata Fouriera $I(s)$ oraz krzywa rozkładu odległości wewnątrzcząsteczkowych

Rozpraszanie promieniowania rentgenowskiego przez cząsteczki o gęstości elektronowej $\rho_{cz}(\vec{r})$ zawieszona w buforze o gęstości elektronowej $\rho_b = \langle \rho_b(\vec{r}) \rangle_V$ (przyjmujemy, że gęstość rozpraszania buforu w skali długości próbkowanej w eksperymencie SAXS jest stała) scharakteryzowane jest przez transformatę Fouriera gęstości rozpraszania $\Delta\rho(\vec{r}) = \rho_{cz}(\vec{r}) - \rho_b$. Wtedy po odjęciu od krzywej rozpraszania dla roztworu białka rozpraszania przez bufor, amplitudę fali rozproszonej możemy zapisać w następującej postaci, analogicznie do równania na funkcję $\phi(\vec{r})$:

$$A = \int_V \Delta\rho(\vec{r}) \exp(i\vec{s}\vec{r}) d\vec{r} \quad (2.16)$$

Mając już wyrażenie na amplitudę rozpraszania możemy zapisać wzór na intensywność rozpraszania przez cząsteczki, uśrednioną po wszystkich jej orientacjach względem wektora \vec{s} , to jest po wszystkich możliwych wartościach $\vec{s}\vec{r}$:

$$I_{cz}(s) = \langle I_{cz}(\vec{s}) \rangle_\Omega = \left\langle \int_V \int_V \Delta\rho(\vec{r}) \Delta\rho(\vec{r}') \exp(i\vec{s}(\vec{r} - \vec{r}')) d\vec{r} d\vec{r}' \right\rangle_\Omega \quad (2.17)$$

wprowadzając nową zmienną $\vec{u} = \vec{r} - \vec{r}'$, uśredniając $\exp(i\vec{s}(\vec{r} - \vec{r}'))$ i całkując we współrzędnych sferycznych po współrzędnych kątowych otrzymamy wyrażenie:

$$I_{cz}(s) = \int_0^{D_{max}} u^2 \gamma(u) \frac{\sin(su)}{(su)} du \quad (2.18)$$

gdzie:

$$\gamma(u) = \left\langle \int \Delta\rho(\vec{u} + \vec{r}) \Delta\rho(\vec{r}) d\vec{r} \right\rangle_\Omega \quad (2.19)$$

jest uśrednioną funkcją autokorelacji a D_{max} jest maksymalną odległością wewnątrzcząsteczkową. Wygodniej jest zmienić zmienną u na r . Gdzie r oznacza teraz odległość między dwoma dowolnymi punktami w cząsteczce. Na podstawie $\gamma(r)$ można zdefiniować funkcję $P(r) = r^2 \gamma(r)$, która jest rozkładem odległości wewnątrzcząsteczkowych. Funkcję $P(r)$ można obliczyć za pomocą odwrotnej transformaty Fouriera:

$$P(r) = \frac{r^2}{2\pi^2} \int s^2 I_{cz}(s) \frac{\sin(sr)}{sr} ds \quad (2.20)$$

Niestety w praktyce funkcję tą nie można obliczyć za pomocą odwrotnej transformaty Fouriera

ze względu na ucięcie wartości wektora s dla których mierzone są dane eksperymentalne.

Do obliczania funkcji $P(r)$ stosuje się metodę IFT (ang. indirect Fourier transform) polegającą na rozkładzie funkcji $P(r)$ na funkcje składowe:

$$P(r) = \sum_i c_i \phi_i(r) \quad (2.21)$$

gdzie $\phi_i(r)$ są funkcjami (krzywymi) B-sklejanymi (ang. B-spline functions). Eksperymentalna intensywność rozpraszania może być zapisana jako liniowa transformata funkcji $P(r)$, gdzie T_1 jest transformatą Fouriera a kolejne $T_2 - T_4$ mogą być związane np. z geometrią wiązki padającej czy rodzajem aparatury pomiarowej:

$$I_{cz}(s) = T_4 T_3 T_2 T_1 p(r) \quad (2.22)$$

także podstawiając do powyższego równania rozkład funkcji $P(r)$ otrzymamy w rezultacie:

$$I_{cz}(s) = \sum_i c_i T_4 T_3 T_2 T_1 \phi_i(r) = \sum_i c_i \chi_i(s) \quad (2.23)$$

problem polega na znalezieniu współczynników c_i poprzez minimalizację funkcjonału L wraz z dodatkowym warunkiem $N - L + \lambda N_e = \min$, gdzie:

$$L = \frac{\sum_i (I_{exp}(s_i) - I_{cz}(s_i))^2}{\sigma^2(s_i)} = \frac{\sum_i (I_{exp}(s_i) - \sum_i c_i \chi_i(s))^2}{\sigma^2(s_i)} \quad (2.24)$$

$$N_e = \sum_i (c_{i+1} - c_i)^2 \quad (2.25)$$

λ – jest tu czynnikiem stabilizującym (mnożnikiem Lagrange'a). Praktycznie, przed przystąpieniem do obliczania współczynników rozwinięcia, należy zdefiniować wartość N czyli ilość używanych funkcji (w praktyce 10-40), wartość D_{max} oraz λ . W celu znalezienia optymalnej wartości parametru λ używa się metody punktu przegięcia wykreślając wartości N_e i L względem różnych wartości λ . Optymalny parametr λ znajdziemy gdy jednocześnie N_e i L będą miały punkt przegięcia dla pewnego λ_{opt} . Błędne określenie D_{max} oraz N będzie skutkowało niemożliwością znalezienia właściwej wartości parametru λ_{opt} . Svergun w swojej pracy (Svergun 1992), dodatkowo wprowadził pewne kryteria bazujące na kształcie krzywej $P(r)$ w celu nałożenia dodatkowych więzów w procesie optymalizacji. Najpopularniejszym programem wykorzystującym metodę IFT wraz ze wspomnianymi więzami jest program GNOM autorstwa D. Sverguna oraz A. Semyenuka, który pozwala na otrzymanie krzywej $P(r)$ (Svergun 1992). Funkcja $P(r)$ jest rozkładem odległości wewnątrzcząsteczkowych i rozkład ten jest charakterystyczny dla danego kształtu. Dla niektórych obiektów geometrycznych znane są analityczne wyrażenia opisujące rozkład $P(r)$. Analiza kształtu krzywej $P(r)$ dostarcza szybkiej informacji na temat przybliżonego kształtu badanej cząsteczki.

2.6.2 Asymptotyczne zachowanie krzywej $I(s)$

2.6.2.1 Wykres Guiniera, promień żyrcji

Jeżeli krzywą rozpraszania $I(s)$ będziemy rozważać w granicy $\lim_{s \rightarrow 0} I(s)$ to możemy rozwinąć funkcję $\sin(sr)/sr \approx 1 - (sr)^2/3! + \dots$ w wyrażeniu 2.18 na intensywność rozpraszania. Otrzymamy wtedy następujący wzór:

$$I(s) = I(0) \left[1 - \frac{1}{3} R_g^2 s^2 + O(s^4) \right] \simeq I(0) \exp \left(-\frac{1}{3} R_g^2 s^2 \right) \quad (2.26)$$

gdzie:

$$R_g = \frac{\int r^2 P(r) dr}{2 \int P(r) dr} \quad (2.27)$$

jest elektronowym promieniem żyrcji opisujący średnią odległość punktów od środka masy (rolę masy odgrywa tutaj gęstość elektronowa), czyli zarazem opisuje też wielkość cząsteczki. Zmiana wartości R_g ma raczej charakter jakościowy i może świadczyć o zmianach konformacyjnych, oligomeryzacji itp. Wykres $\ln(I(s))$ vs. s^2 nazywany jest wykresem Guiniera i powinien być linią prostą, której współczynnik kierunkowy proporcjonalny jest do promienia żyrcji. Jakikolwiek odstępstwa od liniowego przebiegu wykresu Guiniera świadczą o tym, że próbka jest nie homogenna pod względem wielkości cząsteczek rozpraszających lub też występują oddziaływania repulsyjne pomiędzy cząsteczkami. Zarówno z wykresu jak również za pomocą funkcji $P(r)$ można obliczyć wartość R_g .

2.6.2.2 Prawo Poroda – asymptotyczne zachowanie się krzywej $I(s)$ w granicy $s \rightarrow \infty$

Na zachowanie krzywej $I(s)$ w granicy $s \rightarrow \infty$ największy wpływ ma struktura wewnętrzna cząsteczki a najbardziej, jak zostanie to udowodnione, lokalna struktura na powierzchni cząsteczki. Granica $s \rightarrow \infty$ odpowiada $r \rightarrow 0$ w przestrzeni rzeczywistej (lub też na krzywej $P(r)$). Możemy wtedy uśrednioną i unormowaną funkcję autokorelacji $\gamma_0(r) = \gamma(r)/\Delta\rho$ ($\Delta\rho$ średni kontrast gęstości elektronowej) rozwinąć w szereg:

$$\gamma_0(r) = 1 - ar + br^2 + cr^3 + \dots \quad (2.28)$$

gdzie $a = \frac{1}{l}$, jest średnią wartością przecięcia cząsteczki (przecięciem cząsteczki nazywamy odcinek przechodzący przez cząsteczkę łączący dwa punkty na jej powierzchni). Z definicji $\gamma_0(r)$ wynika, że dla bardzo małych r zerowy wkład do funkcji autokorelacji będzie pochodził od powierzchni cząsteczki. Dla jednostki powierzchni dS będzie on wynosił $rdS \cos(\nu)$, gdzie ν jest kątem pomiędzy normalną do elementu dS a wektorem przesunięcia \vec{r} . Jeżeli uśrednimy po wszystkich kierunkach wektora \vec{r} otrzymamy wtedy $rdS \frac{1}{2}$. Każdy element dS będzie dawał wkład tylko dla jednego zwrotu \vec{r} czyli poprzednie wyrażenie możemy pomnożyć przez kolejny czynnik $\frac{1}{2}$ otrzymując ostatecznie wzór na wspólną objętość oraz funkcję autokorelacji:

$$\langle V_c(\vec{r}) \rangle_\Omega = V - \frac{1}{4} Sr \quad (2.29)$$

$$\gamma_0(r) \simeq 1 - \frac{S}{4V}r \quad (2.30)$$

oraz wstawiając powyższe wyrażenie do wyrażenia na $I(s)$ otrzymamy:

$$I(s) \simeq 4\pi\Delta\rho V \int r^2 \left(1 - \frac{S}{4V}r + \dots\right) \frac{\sin(sr)}{sr} dr \quad (2.31)$$

całkując przez części oraz rozważając wynik w granicy $s \rightarrow \infty$ otrzymamy:

$$\lim_{s \rightarrow \infty} I(s) = \Delta\rho^2 \frac{2\pi}{s^4} S \quad (2.32)$$

Ostatni wzór zwany jest prawem Poroda (Porod 1951, Debye i wsp., 1957) i opisuje zachowanie krzywej rozpraszania $I(s)$ dla dużych kątów rozpraszania. Spadek intensywności rozpraszania będzie proporcjonalny do s^{-4} . Znając $\Delta\rho$ można otrzymać wartość S . Wygodniej jednak jest zdefiniować wartość Q – niezmiennik Poroda, posługując się twierdzeniem Parsewala:

$$Q = \int s^2 I(s) ds = 2\pi^2 \int (\Delta\rho(\vec{r}))^2 d\vec{r} \quad (2.33)$$

$$Q = 2\pi^2 \gamma_0(0) = 2\pi^2 \Delta\rho^2 V \quad (2.34)$$

biorąc pod uwagę, że $I(0) = \Delta\rho^2 V^2$ otrzymamy wyrażenie na objętość: $V = 2\pi^2 I(0) Q^{-1}$. Mając już wzór na objętość oraz graniczną wartość $\lim_{s \rightarrow 0} I(s)$ jesteśmy w stanie wyliczyć również S . Razem powyższe rozważania możemy podsumować za pomocą układu dwóch równań:

$$\begin{cases} \frac{\pi}{Q} \lim_{s \rightarrow \infty} s^4 I(s) = \frac{S}{V} \\ Q = 2\pi^2 I(0) V^{-1} \end{cases} \quad (2.35)$$

Asymptotyczne zachowanie krzywej $I(s)$ razem z niezmiennikiem Poroda Q pozwalają więc na oszacowanie objętości oraz powierzchni rozpraszających cząsteczek. Dla cząsteczek charakteryzujących się niejednorodną powierzchnią zależność $I(s) \sim Ss^{-4}$ nie jest już spełniona. Także, obserwując zachowanie się funkcji $I(s)$ dla $s \rightarrow \infty$ możemy nie tylko oszacować objętość i powierzchnie ale także stwierdzić jaki charakter ma powierzchnia cząsteczki.

2.6.3 Czynniki kształtu

Zajmując się rozpraszaniem od identycznych obiektów (cząsteczek) można założyć, że kontrast wewnątrz cząsteczki jest stały i wynosi $\Delta\rho$. Możemy zapisać wtedy wzór na rozpraszanie przez roztwór o stężeniu n (liczba cząsteczek o na jednostkę objętości) w następujący sposób:

$$I(s) = n (\Delta\rho)^2 V^2 |A(s)|^2 \quad (2.36)$$

$$|A(s)|^2 = P(s) \quad (2.37)$$

$P(s)$ jest tak zwanym czynnikiem kształtu (ang. form factor). Czynniki kształtu opisuje nam rozpraszanie przez obiekt o zadanym kształcie i jest znormalizowany do jedności tj. $P(0) = 1$. Nie

zależnie od kontrastu każda cząsteczka o danym kształcie będzie miała ten sam czynnik kształtu. Dla kuli o promieniu R można wyznaczyć analityczną formułę na czynnik kształtu:

$$P(s) = F_{sfera}^2 = \left(\frac{3 [\sin(sR) - sR \cos(sR)]}{(sR)^3} \right)^2 \quad (2.38)$$

dla cząsteczek o innych kształtach nie można otrzymać analitycznych wzorów opisujących czynnik kształtu. Dla wielu brył jednak udało się wyprowadzić pół-analityczne wyrażenia za pomocą, których stosunkowo łatwo obliczyć czynnik kształtu i dopasować parametry opisujące wielkość cząstek do danych eksperymentalnych w celu weryfikacji przyjętego kształtu cząsteczek rozpraszających. Poniżej zebrane są wyrażenia na czynnik kształtu dla wybranych rodzajów obiektów. Warto zauważyć, że dany rodzaj cząsteczek rozpraszających będzie miał charakterystyczny czynnik kształtu pozwalający nawet bez dopasowania modelu do danych eksperymentalnych na stwierdzenie jaki kształt ma cząsteczka w roztworze.

Dla sferycznej powłoki o promieniu zewnętrznym R_1 , promieniu wewnętrznym R_2 oraz ze zdefiniowanymi objętościami $V_1 = \frac{4}{3}\pi R_1^3$, $V_2 = \frac{4}{3}\pi R_2^3$ czynnik struktury wynosi:

$$P(s) = \left(\frac{V_1(R_1) F_{sfera}(R_1, s) - V_2(R_2) F_{sfera}(R_2, s)}{V_1(R_1) - V_2(R_2)} \right)^2 \quad (2.39)$$

Ponieważ sferyczną powłokę można opisać jako różnicę między dwoma kulami o promieniach R_1 i R_2 także czynnik kształtu jest podniesioną do kwadratu różnicą w rozpraszaniu przez dwie kule.

Czynnik struktury dla cząsteczki składającej się z N kulistych podjednostek pierwszy wyprowadził Debye w 1915 roku (Debye 1915). Równanie to jest jednym z ważniejszych w teorii małokątowego rozpraszania ponieważ w pewnym przybliżeniu każdy badany obiekt można przedstawić w postaci zbioru N kul.

$$P(s) = \frac{1}{M^2} \sum_{i=1}^M \sum_{j=1}^M F_{sfera}^i(R_i, s) \cdot F_{sfera}^j(R_j, s) \frac{\sin(sr_{ij})}{sr_{ij}} \quad (2.40)$$

gdzie: M jest sumą mas N kul i służy jako czynnik normalizujący, r_{ij} jest wektorem łączącym środek masy i -tej i j -tej sfery. Wzór ten stanowi podstawę do obliczania teoretycznych krzywych rozpraszania dla zbioru N -atomów.

Innymi ważnymi wzorami, które często są stosowane są wyrażenie na czynnik kształtu cylindra o promieniu podstawy R oraz długości L podany przez Fourneta w 1951 roku (Fournet 1951):

$$P_{cylinder}(s) = \int \left[\frac{2J_1(sR \sin \alpha)}{sR \sin \alpha} \frac{\sin(sL \cos(\alpha)/2)}{sL \cos(\alpha)/2} \right]^2 \sin \alpha d\alpha \quad (2.41)$$

Dla łańcucha polimerowego o gaussowskiej statystyce (kłębek statystyczny) czynnik struktury obliczył Debye (1947):

$$P_{polimer}(s) = 2 \frac{[\exp(-\langle R_g^2 \rangle s^2) + \langle R_g^2 \rangle s^2 - 1]}{(\langle R_g^2 \rangle s^2)^2} \quad (2.42)$$

gdzie: $\langle R_g^2 \rangle$ jest uśrednionym po zespole statystycznym promieniem żyrcji łańcucha polimerowego.

2.6.4 Metody *ab-initio* w modelowaniu niskorozdzielczych struktur białek i kwasów nukleinowych na podstawie danych SAXS

W poprzednich rozdziałach zostały opisane podstawy teoretyczne małokątowego rozpraszania promieniowania rentgenowskiego. Znając rozkład atomów modelu jesteśmy w stanie obliczyć krzywą rozpraszania i porównać ją z krzywą rozpraszania otrzymaną w eksperymencie. Każdą cząsteczkę rozpraszającą możemy opisać w przybliżeniu jako układ N kul. Im większa będzie liczba N tym z większą dokładnością będziemy mogli obliczyć krzywą rozpraszania. Intensywność rozpraszania można obliczyć za pomocą wzoru Debye'a. Jednakże metoda ta prosta w swoich założeniach ma dwie istotne wady. Po pierwsze szybkość obliczenia intensywności rozpraszania zależy $\sim N^2$ od ilości kul reprezentujących cząsteczkę. Obecnie moc obliczeniowa komputerów jest bardzo duża i na pierwszy rzut oka nie stanowi to już większego problemu. Jednak gdy obliczenia trzeba powtarzać wielokrotnie w ciągu pracy programu problemy z mocą obliczeniową mogą stanowić pewne utrudnienie. Drugą wadę stanowi fakt, że musimy najpierw zaproponować pewien model wykorzystując znane nam informacje (stopień oligomeryzacji, symetria lokalna) na temat badanej cząsteczki. Kiedy jednak nie dysponujemy żadnymi danymi na temat przybliżonego kształtu musimy losowo generować układ kul, obliczyć krzywą rozpraszania i porównać ją z danymi eksperymentalnymi. Załóżmy, że nasz obiekt składa się z N kul a dyskretna przestrzeń po której możemy przemieszczać kule składa się z M punktów. Wtedy ilość konfiguracji do sprawdzenia wynosi zgodnie z regułami kombinatoryki:

$$\binom{M}{N} = \frac{M!}{N!(M-N)!} \quad (2.43)$$

Jeśli za N podstawimy 100 natomiast za M 1000, wtedy dokładna liczba konfiguracji wyniesie: $6,385051192630514 \times 10^{139}$. Nawet dla M równego 200 czyli dla objętości obszaru przeszukiwania konfiguracji tylko dwa razy większego od objętości modelu liczba konfiguracji wynosi $9,054851465610328 \times 10^{58}$. Jeżeli nawet uda nam się nałożyć jakiegokolwiek więzy nadal liczba kombinacji do sprawdzenia będzie ogromna. Dlatego też niezbędne było znalezienie szybszych metod obliczania intensywności rozpraszania za pomocą wzoru Debye'a jak również wprowadzenie metod optymalizacyjnych w celu znalezienie odpowiedniej konfiguracji kul dla której krzywa rozpraszania jest jak najbardziej zgodna z krzywą eksperymentalną. Inną stroną zagadnienia stanowi znalezienie unikalnego modelu. Obliczanie krzywej rozpraszania oczywiście jest zawsze unikalne tzn. dla każdego obiektu istnieje tylko jedna krzywa rozpraszania. Następuje tutaj "redukcja" wymiarów przestrzeni z 3D do 1D. Natomiast modelowanie modelu cząsteczki na podstawie krzywej rozpraszania zwiększa nam wymiar przestrzeni z 1D do 3D. Dlatego też kilka różnych konfiguracji może mieć taką samą krzywą rozpraszania. Wprowadzenie dodatkowych więzów: np. dostępnych eksperymentalnie parametrów np. promień żyrcji czy maksymalny rozmiar cząsteczki oraz generalnych właściwości struktur makromolekuł tj. kompaktowy kształt, brak nieciągłości w strukturze itp., umożliwiłoby osiągnięcie większego stopnia unikalności otrzymanywanych modeli.

Pod koniec lat dziewięćdziesiątych pojawiły się pierwsze algorytmy i metody pozwalające na szybsze liczenie krzywej rozpraszania oraz modelowanie niskorozdzielczych struktur makromolekuł bez jakichkolwiek informacji *a priori*. W poniższych podrozdziałach zostaną opisane najważniejsze algorytmy i metody obliczania krzywej rozpraszania oraz modelowania *ab-initio* struktur makromo-

lekuł.

2.6.5 Obliczanie krzywej rozpraszania dla makrocząsteczek w roztworze

2.6.5.1 Metoda harmonijek sferycznych

W 1970 roku Stuhrmann (Stuhrmann 1970) zaproponował efektywną metodę obliczania krzywej rozpraszania dla układu N-atomów. Procedura ta została zaimplementowana w programie CRY SOL przez Sverguna (Svergun i wsp., 1995). Metoda zakłada rozwinięcie funkcji eksponencjalnej występującej w wyrażeniu na amplitudę rozpraszania za pomocą harmonijek sferycznych. Amplitudę oraz intensywność rozpraszania przez N-atomów możemy zapisać w postaci następujących wzorów:

$$A(\vec{s}) = \sum_{i=1}^N f_i(s) \exp(i\vec{s}\vec{r}_i) \quad (2.44)$$

$$I(\vec{s}) = \sum_{i=1}^N \sum_{j=1}^N f_i(s) f_j(s) \exp[i\vec{s}(\vec{r}_i - \vec{r}_j)] \quad (2.45)$$

Jeżeli teraz zapiszemy $\exp[i\vec{s}(\vec{r}_i - \vec{r}_j)]$ w postaci rozwinięcia otrzymamy następującą formułę na amplitudę rozpraszania:

$$A(\vec{s}) = \sum_{l=0}^{\infty} \sum_{m=-l}^l a_{lm}(s) Y_{lm}(\Omega) \quad (2.46)$$

gdzie:

$$a_{lm}(s) = 4\pi i^l \sum_{j=1}^N f_j(s) j_l(sr_j) Y_{lm}^*(\omega_j) \quad (2.47)$$

a $\vec{s} = (s, \Omega)$, $\vec{r}_j = (r_j, \omega_j)$, $j_l(sr_j)$ -są sferycznymi funkcjami Bessela l-ego rzędu, Y_{lm} są harmonijkami sferycznymi. Ponieważ harmonijki sferyczne są funkcjami ortogonalnymi tj.

$\int_{\varphi=0}^{2\pi} \int_{\theta=0}^{\pi} Y_{lm} Y_{l'm'}^* d\Omega = \delta_{mm'} \delta_{ll'}$ uśredniając wyrażenie na $I(\vec{s})$ tj. $\langle I(\vec{s}) \rangle_{\Omega} = I(s)$ człony dla różnych par (m,l) będą równe 0. Tak więc problem obliczenia funkcji rozpraszania za pomocą wzoru Debye'a można zredukować do złożoności O(N). Ostatecznie wzór na intensywność rozpraszania możemy zapisać w postaci:

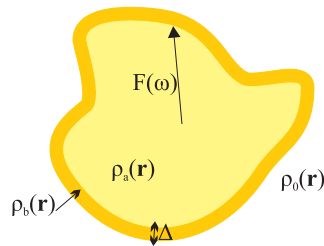
$$I(s) = \sum_{l=0}^{l_{max}} \sum_{m=-l}^l |a_{lm}(s)|^2 \quad (2.48)$$

Podobne wyrażenie lecz z uwzględnieniem warstwy hydratacyjnej wykorzystuje program CRY-SOL do obliczania rozpraszania przez cząsteczkę wraz z otaczającą ją sferą hydratacyjną:

$$I(s) = \langle |A_a(\vec{s}) - \rho_0 A_c(\vec{s}) + \delta\rho_b A_b(\vec{s})|^2 \rangle_{\Omega} = \quad (2.49)$$

$$= \sum_{l=0}^{l_{max}} \sum_{m=-l}^l |A_{lm}(s) - \rho_0 B_{lm}(s) + \delta\rho_b C_{lm}(s)|^2 \quad (2.50)$$

gdzie: $A_a(\vec{s})$ - jest amplitudą rozpraszania przez cząsteczkę w próżni, $\rho_0 A_c(\vec{s})$ - amplitudą rozpraszania przez objętość wyłączoną, ρ_0 gęstością elektronową rozpuszczalnika, $\delta\rho A_b(\vec{s})$ jest amplitudą rozpraszania przez warstwę hydratacyjną cząsteczki o różnicowej gęstości elektronowej $\delta\rho = \rho_b - \rho_0$, gdzie ρ_b jest gęstością elektronową warstwy hydratacyjnej. W równaniu 2.50 współczynniki $A_{lm}(s)$, $B_{lm}(s)$, $C_{lm}(s)$ odpowiadają współczynnikom a_{lm} w równaniu 2.48. Warstwa hydratacyjna modelowana jest w następujący sposób. Najpierw definiowana jest funkcja $F(\omega)$ (rysunek 2.14), która opisuje nam powierzchnię cząsteczki. Dla każdej wartości kąta ω , funkcja $F(\omega)$ definiuje nam promień r od środka masy cząsteczki.



Rysunek 2.14 Definicja funkcji $F(\omega)$. Na podstawie (Svergun i wsp., 1995).

Za pomocą funkcji $F(\omega)$ możemy teraz opisać gęstość elektronową warstwy hydratacyjnej ρ_b :

$$\rho_b = \begin{cases} 1 & F(\omega) \leq r \leq F(\omega) + \Delta \\ 0 & 0 < r < F(\omega) \text{ lub } r > F(\omega) + \Delta \end{cases} \quad (2.51)$$

Niestety metoda ta nie może opisać warstwy hydratacyjnej np. we wnękach ponieważ funkcja $F(\omega)$ opisuje jedynie zewnętrzną powierzchnię cząsteczki. Nie pozwala również różnicować lokalnej grubości warstwy hydratacyjnej.

2.6.5.2 Metoda wielomianów Zernike'a

Innym alternatywnym podejściem do obliczania krzywej rozpraszania wraz ze sferą hydratacyjną jest zastosowanie rozwinięcia funkcji eksponencjalnej za pomocą wielomianów Zernike'a (Liu i wsp., 2012a). Metoda ta została wykorzystana w programie SASTBX (Liu i wsp., 2012b). W odróżnieniu od metody harmonijek sferycznych metoda oparta na wielomianach Zernike'a nie tylko pozwala na obliczanie krzywej rozpraszania ze złożonością $O(N)$, ale pozwala także na obliczenie wkładu od sfery hydratacyjnej także dla wnęk w białkach np. kanałach chaperonów białkowych. Wielomiany Zernike'a są następującej postaci:

$$Z_{nlm} = R_{nl}(r)Y_{lm}(\omega) \quad (2.52)$$

gdzie:

$$R_{nl}(r) = \sum_{k=0}^{(n-l)/2} N_{nlk} r^{n-2k} \quad (2.53)$$

$$N_{nlk} = (-1)^k 2^{l-n} (2n+3)^{1/2} \times \frac{(2n-2k+1)! [(1/2)(n+l)-k]!}{[(1/2)(n-l)-k]! (n+l-2k+1)! (n-k)! k!} \quad (2.54)$$

a $Y_{lm}(\omega)$ są harmonijkami sferycznymi. Wskaźniki n , l , m muszą spełniać następujące reguły: $n \geq l$ i wartość $(n-l)$ musi być parzysta, $-l \leq m \leq l$. Funkcje Z_{nlm} są funkcjami ortogonalnymi tj.:

$$\int_{r \leq 1} Z_{nlm}(\vec{r}) Z_{n'l'm'}(\vec{r}) d\vec{r} = \delta_{nn'} \delta_{ll'} \delta_{mm'} \quad (2.55)$$

Ponieważ wielomiany Zernike'a są ortogonalne można więc rozwinąć dwukrotnie różniczkowalną funkcję za ich pomocą. Gęstość elektronową możemy zapisać w postaci szeregu wielomianów Zernike'a w następujący sposób:

$$\rho(\vec{r}) = \sum_{n=0}^{\infty} \sum_{l=0}^n \sum_{m=-l}^l c_{nlm} Z_{nlm}(\vec{r}) \quad (2.56)$$

Współczynniki c_{nlm} są nazywane trójwymiarowymi momentami Zernike'a i mogą być obliczone za pomocą algorytmu zaproponowanego przez Novotniego i Kleina (Novotni i wsp., 2003). Jeżeli teraz obliczymy transformatę Fouriera funkcji $\rho(\vec{r})$ z użyciem rozwinięcia w postaci równania 2.56 oraz funkcji eksponencjalnej za pomocą harmonijek sferycznych otrzymamy wzór na amplitudę rozpraszania:

$$A(\vec{s}) = 4\pi \sum_{n=0}^{\infty} \sum_{l=0}^n \sum_{m=-l}^l i^l (-1)^{(n-l)/2} c_{nlm} Y_{lm}^*(\Omega) b_n(s) \quad (2.57)$$

$$b_n(s) = \frac{j_n(s) + j_{n+2}(s)}{2n+3} \quad (2.58)$$

gdzie j_n są funkcjami Bessela. Ostatecznie uśredniając kwadrat amplitudy po wszystkich orientacjach wektora s , otrzymamy wyrażenie na intensywność rozpraszania:

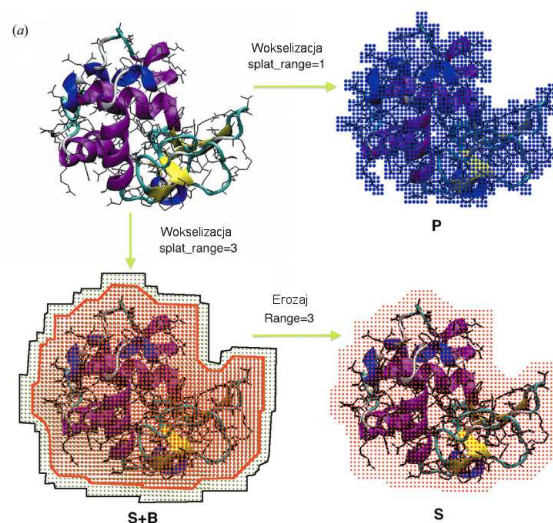
$$I(s) = 16\pi^2 \sum_{n=0}^{\infty} \sum_{n'=0}^{\infty} b_n(sr_{max}) b_{n'}(sr_{max}) F_{nn'} \quad (2.59)$$

$$F_{nn'} = \sum_{l=0}^n k_{nn'l} \sum_{m=-l}^l c_{nlm} c_{n'lm}^* \quad (2.60)$$

$$k_{nn'l} = (-1)^{(n+n')/2-l} \quad (2.61)$$

gdzie r_{max} jest największym rozmiarem cząsteczki.

Algorytm Novotniego i Kleina opiera się na podziale ciągłego obiektu trójwymiarowego (gęstości elektronowej), w naszym przypadku cząsteczki biologicznej, na dyskretne woksele. Na podstawie tego podziału obliczane są współczynniki c_{nlm} .



Rysunek 2.15 Proces tworzenia zbiorów **P**, **S+B** oraz **S**. Zaczepnięte i zmodyfikowane na podstawie Liu i wsp., 2012a.

W procedurze tej (rysunek 2.15) otrzymujemy zbiór **P**, który reprezentuje gęstość elektronową cząsteczki. Następnie otrzymujemy zbiór **S+B**, który reprezentuje nam obszar wyłączanego rozpuszczalnika **S** oraz otoczki hydratacyjnej **B** poprzez dodanie kolejnych wokseli, znajdujących się 3 Å od powierzchni zbioru **P**, do zbioru **P**. Następnie sam zbiór **S** otrzymuje w procedurze nazwanej przez autorów erozją, polegającą na zmniejszeniu obszaru **S+B**. Zaletą tej metody, jest to że, niezależnie czy powierzchnia znajduje się na zewnątrz czy też wewnątrz cząsteczki będzie dawała wkład do zbioru **S+B**. Kolejną zaletą np. w porównaniu z metodą tzw. pseudo-atomów, które reprezentują otoczkę hydratacyjną oraz objętość wyłączoną rozpuszczalnika, jest że w metodzie wielomianów Zernike'a unika się powstawania niejednorodności i efektów wynikających z nakładania się pseudo-atomów.

2.6.6 Metody *ab-initio* do modelowanie niskorozdzielczej struktury makromolekuł

2.6.6.1 Algorytm genetyczny

Historycznie pierwszą metodą służącą do uzyskiwania niskorozdzielczych modeli na podstawie danych z małokątowego rozpraszania promieniowania rentgenowskiego był algorytm genetyczny zaproponowany przez Chacóna i współpracowników w 1998 roku (Chacón i wsp., 1998). W metodzie tej jednym z usprawnień jest zmodyfikowana metoda liczenia krzywej rozpraszania za pomocą wzoru Debye'a. Zamiast bezpośrednio liczyć wyrażenie $\sin(sr_{ij})/(sr_{ij})$ dla każdej pary atomów najpierw tworzony jest histogram odległości między atomami $g(r_i)$, który dzieli wszystkie odległości na N przedziałów. Zmniejsza to istotnie ilość liczonych wyrażeń $\sin(sr_{ij})/(sr_{ij})$ a zarazem zmniejszony jest także czas potrzebny na obliczenie krzywej $I(s)$. Intensywność rozpraszania dla układu N_s atomów o amplitudzie rozpraszania $F_s(s)$ z wykorzystaniem histogramu $g(r_i)$ możemy zapisać następująco:

$$I(s) = N_s F_s^2(s) + F_s^2(s) \sum_{i=1}^N g(r_i) \frac{\sin(sr_i)}{sr_i} \quad (2.62)$$

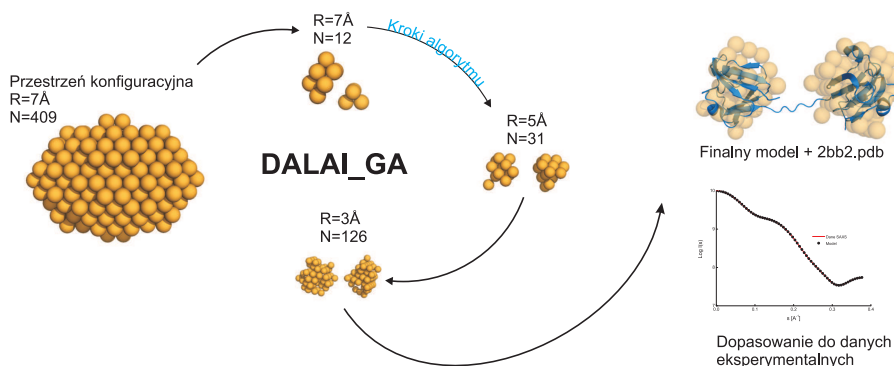
W metodzie tej zaimplementowanej w programie DALAI_GA (Chacón i wsp., 1998) modelowanie struktury niskorozdzielczej białka odbywa się poprzez tworzenie modelu cząsteczki za pomocą kul o zadanej średnicy. Schemat działania programu DALAI_GA przedstawiony jest na rysunku 2.16. Startując z L losowych konfiguracji zbudowanych z M kul (pseudo-atomów) program za pomocą algorytmu genetycznego modyfikuje tworzoną przez nie strukturę i optymalizuje, tak aby znaleźć jedną konfigurację kul, której profil rozpraszania będzie zgadzał się jak najlepiej z danymi eksperymentalnymi. Każda konfiguracja nazywana jest chromosomem. Chromosom zapisany jest w postaci ciągu binarnego gdzie każdy bit reprezentuje jeden punkt w przestrzeni. Natomiast wartość bitu mówi nam czy w danym punkcie przestrzeni znajduje się kula - wartość 1 czy też nie - wartość 0. Algorytm genetyczny w celu optymalizacji opiera się na dwóch operatorach genetycznych: krzyżowania (crossing-over) polegającej na częściowej wymianie bloków bitów pomiędzy dwoma chromosomami (nazwa pochodzi od zjawiska krzyżowania w biologii gdzie podczas podziału komórki dochodzi do wymiany części chromatyd pomiędzy chromosomami homologicznymi) oraz mutacji czyli losowej zmianie pojedynczego bitu. W procesie krzyżowania z dwóch chromosomów rodziców powstaje dwójka chromosomów potomnych. Współczynnik krzyżowania wynosi 0,2 tzn. 20 % materiału genetycznego (bitów) pochodzi od jednego rodzica a 80 % od drugiego. Mutacja polega na kopiowaniu zawartości chromosomu z zadaniem prawdopodobieństwem błędu wynoszącym 0,01. Dla każdego chromosomu (konfiguracji) liczona jest krzywa rozpraszania i porównywana z danymi eksperymentalnymi. Każdej konfiguracji przypisuje się wartość funkcji F , która liczbowo opisuje zgodność modelu z danymi:

$$F = \left(\sum_i (\log(I_{exp}(s_i)) - \log(I_{model}(s_i)))^2 \right)^{-\frac{1}{2}} \quad (2.63)$$

Następnie dokonywana jest klasyfikacja chromosomów ze względu na wartość F , dzieli się je na dwie grupy: P_{elite} do której będą zapisywane chromosomy o najniższej wartości F . Liczebność P_{elite} wynosi 0,5 P_{total} (liczba wszystkich chromosomów). Do modyfikacji losowane są najlepsze chromosomy (frakcja 0,4 P_{total}). Wyboru rodzaju operatora genetycznego dokonuje się losowo z prawdopodobieństwem 0,6 dla mutacji i 0,4 dla krzyżowania. Dodatkowo sprawdzane jest czy powstałe chromosomy nie duplikują się z innymi. Po przeprowadzeniu operacji genetycznych i usunięciu powtarzających się chromosomów nowo powstałe chromosomy zastępują drugą grupę chromosomów ($1 - P_{elite}$). Tak wygląda jeden cykl algorytmu. Proces ten powtarzany jest dopóki przez pewien okres czasu (liczbę cykli) wartość F dla najlepszej konfiguracji nie ulegnie zmianie. W celu szybszego działania algorytmu program startuje z dużą wartością promienia kuli np. $r = 6 \text{ \AA}$, po pewnym czasie promień kuli zostaje zmniejszony, a model zostaje umieszczony w heksagonalnej sieci kul o promieniu $r_2 < r_1$. Kule o promieniu r_2 które znajdują się w odległości $r < 2r_1$ od środka masy kuli r_1 w następnych iteracjach stanowią nową przestrzeń do szukania konfiguracji. Taka zamiana w trakcie symulacji pozwala na efektywniejsze działanie algorytmu oraz ma swoje korzyści gdy nie możemy oszacować wielkości cząsteczki. Ostatecznie możemy zapisać algorytm w postaci bloków:

1. Losowanie początkowego zestaw chromosomów z zadanymi parametrami wejściowymi: r - promień kul, (a, b, c) wymiary przestrzeni poszukiwania np. półosie elipsy, R_g - dozwolony zakres wartości promienia żyracji dla początkowych populacji, M - dozwolony zakres masy dla początkowej populacji.

2. Obliczenie wartości funkcji F dla chromosomów.
3. Podziału na grupę P_{elite} oraz $(1 - P_{elite})$.
4. Na grupie P_{elite} prowadzone są operacje genetyczne z zadaniem prawdopodobieństwem (powstaną nowe chromosomy P_{new}).
5. Selekcja zduplikowanych chromosomy i ewentualne ich usuwanie.
6. Zamiana $(1 - P_{elite})$ na P_{new} tzn. $(1 - P_{elite}) = P_{new}$.
7. Kontrola zbieżności F dla najlepszego chromosomu. Jeśli zbieżność jest osiągnięta program kończy obliczenia. Jeśli nie procedura powtarzana jest od punktu 2.



Rysunek 2.16 Schemat działania algorytmu DALAI_GA. Na rysunku schematycznie pokazane są kolejne kroki i ewolucja modelu oraz jego dopasowanie do struktury krystalograficznej (kolor niebieski) oraz dopasowanie do krzywej eksperymentalnej.

2.6.6.2 Algorytm symulowanego wyżarzania

Inną zaproponowaną metodą do modelowania niskorozdzielczych struktur makrocząsteczek na podstawie danych z eksperymentów małąkątowego rozpraszania promieniowania rentgenowskiego jest metoda na bazie symulowanego wyżarzania zaproponowany przez Sverguna (Svergun 1999). Symulowane wyżarzanie (Kirkpatrick i wsp., 1983) jest algorytmem służącym do poszukiwania minimum globalnego problemu w sposób zoptymalizowany, szczególnie kiedy przestrzeń poszukiwań globalnego minimum jest bardzo duża i jej całkowite przeszukanie byłoby bardzo czasochłonne. Algorytm ten jest modyfikacją algorytmu Metropolis (Metropolis i wsp., 1953) będący adaptacją metody Monte Carlo do rozwiązywania problemów wielu ciał w fizyce. W ogólnym zarysie w metodzie Monte Carlo wykorzystuje się zadany wcześniej rozkład prawdopodobieństwa w celu wylosowania konfiguracji (rozwiązania) obliczenie wartości interesującej nas wartości wielkości fizycznej i powtórzenie tych kroków N razy w celu obliczenia średniej wartości tej wielkości. W algorytmie Metropolis startując z losowej konfiguracji (stanu układu) losowo zmieniamy ten układ np. zmianę położenia atomu i zliczymy różnicę energii ΔE pomiędzy dwoma konfiguracjami. Jeżeli $\Delta E \leq 0$ to zmiana jest akceptowana i nowa konfiguracja jest wykorzystywana w następnym kroku symulacji. Natomiast w przypadku gdy $\Delta E > 0$ nowa konfiguracja jest akceptowana z prawdopodobieństwem $\exp(-\Delta E/k_b T)$, gdzie k_b jest stałą Boltzmanna. Schemat Metropolis oryginalnie wykorzystywany do symulacji układów fizycznych można zaadoptować do innych problemów definiując odpowiedniki energii oraz tempera-

ture. Algorytm Metroplisa pozwala na uniknięcie problemu minimum lokalnego badanego systemu. W metodzie symulowanego wyżarzania stosuje się schemat w którym na samym początku symulacji wykorzystuje się symboliczną wysoką “temperaturę” i obniża się ją w trakcie trwania algorytmu co odpowiada zmniejszaniu prawdopodobieństwa akceptacji konfiguracji o $\Delta E > 0$. Nazwa metody pochodzi od procesu wygrzewania stopów i powolnego ich schładzania w celu otrzymania próbki o wysokim uporządkowaniu. Tak samo w tej metodzie stosowanie “powolnego chłodzenia” układu pozwala na optymalne przeszukanie przestrzeni rozwiązań i znalezienie optymalnego rozwiązania.

Svergun wykorzystał schemat algorytmu symulowanego wyżarzania do symulacji kształtu makrocząstek w programie DAMMIN. Program ten wykorzystuje podobnie jak program DALAI.GA model kulkowy (pseudo-atomy) do generowania modeli. W odróżnieniu od poprzedniego podejścia, metoda zaproponowana przed Sverguna pozwala na otrzymanie modelu cząsteczki składającej się z K faz np. białek i kwasów nukleinowych. Dla cząsteczki składającej się z K faz, każdej o kontraście $\Delta\rho_k$, możemy zapisać wzór na intensywność rozpraszania w następujący sposób:

$$I(s) = \left\langle \left[\sum_{k=1}^K \Delta\rho_k A_k(s) \right]^2 \right\rangle_{\Omega} \quad (2.64)$$

wykorzystując metodę harmonijek sferycznych można zapisać amplitudę k -tej fazy oraz intensywność rozpraszania dla cząsteczki w postaci wzorów:

$$A_k(s) = \sum_{l=0}^{\infty} \sum_{m=-l}^l A_{lm}^k(s) Y_{lm}(\Omega) \quad (2.65)$$

$$I(s) = 2\pi^2 \sum_{l=0}^{\infty} \sum_{m=-l}^l \left\{ \sum_{k=1}^K [\Delta\rho_k A_{lm}^k(s)]^2 + 2 \sum_{n>k} \Delta\rho_k A_{lm}^k(s) \Delta\rho_n [A_{lm}^n(s)]^* \right\} \quad (2.66)$$

$$A_{lm}^k(s) = i^l \sqrt{2l\pi} f(s) \sum_{j=1}^{N_k} j_1(sr_j) Y_{lm}^*(\omega_j) \quad (2.67)$$

gdzie: j_l - funkcją Bessela l -tego rzędu, $f(s)$ - czynnikiem rozpraszania dla pseudo-atomu, N_k - liczba kul - pseudo-atomów w modelu. W celu znalezienia najlepszego modelu należy zminimalizować funkcję χ^2 :

$$\chi^2 = \frac{1}{M} \sum_{i=1}^M \sum_{j=1}^{N(i)} \left[\frac{I_{exp}^i(s_j) - I^i(s_j)}{\sigma^i(s_j)} \right]^2 \quad (2.68)$$

Funkcję χ^2 definiuje się dla $M \geq 1$ krzywych eksperymentalnych (krzywe z eksperymentów wariacji kontrastu) o $N(i)$ punktach i błędzie pomiarowym $\sigma^i(s_j)$.

W przypadku modeli otrzymywanych w programie DALAI.GA można było zauważyć, że często pojawiały się przerwy w modelach lub też występowały luźno związane atomy. Mając na uwadze rozdzielczość danych SAXS oraz fizyczność modelu Svergun wprowadził dla danej konfiguracji atomów X parametr $P(X)$ opisujący upakowanie atomów. Dla sieci heksagonalnej liczba sąsiadów N_c wynosi 12 lub mniej gdy dany pseudo-atom znajduje się na powierzchni modelu. Dla danej fazy kompaktowość struktury opisana jest poprzez zliczenie dla każdego pseudo-atomu danej fazy liczby

kontaktów N_e z innymi pseudo-atomami danej fazy. Funkcja $C(N_e)$ opisuje kompaktowość w sposób bardziej ilościowy:

$$C(N_e) = 1 - Z(N_e) = 1 - [\exp(-0.5N_e) - \exp(-0.5N_c)] \quad (2.69)$$

Wartość funkcji dla $N_e = 12$ do $N_e = 6$ zmienia się łagodnie od 1 do 0,943 ($N_e = 6$ może oznaczać położenie na powierzchni) i spada szybko dla $N_e < 6$ także $C(0) = 0,002$. Parametr $P(X)$ można zdefiniować jako:

$$P(X) = 1 - \langle C(N_e) \rangle_{N_k} \quad (2.70)$$

gdzie $\langle C(N_e) \rangle_{N_k}$ jest uśrednioną po wszystkich atomach wartością funkcji $C(N_e)$. Wprowadzając funkcje $P(X)$ możemy zmodyfikować funkcję χ^2 aby oprócz dobrego dopasowania do danych eksperymentalnych narzucała również więzy na kompaktowość modelu:

$$f(X) = \chi^2 + \alpha P(X) \quad (2.71)$$

gdzie $f(X)$ jest nową funkcją do minimalizowania, α jest wagą kryterium kompaktowości. Wartość $P(X)$ zależy od kształtu cząsteczki dlatego też, wiedząc czy kształt cząsteczki jest mniej czy więcej anizotropowy (np. z kształtu krzywej $P(r)$) możemy zmieniać wartość α . Mając zdefiniowaną funkcję $f(X)$ można zapisać algorytm symulowanego wyżarzania za pomocą schematu:

1. Generowanie losowej konfiguracji kul X_0 w wysokiej temperaturze T_0 .
2. Wylosowanie jednego pseudo-atomu i losowa zmiana jego fazy (nowa konfiguracja X'). Obliczanie $\Delta = f(X') - f(X)$.
3. Jeśli $\Delta < 0$ akceptacja zmiany ($X \rightarrow X'$). Jeśli $\Delta > 0$ akceptacja zmiany z prawdopodobieństwem $\exp(-\Delta/T)$. Powtórzenie drugiego kroku z konfiguracji X lub X' (jeśli została zaakceptowana).
4. Utrzymanie temperatury przez 100 N zmian konfiguracji lub 10 N zmian konfiguracji zakończonych sukcesem ($\Delta < 0$). Następnie zmniejszenie temperatury: $T' = 0,9 T$. Kontynuacja schładzania do ustabilizowania się funkcji $f(X)$.

2.6.6.3 Modelowanie struktury domenowej białek za pomocą algorytmu symulowanego wyżarzania

Algorytm symulowanego wyżarzania został także zaimplementowany do modelowania modeli białek w postaci pseudo-łańcucha (Svergun i wsp., 2001). W modelu tym łańcuch białkowy przedstawiony jest w postaci kul odpowiadającym pojedynczym merom (resztom aminokwasowym) łańcucha polipeptydowego. Krzywe rozpraszania dla różnych białek o różnej masie molekularnej wykazują spore różnice w kształcie do wartości wektora $s = 12 \text{ nm}^{-1}$ co odpowiada rozdzielczości 0,5 nm. Małokątowe rozpraszanie nie nadaje się więc do określania struktur drugorzędowych. Jednak nawet rozdzielczość 0,5 nm znacznie poprawiłaby rozdzielczość względem modeli otrzymanywanych z zastosowaniem modeli pseudo-atomów. Z tą rozdzielczością można teoretycznie modelować strukturę białek i kwasów nukleinowych na poziomie indywidualnych domen. Z właściwości statystycznych

wynika, że dla białek rozkład prawdopodobieństwa znalezienia innego aminokwasu w odległość R posiada charakterystyczny kształt. Oczywiście najbliższy sąsiad znajduje się w odległości 0,38 nm odpowiadającej odległości pomiędzy atomami C_α dwóch kolejnych aminokwasów w łańcuchu polipeptydowym. Wykorzystując powyższe właściwości można nałożyć więzy na łańcuch pseudo-reszt, tak aby jego właściwości strukturalne jak najbardziej przypominały te dla prawdziwych łańcuchów białkowych (otrzymanych ze struktur białkowych zdeponowanych w bazie danych PDB). Program GASBOR napisany przez Sverguna (Svergun i wsp., 2001) implementuje algorytm symulowanego wyżarzania z dodatkowymi więzami opisanymi powyżej. W programie minimalizuje się funkcję $E(X)$ dla danej konfiguracji łańcucha X_j :

$$E(X) = \chi^2 + \alpha P(X) \quad (2.72)$$

$$\chi^2 = \frac{1}{n-1} \sum_{j=1}^n \left[\frac{\langle c(s) \rangle I_{DR}(s_j) - c I_{exp}(s_j)}{\sigma(s_j)} \right]^2 \quad (2.73)$$

gdzie $I_{exp}(s)$ krzywa eksperymentalna o n punktach, $\langle c(s) \rangle$ jest czynnikiem korygującym błąd krzywej rozpraszania dla modelu pseudo-reszt $I_{DR}(s_j)$, dla dużych wartości s gdzie istotny wpływ ma struktura reszt bocznych aminokwasów, $\sigma(s_j)$ - błąd danych eksperymentalnych, c - czynnik skalujący.

$$P(X) = \sum_k [W(R_k) (N_{DR}(R_k) - \langle N(R_k) \rangle)]^2 + G(X) + [\max\{0, (|r_c^z| - r_0)\}]^2 \quad (2.74)$$

gdzie $W(R_k)$ jest odwrotnie proporcjonalna do wariacji na krzywej, $N_{DR}(R_k)$ liczba sąsiadów dla modelu pseudo-reszt w odległości R_k , $\langle N(R_k) \rangle$ - liczba sąsiadów w odległości R_k odczytana z histogramu uśrednionego dla kilku struktur z bazy danych PDB, $G(X) = \ln(N_G/N)$ - człon ten zapewnia nam łańcuch bez przerw, N_G długość najdłuższego grafu powstałego z połączenia pseudo-reszt. Ostatni człon zapewnia, że model będzie znajdował się blisko początku układu współrzędnych. Schemat blokowy algorytmu wygląda bardzo podobnie jak w przypadku algorytmu programu DAM-MIN:

1. W sferze o promieniu $D_{max}/2 + r_0$ losowo rozmieszczane jest N pseudo-reszt (N równe liczbie aminokwasów w białku) oraz M cząsteczek rozpuszczalnika. Liczona jest wartość funkcji $E(X)$ oraz wybierana jest temperatura T_0 .
2. Wybieranie losowo jednej pseudo-reszty i przenoszenie jej na odległość 0,38 nm od drugiej pseudo-reszty. Jeżeli nowe położenie znajduje się poza obszarem przeszukiwania powtarza się krok 2.
3. Jeśli jest potrzeba, następuje zmiana położenia cząsteczek rozpuszczalnika. Obliczana jest różnica $\Delta E = E(X') - E(X)$. Jeśli $\Delta E \geq 0$ ruch jest akceptowany jeśli $\Delta E < 0$ ruch akceptowany jest z prawdopodobieństwem $\exp(-\Delta E/T)$.
4. Kroki 2 i 3 są powtarzane N_T razy dopóki układ się nie ustabilizuje. Po N_T krokach zmienia się temperatura: $T' = 0,9 T$. Schładzanie systemu kontynuowane jest do momentu ustabilizowania się wartości funkcji $E(X)$.

2.6.6.4 Modelowanie bryły sztywnej - wykorzystanie dostępnych struktur wysokorozdzielczych

Metody *ab-initio* pozwalają wymodelować niskorozdzielczą strukturę makrocząsteczek *de novo*. Jest to szczególnie przydatne gdy nie dysponujemy żadnymi danymi strukturalnymi. Pierwsza generacja programów takich jak DALAI.GA, DAMMIN i GASBOR stała się standardem w określaniu niskorozdzielczej struktury na podstawie danych z eksperymentów SAXS. Jednak dają one jedynie niskorozdzielczy opis kształtu makrocząsteczki na poziomie precyzji odpowiadającym danym z mikroskopii elektronowej (TEM). Po zakończeniu projektów sekwencjonowania ludzkiego genomu oraz genomów wielu organizmów (okres ten był nazywany erą genomiki) nastąpił wzrost zainteresowania tym co jest zakodowane w genomie czyli min. białkami. Często używa się terminu ery post-genomowej. Od kilku lat działają na całym świecie konsorcja tzw. genomiki strukturalnej mające za swój cel określenie struktury przestrzennej określonej grupy białek czy też białek z całego organizmu (np. bakterii *Thermotoga maritima* - konsorcjum Joint Center for Structural Genomics w USA). Do określania struktury przestrzennej stosuje się metody krystalografii rentgenowskiej oraz spektroskopii magnetycznego rezonansu jądrowego (NMR). Podejście to w przypadku metody NMR niesie ze sobą ograniczenie co do wielkości cząsteczek. Standardowo za pomocą spektroskopii NMR można badać białka o masie do 20 kDa. W przypadku krystalografii dla wielu białek można było uzyskać strukturę jedynie dla pojedynczych domen lub też skróconych wersji badanych białek. Jednocześnie im większe białko, z większą ilością domen połączonych giętkimi fragmentami, tym trudniej jest otrzymać jego kryształy. Podobnie w przypadku badania struktur kompleksów makromolekuł, otrzymanie kryształów jest niejednokrotnie bardzo trudne a metoda NMR narzuca ograniczenie na wielkość badanego kompleksu. Jak już wspomniano wcześniej metoda SAXS nie niesie ze sobą ograniczenia co do wielkości cząsteczek. Mając do dyspozycji strukturę atomową badanych białek lub ich fragmentów można teoretycznie określić strukturę całego białka/kompleksu tak jak miało to miejsce gdy zamiast modeli atomowych miało się do dyspozycji model kulkowych. Jednak w przypadku dużych białek/kompleksów liczba wzajemnych orientacji do przetestowania w celu znalezienia odpowiedniej, pasującej do danych eksperymentalnych jest bardzo czasochłonna. Dlatego też opracowano szereg metod i algorytmów, które umożliwiają nam modelowanie białek czy kompleksów makromolekuł na podstawie danych SAXS (Svergun i wsp., 2005, Mertens i wsp., 2010).

Metoda modelowania bryły sztywnej polega na translacji i rotacji K podjednostek (traktowanych jak bryła sztywna struktur białek lub kwasów nukleinowych) tak aby obliczona krzywa rozpraszania była jak najbardziej zgodna z danymi eksperymentalnymi. Dla układu K podjednostek można zapisać wzór na krzywą rozpraszania:

$$I(s) = \left\langle \left| \sum_{k=1}^K A^{(k)}(\vec{s}) \right|^2 \right\rangle_{\Omega} \quad (2.75)$$

$$A^{(k)}(\vec{s}) = \exp(i\vec{s}\vec{r}_k) \prod (\alpha_k \beta_k \gamma_k) \left[C^{(k)}(\vec{s}) \right] \quad (2.76)$$

gdzie: $\exp(i\vec{s}\vec{r}_k)$ jest operatorem translacji o wektor \vec{r}_k przestrzeni rzeczywistej, $\prod (\alpha_k \beta_k \gamma_k)$ - jest operatorem obrotu o kąty Eulera α , β , γ , $C^{(k)}(\vec{s})$ - jest krzywą rozpraszania dla k-tej podjednostki w położeniu referencyjnym. Także amplitudy $A^{(k)}(s)$ zależą jedynie od sześciu parametrów:

$x_k, y_k, z_k, \alpha_k, \beta_k, \gamma_k$. Jeżeli wyrazimy amplitudę $A^{(k)}(\vec{s})$ za pomocą harmonijek sferycznych (wzór 2.65) to współczynniki $A_{lm}^{(k)}(s)$ będzie można wyrazić za pomocą współczynników $C_{lm}^{(k)}(s)$ (Svergun i wsp., 1997). Dodatkowo jeżeli układ charakteryzują się pewną symetrią lokalną np. osią dwukrotną (w przypadku dimerów) lub inną symetrią niekrytalograficzną odpowiadającą np. grupie P_n lub P_n^2 to w rozwinięciu w harmonijki sferyczne wkład będą dawać symetrycznie niezależne podjednostki oraz jedynie współczynniki $A_{lm}^{(k)}(s)$ o $m = 0, m = i * n$. Dodatkowo dla grup punktowych P_n^2 , człony $A_{l0}^{(k)}(s)$ z nieparzystym l znikają. Dlatego dla symetrycznych oligomerów obliczenia stają się znacznie prostsze. Dodatkowo można narzucić więzy na sprecyzowane miejsce oddziaływania (np. informacje z innych technik np. spektroskopii NMR lub mutagenezy). W ogólności do opisu krzywej rozpraszania potrzebne są $6K - 6$ parametrów ($6K/n - 2$ dla P_n , $3K/n$ dla P_n^2), które zmienia się w trakcie działania algorytmu. W zależności od tego jak dużo jest tych parametrów różne algorytmy wykorzystują różne metody minimalizacji.

Z najprostszym przypadkiem mamy do czynienia podczas modelowania homo- i heterodimeru. Dla takiego przypadku mamy dwie różne lub tylko jedną podjednostkę. Svergun zaproponował metodą modelowania opierającą się na funkcji $F(\omega)$. Najpierw generowane jest N kierunków - siatka (ω_i) w przestrzeni za pomocą liczb Fibonacciego. W metodzie tej wybierany jest kierunek (ω_i) i obracana jest pierwsza podjednostka tak aby $\omega_i = z$. Następnie dla drugiej podjednostki obracamy jej siatkę (ω_j) tak aby jej ω_j -ty kierunek był antyrównoległy do kierunku ω_i . W przypadku homodimerów $j = i$. Drugi monomer następnie przesuwany jest wzdłuż osi z o wartość $F^{(1)}(\omega_j) + F^{(2)}(\omega_i) + \delta$ (parametr $\delta = 0, 3nm$ dobrany jest aby zapobiec zawodom sferycznym) i obracany o kąt $0 < \psi < 2\pi$. Algorytm ten jest zaimplementowany w programie DIMFOM (Svergun i wsp., 2005), w którym ewaluowane są wszystkie pary ω_i, ω_j oraz obroty o kąt ψ .

2.6.6.5 Modelowanie homooligomerów o określonej symetrii

W przypadku modelowania symetrycznych oligomerów o symetrii innej niż P_2 powyższy algorytm się już nie sprawdza. Jednak w tym przypadku możliwe jest dosyć szybkie znalezienie najlepszej struktury pasującej do danych eksperymentalnych stosując nieco inne podejście. Układ oligomeru zdefiniowany jest przez 6 parametrów: przesunięcie względem początku układu współrzędnych $\vec{r} = (r, \psi, \phi)$ oraz obrotów o kąty Eulera α, β, γ . Dodatkowo właściwości harmonijek sferycznych względem symetrii znacznie upraszczają liczenie funkcji $I(s)$. Długość wektora \vec{r} zdefiniowana jest w granicach: $0,95 \times r_0 < |\vec{r}| < 1,05 \times r_0$, gdzie $r_0 = \sqrt{(R_g^{\text{exp}})^2 + (R_g^{\text{mon}})^2}$ a $R_g^{\text{exp}}, R_g^{\text{mon}}$ są promieniami żyracji dla oligomeru (z eksperymentu) i dla monomeru (obliczony teoretycznie) odpowiednio. Wartości (ψ, ϕ) brane są z siatki Fibonacciego. Aby zapobiec otrzymaniu struktury oligomeru o monomerach nie będących w kontakcie między sobą lub mających zawady steryczne wprowadza się funkcję, która opisuje czy spełnione są warunki na brak zawad sterycznych i oddziaływanie pomiędzy monomerami. Dla każdego atomu węgla C_α jednego monomeru, kreślona jest sfera o promieniu $0,76 \text{ nm}$ ($\sim 2 \times$ odległość pomiędzy dwoma atomami C_α w łańcuchu peptydowym). Identyfikowana jest liczba atomów C_α należących do innego monomeru wewnątrz każdej sfery i dla każdej pary liczona jest odległość $d = C_\alpha - C_{\alpha'}$. Jeśli $d < 0,38 \text{ nm}$ para daje wkład $(1/d - 1/0,38)$ do funkcji opisującej kryterium kontaktu $P_{interfejs}$. Jeśli $0,38 \text{ nm} < d < 0,76 \text{ nm}$ para daje wkład $1/d$. Przyjmuje się, że dwa monomery oddziałują ze sobą gdy wartość funkcji $P_{interfejs} > 2,5 \text{ nm}^{-1}$. Obliczona na całego oligomeru

funkcja $P_{interfejs}$ jest normalizowana do liczby monomerów. Wygenerowane struktury oligomeru, w której cząsteczki monomerów nie oddziałują między sobą oraz te, dla których $P_{interfejs} > 50 \text{ nm}^{-1}$ są odrzucane bez liczenia wartości funkcji $I(s)$. Dodatkowo, można narzucić więzy związane ze zdefiniowaniem odległości pomiędzy parą atomów C_α . Jeżeli dla N par spodziewamy się odległości między nimi d_{n0} , to możemy zdefiniować funkcję: $P_{kontakty} = \sum_{n=1}^N (\max(0, d_n - d_{n0}))^2$, gdzie d_n jest odległością dla n -tej pary w danej konfiguracji. Tak zdefiniowane więzy można zebrać w postaci funkcji E :

$$E = \chi^2 + \alpha_{interfejs} P_{interfejs} + \alpha_{kontakty} P_{kontakty} \quad (2.77)$$

z wagami $\alpha_{interfejs}$, $\alpha_{kontakty}$ dobranymi tak aby człon z więzami dawał wkład 10 % – 50 % do funkcji E . Funkcja χ^2 ma analogiczną postać jak wzór 2.68 i określa dopasowanie modelu do danych eksperymentalnych. W trakcie działania algorytmu minimalizowana jest właśnie funkcja E .

2.6.6.6 Modelowania hetero-oligomerów

W przypadku kompleksów o różnych podjednostkach przeszukanie wszystkich możliwych konfiguracji jest bardzo czasochłonne, gdyż nie obowiązują już uproszczenia opisane powyżej. W tym przypadku trzeba zastosować metody optymalizacji w celu znalezienia odpowiedniej konfiguracji. W przypadku hetero-oligomerów Svergun i współpracownicy po raz kolejny posłużyli się algorytmem symulowanego wyżarzania. W programie SASREF zamiast zmieniać pozycję pseudo-atomów, losuje się podjednostkę oraz kierunek, w którym dokonuje się translacji o zadaną długość oraz oś wokół której dokonuje się rotacji o zadany kąt (Petoukhov i wsp., 2005). Za każdym razem liczone są amplitudy $A_{lm}^{(k)}$ jedynie dla podjednostki, która zmienia swoje położenie. Na początku działania algorytmu ustawiana jest wysoka temperatura a ruchy są niemal dowolne. Natomiast w trakcie działania programu temperatura jest stopniowo obniżana by osiągnąć minimalną wartość funkcji E (wzór 2.78). Dodatkowo można dopasowywać modele do kilku zestawów danych eksperymentalnych np. dla konstruktów zawierających tylko kilka domen z całego kompleksu. Dzięki temu możliwe jest dokładniejsze wymodelowanie struktury kompleksu. W programie minimalizowana jest funkcja E podobna do wzoru 2.77:

$$E = \sum_{i=1}^M \chi_i^2 + \alpha_{dist} P_{dist} + \alpha_{interfejs} P_{interfejs} + \alpha_{kontakty} P_{kontakty} \quad (2.78)$$

Gdzie χ_i^2 - są funkcjami nieosiującymi dopasowanie do i -tego zestawu danych, waga α_{dist} oraz funkcja $P_{dist} = \ln(K/K_G)$ (K_G liczba podjednostek w najdłuższym grafie reprezentującym oddziałujące podjednostki) zapewnia, że wszystkie podjednostki w kompleksie powinny oddziaływać ze sobą. Wagi i funkcje: $\alpha_{interfejs} P_{interfejs}$, $\alpha_{kontakty} P_{kontakty}$ zdefiniowane są podobnie jak dla symetrycznych oligomerów.

W przypadku, gdy dysponujemy jedynie strukturą dla fragmentu badanego białka, możliwe jest połączenie dwóch metod: metody pseudo-łańcucha oraz metody bryły sztywnej. W metodzie tej brakujące części białka zastępowane są za pomocą pseudo-łańcuchów. A krzywa rozpraszania dla K podjednostek oraz N łączników ma następującą postać:

$$I(s) = 2\pi^2 \sum_{l=0}^{\infty} \sum_{m=-l}^l \left| \sum_{k=1}^K A_{lm}^{(k)}(s) + \sum_{j=1}^N D_{lm}^{(j)}(s) \right|^2 \quad (2.79)$$

Gdzie $A_{lm}^{(k)}(s)$, $D_{lm}^{(j)}(s)$ są amplitudami dla modeli atomowych i pseudo-łańcuchów odpowiednio. A zmodyfikowana funkcja E ma postać:

$$E = \sum_{i=1}^M \chi_i^2 + \alpha_{dist} P_{dist} + \alpha_{interfejs} P_{interfejs} + \alpha_{kontakty} P_{kontakty} + \alpha_{k\acute{a}ty\ torsyjne} P_{k\acute{a}ty\ torsyjne} + \alpha_{wi\acute{a}zania} P_{wi\acute{a}zania} + \alpha_{kompakt} P_{kompakt} \quad (2.80)$$

Gdzie: $P_{kompakt} = \left[\sum_j (\max(0, R_g^j - 3\sqrt[3]{M_j}))^2 \right] / \left[9 \sum_j \left(\sqrt[3]{M_j^2} \right) \right]$ zapewnia, że pseudo-łańcuchy nie będą miały zbyt rozciągniętej konformacji (R_g^j -promień żyracji j-tego pseudo-łańcucha o M_j pseudo-resztach, $3\sqrt[3]{M_j}$ -spodziewany promień żyracji globularnego białka o M_j aminokwasach), $P_{k\acute{a}ty\ torsyjne}$, $P_{wi\acute{a}zania}$ są więzami na długość wiązań i kąty torsyjne pomiędzy pseudo-łańcuchami odpowiednio. Metoda ta zaimplementowana w programach BUNCH (modelowanie pojedynczych łańcuchów polipeptydowych przyłączonych do sztywnych domen) i CORAL (modelowanie kompleksów o kilku łańcuchach polipeptydowych) jest warta uwagi, ponieważ wiele struktur zdeponowanych w bazie danych PDB posiada nieuporządkowane pętle czy fragmenty w strukturze, które można modelować w tych programach poprawiając dopasowanie do danych eksperymentalnych (Petoukhov i wsp., 2005).

2.6.6.7 Inne metody modelowania na podstawie danych z małokątowego rozpraszania promieniowania rentgenowskiego

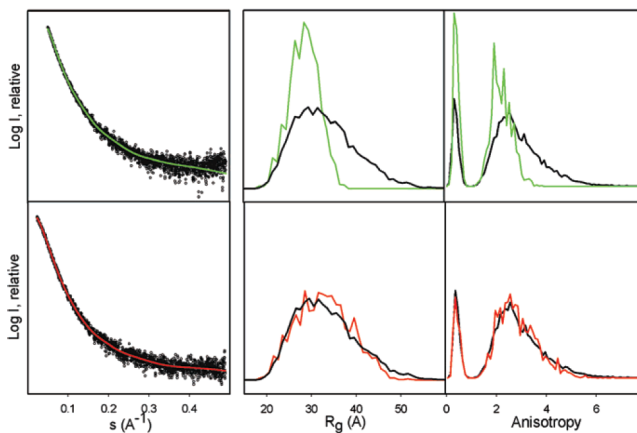
W laboratorium Yanga Zhanga na Uniwersytecie w Michigan stworzono pakiet do przewidywania struktur białek SAXSTER (<http://zhanglab.ccmb.med.umich.edu/SAXSTER/>). SAXSTER jest połączeniem metod przewidywania struktur białek opracowanych w laboratorium Zhanga takich jak I-TASSER czy Quark służących do modelowania homologicznego i *ab-initio* odpowiednio. SAXSTER na podstawie krzywej rozpraszania lub krzywej P(r) filtruje modele otrzymane w procesie modelowania tak aby obliczona krzywa rozpraszania dla modelu zgadzała się z danymi eksperymentalnymi. Ciekawym pomysłem jest połączenie metod dynamiki molekularnej bazujących na zredukowanym modelu (ang. coarse grained) białek i kwasów nukleinowych. Sprzężenie metod takich jak: UNRES (Gołaś i wsp., 2012, Liwo i wsp., 2014) czy też metoda Kolińskiego (Koliński 2004) oparta na algorytmach Monte Carlo do *de-novo* przewidywania struktur białek, z możliwością selekcji na podstawie danych otrzymanych z eksperymentu SAXS mogłoby by doskonale sprawdzić się w badaniu struktury białek z wykorzystaniem danych z eksperymentów SAXS.

Do badania białek natywnie nieuporządkowanych, jak również posiadających giętkie sekwencje łączące domeny metody opisane powyżej nie sprawdzają się już tak dobrze jak w przypadku białek globularnych. Dzieje się tak dlatego, że białka nieuporządkowane z definicji nie posiadają jednej struktury lecz występują w postaci zbioru kilku konformerów. Również białka posiadające tylko małe fragmenty nieuporządkowane często wykazują podobny dynamiczny charakter. Do opisu tego typu

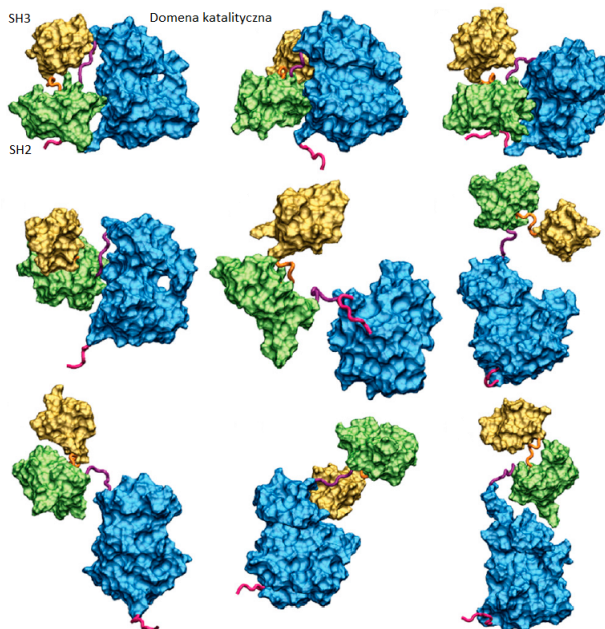
białek było niezbędne zastosowanie innego podejścia. Bernadó i współpracownicy opracowali metodę do badania białek nieuporządkowanych, którą nazwali EOM (ang. ensemble optimization method) (Bernadó i wsp., 2007). Metoda ta opiera się na generowaniu losowych konformacji badanego białka (w reprezentacji za pomocą pseudo-reszt), w celu stworzenia ich zestawu. Wygenerowana pula modeli posiada pewien rozkład, w którym zawarte są zarówno rozciągnięte jak i bardziej kompaktowe struktury. Program RanCh służy do generowania puli modeli. Umożliwia on generowanie oligomerycznych struktur np. dimerów ze zdefiniowanym przez użytkownika interfejsie oddziaływania. Dodatkowo, można wykorzystać modele atomowe, które będą traktowane jak bryły sztywne, gdzie tylko fragmenty pomiędzy domenami będą wymodelowane w postaci pseudo-reszt. Generowane modele są odrzucane jeśli będą posiadały zawady steryczne. Tak przygotowana pula modeli, lub też pula modeli dostarczona przez użytkownika służy następnie w programie Gajoe (Bernadó i wsp., 2007) do wyznaczenia minimalnego zespołu, który opisuje nam zachowanie się białka w roztworze i najlepiej dopasowuje się do eksperymentalnych danych z małokątowego rozpraszania promieniowania rentgenowskiego. Program Gajoe wykorzystuje algorytm genetyczny do dopasowania zespołu konformacji do danych eksperymentalnych. Z wygenerowanych modeli tworzone są chromosomy o liczbie konformerów $N > 10$ (autorzy programu sugerują liczbę $N \approx 50$). Najczęściej $K = 50$ chromosomów jest losowo generowanych z losowego zespołu struktur. W algorytmie stosuje się dwie operacje genetyczne: mutacje (gdzie 20 % genów (gen jest w tym przypadku jedną z konformacji badanej makromolekuły) jest wymieniana na inne lub 50 % genów jest na nowo losowanych z losowej puli lub 50 % genów jest losowanych z puli reszty $K-1$ chromosomów) oraz krzyżowania (losowane są dwa chromosomy i wymieniane są co najmniej dwa geny pomiędzy nimi). Na koniec jednej rundy operacji powstaje $3K$ chromosomów z czego wybierane jest K najlepszych. Przeważnie do 5000 generacji jest powtarzanych w toku działania algorytmu. Na sam koniec działania algorytmu otrzymujemy wybrany rozkład konformacji badanej makrocząsteczki, których rozkład promienia żyracji R_g oraz maksymalnego rozmiaru cząsteczki D_{max} jest porównywany z rozkładem dla losowej puli. Z rozkładu można wywnioskować czy badane białko jest bardziej sztywne (wąski rozkład R_g) czy też jest niepofałdowane (szeroki rozkład R_g podobny do losowej puli). Metoda EOM pozwala na wzbogacenie zespołu w struktury, które są reprezentowane na niskim poziomie w losowej puli (skrajne przypadki z rozkładu losowej puli modeli). Metoda ta doskonale się sprawdziła w przypadku badań nad zdenaturowanym lizozymem w 8 M moczniku (testowe białko niepofałdowane) jak również dla wielodomenowego białka z dynamicznymi fragmentami pomiędzy nimi - kinaza tyrozynowa Brutona (Bernadó i wsp., 2007). Na rysunku 2.17, zaprezentowane są wyniki modelowania lizozymu z jaja kurzego dla białka ze zredukowanymi mostkami di-siarczkowymi (kolor czerwony) i z niezredukowanymi mostkami (kolor zielony). Zauważalna jest różnica pomiędzy tymi dwoma białkami. Zredukowany lizozym ma bardziej dynamiczną strukturę ponieważ jego profil R_g jest praktycznie taki sam jak dla losowej puli modeli.

Innym podejściem do modelowania białek o dynamicznej strukturze jest metoda BSS-SAXS (ang. basis-set supported SAXS) (Yang i wsp., 2010). Metoda ta w skrócie opiera się na symulacji dynamiki molekularnej opartej na gruboziarnistym modelu białka. Dzięki takiej reprezentacji możliwe jest symulowanie dłuższych czasowo procesów oraz pozwala na uchwycenie nisko reprezentowanych konformacji. W metodzie tej symulacja dynamiki molekularnej zaimplementowana jest z zastosowanie

równania Langevina oraz zredukowanego modelu białka – modelu GO (Hills i wsp., 2009).



Rysunek 2.17 Zastosowanie metody EOM do badania białek zdenaturowanych. Panel górny (w kolejności od lewej): krzywa rozpraszania dla lizozymu w 8M moczniku wraz z dopasowaną krzywą dla modeli otrzymanych w metodzie EOM, wykres przedstawiający rozkład promienia żyracji dla losowej puli (kolor czarny) i dla dopasowanego zestawu konformacji (kolor zielony), rozkład anizotropii kształtu modeli dla losowej puli (kolor czarny) i wybranego zestawu konformacji (kolor zielony). W panelu dolnym przedstawione są te same dane lecz dla próbki zredukowanej za pomocą DTT. Rysunek zapożyczono z pracy Bernadó i wsp., 2007.



Rysunek 2.18. Dziewięć różnych konformacji kinazy Hck otrzymanych w metodzie BSS-SAXS. Rysunek zapożyczony i zmodyfikowany z pracy (Yang i wsp., 2010).

Nazwa pochodzi od autora pierwszych prac symulacyjnych (Ueeda i wsp., 1978), w których użyto tego typu zredukowany model białka w którym białko reprezentowane jest za pomocą atomów C_α . Zmodyfikowany jest także potencjał, w którym energetycznie korzystne są natywne, znane wcześniej kontakty pomiędzy aminokwasami. Model ten używa się do badania procesu fałdowania się białek. W pierwszym kroku przeprowadza się kilka symulacji dynamiki molekularnej w celu otrzymania zestawu struktur.

W dynamice molekularnej można zmieniać parametry oddziaływania tak aby otrzymać jak największą przestrzeń konformacyjną. Modele są klastrowane a następnie grupowane na podstawie podobieństwa krzywych rozpraszania obliczonych za pomocą metody FAST-SAXS zoptymalizowanej dla modeli gruboziarnistych (Yang i wsp., 2009). W ten sposób przygotowuje się zestaw krzywych, których kombinacje liniową przygotowuje się tak aby pasowała do danych eksperymentalnych przy pomocy metody Bayesian Monte Carlo. Metoda ta po raz pierwszy została zaimplementowana do określenia struktury przestrzennej i dynamiki konformacyjnej kinazy tyrozynowej Hck biorącej udział w aktywacji układu odpornościowego (Martin 2001) (rysunek 2.18).

Autorzy pracy wykazali, że w roztworze kinaza Hck występuje w równowadze co najmniej dwóch konformacji: nieaktywnej-zamkniętej, gdzie domeny SH2 i SH3 oddziałują z domeną kinazy oraz aktywnej konformacji gdzie domeny SH2 i SH3 z nią nie oddziałują a samo białko ma bardziej rozciągniętą konformację. Po związaniu peptydów przez domeny SH2 i SH3 białko Hck istnieje w wielu różnych konformacjach. Za pomocą SAXS i metody BSS udało się zbadać jak dynamika konformacyjna może wpływać na modulowanie funkcji kinaz z rodziny Src.

Podobne podejście stosuje metoda EROS (ang. ensemble refinement of SAXS) (Różycki i wsp., 2011). Metoda ta składa się z dwóch kroków. W pierwszym kroku przeprowadzana jest symulacja metodą Replica Exchange Monte Carlo (REMC) w temperaturze 300 K z gruboziarnistym potencjałem oddziaływań (oddziaływania pomiędzy resztami bocznymi aminokwasów), który składa się z dwóch członów (Karanicolas i wsp., 2002, Kim i wsp., 2008): pierwszy człon opisuje oddziaływania krótko-zasięgowe:

$$u_{ij}(r) = |\varepsilon_{ij}| \left[13 (\sigma_{ij}/r)^{12} - 18 (\sigma_{ij}/r)^{10} + 4 (\sigma_{ij}/r)^6 \right] f(r) \quad (2.81)$$

gdzie:

$$f(r) = \frac{1}{1 + (r/1.5 \sigma_{ij})} \quad (2.82)$$

Funkcja $f(r)$ - zapewnia gładkie odcięcie potencjału. Potencjał ten bazuje na statystycznym potencjale otrzymanym za danych eksperymentalnych. Potencjał ten różni się od potencjału Lennarda-Jonesa tym, że zawiera w sobie dodatkowy człon, który daje wkład pochodzący od cząsteczek rozpuszczalnika związanego z białkiem (w postaci bariery energetycznej), które muszą zostać przemieszczone w celu utworzenia kontaktu pomiędzy aminokwasami. Drugim członem w potencjale są daleko-zasięgowe oddziaływania elektrostatyczne modelowane za pomocą potencjału Debye'a-Hückela:

$q_i q_j \exp(-r/\xi) / \varepsilon_r r$ (ξ - długość ekranowania Debye'a). Na samą symulację REMC składa się 20 równoległe prowadzonych symulacji Monte Carlo w temperaturze od 300 K do 485 K (T_1, T_2, \dots, T_{20}). W symulacji typu REMC co jakiś czas pomiędzy symulacjami o sąsiednich temperaturach np. T_5 i T_6 , następuje wymiana aktualnych konfiguracji np. konfiguracja z symulacji w temperaturze T_5 stają

się nową konfiguracją w temperaturze T_6 a prawdopodobieństwo przyjęcia nowej konfiguracji określa algorytm Metropolisa. Takie podejście ma na celu przyspieszenie symulacji i ominięcie problemu minimum lokalnego. Symulacja trwa $2 * 10^8$ kroków Monte Carlo: 10^8 kroków stanowi równoważenie systemu a przez kolejne 10^8 kroków prowadzona jest sama symulacja. W trakcie samej symulacji REMC wybieranych jest 10^4 modeli do dalszych analiz, każdy po 10^4 kroku symulacji. W symulacji domeny/białka traktowane są jako bryły sztywne. W drugim korku wybrane do analizy modele są klastrowane, a same klastry następnie służą do otrzymania krzywej rozpraszania podobnie jak miało to miejsce w metodzie EOM. Metoda EROS wykorzystuje metodę maksymalnej entropii w celu uniknięcia zbyt dokładnego, nierealistycznego dopasowania do danych eksperymentalnych. Zaletą tej metody jest zastosowanie „fizycznego” jak określają autorzy pracy podejścia zastosowanego w konstrukcji potencjału oddziaływania. Wprowadzenie potencjału Debye’a–Hückela pozwala na wybranie odpowiedniej siły jonowej dla danej symulacji, żeby otrzymać warunki jak najlepiej zgodne z warunkami w eksperymencie. W odróżnieniu od innych metod, metoda EROS zakłada symulację w bardziej fizjologicznych warunkach niż inne podejścia.

Podejściem, które coraz bardziej zyskuje na popularności jest łączenie danych z spektroskopii magnetycznego rezonansu jądrowego z danymi z eksperymentów SAXS/WAXS (Grishaev i wsp., 2005). Dane SAXS mogą zostać użyte w postaci więzów w procesie obliczania struktury na podstawie danych NMR. Innym podejściem może być dyskryminowanie już otrzymanych za metody NMR modeli na podstawie danych SAXS. W podejściu tym ze zbioru modeli wybierane są te, których obliczona krzywa rozpraszania lub funkcja $P(r)$ najlepiej pasują do eksperymentalnych danych SAXS (Kozak i wsp., 2010).

Jak wspomniano wyżej dane SAXS mogą zostać użyte jako więzy w celu obliczania struktury białek i kwasów nukleinowych za pomocą metody NMR. Najbardziej popularne jest użycie danych SAXS z danymi z eksperymentów RDC. Resztkowe sprzężenie dipolowe RDC (ang. residual dipolar coupling) jest sprzężeniem pomiędzy dwoma dipolami magnetycznymi występujące gdy nie jest uśrednione z powodu ruchów molekularnych sprzężenie dipolowe. Hamiltonian dla dwóch oddziałujących spinów jądrowych ma następującą postać (Levitt i wsp., 2008):

$$\hat{H} = -\frac{\mu_0}{4\pi} \frac{\gamma_I \gamma_S \hbar}{r_{IS}^3} \left(3 \left(\vec{I} \cdot \vec{e}_{IS} \right) \left(\vec{S} \cdot \vec{e}_{IS} \right) - \vec{I} \vec{S} \right) \quad (2.83)$$

gdzie: r_{IS} jest odległością pomiędzy dwa spinami jądrowymi o spinach \vec{S} i \vec{I} odpowiednio, a \vec{e}_{IS} jest wektorem jednostkowym równoległym do odcinka łączącego dwa jądra. Sprzężenie dipolowe jest więc odwrotnie proporcjonalne do r^3 odległości pomiędzy dwoma spinami. Powyższy wzór można zapisać w prostszej postaci:

$$\hat{H} = D_{IS}(\theta) [2I_z S_z - (I_x S_x + I_y S_y)] \quad (2.84)$$

$$D_{IS}(\theta) = \frac{\mu_0}{4\pi} \frac{\gamma_I \gamma_S \hbar}{r_{IS}^3} [1 - 3 \cos^2 \theta] \quad (2.85)$$

gdzie: θ - jest kątem pomiędzy wektorem indukcji magnetycznej \vec{B} a wektorem łączącym dwa jądra. W przypadku makromolekuł w roztworze sprzężenie dipolowe uśredniane jest do zera ze względu na dynamikę ruchów Browna. Jednak jest możliwe cząstkowe uporządkowanie np. poprzez dodanie

do rozтворu cząsteczek porządkujących układ (biceli lub np. wirusów o pałeczkowatym kształcie), które orientują się w polu magnetycznym spektrometru NMR. Udać się wtedy zmierzyć sprzężenie dipolowe - nazywane wtedy resztkowym sprzężeniem dipolowym. Wartość RDC jest jak wspomniano zależna od r^{-3} , dostarcza więc informacji na temat daleko zasięgowych korelacji w strukturze białka w porównaniu z odległościami próbkowanymi za pomocą jądrowego efektu Overhausera - NOE (ang. nuclear Overhauser effect), którego siła zależy od r^{-6} (Cavanagh i wsp., 2007). Informację na temat RDC można połączyć z danymi SAXS w celu modelowania struktur białek i kwasów nukleinowych (Grishaev i wsp., 2005, Wang i wsp., 2010b). Żeby w pełni scharakteryzować cząsteczkę potrzeba dużej liczby zmierzonych RDC, a dodatkowe więzy z eksperymentu SAXS pozwalają na zwiększenie precyzji modelowania. Jednocześnie w równaniu (2.85) dla czterech różnych wartości kąta θ otrzyma się tą samą wartość $D_{IS}(\theta)$. Wprowadza to niejednoznaczności mogące prowadzić do otrzymania błędnych struktur. Zastosowanie tutaj danych SAXS pozwala znieść wspomniane niejednoznaczności, ponieważ krzywa SAXS będzie się różniła dla każdego kąta θ dającego tą samą wartość D_{IS} . Dane SAXS można także wykorzystać do obliczeń struktury za pomocą pakietów do modelowania NMR np. NIH-Xplor (Schwieters i wsp., 2014). Dodatkowy potencjał ma następującą postać:

$$E_{SAXS} = k_{SAXS} \sum_i^N \omega_i (I_i^{CALC} - I_i^{EXP})^2 \quad (2.86)$$

Gdzie: k_{SAXS} zapewnia odpowiednią wagę więzów z eksperymentu SAXS, $\omega_i = 1/(N\sigma_i^2)$ - N liczba punktów na krzywej SAXS, σ_i -błąd pomiaru SAXS, I_i^{CALC} - obliczona krzywa rozpraszania dla modelu, I_i^{EXP} - eksperymentalna krzywa rozpraszania. Potencjał E_{SAXS} jest podobny do funkcji χ^2 . Metodę tą można stosować zarówno do białek jak również do kwasów nukleinowych. Zuo i współpracownicy określili strukturę 30 kDa dimeru receptora pętli czteronukleotydowej RNA (Zuo i wsp., 2008). RMSD pomiędzy oryginalną strukturą obliczoną na podstawie eksperymentu NMR a strukturą obliczoną z więzami SAXS wyniósł 3,2 Å. Metoda łącząca NMR i SAXS/WAXS pozwala na otrzymanie poprawnych struktur białek i kwasów nukleinowych nawet dla relatywnie dużych cząsteczek, trudnych do badania za pomocą metody NMR.

2.6.7 Metoda SVD

Metoda rozkładu względem wartości osobliwych SVD (ang. singular value decomposition) jest metodą rozkładu macierzy na trzy ortogonalne macierze (<http://mathworld.wolfram.com/SingularValueDecomposition.html>): U, S, V^T tj:

$$A = USV^T \quad (2.87)$$

gdzie A jest macierzą o wymiarach $N \times M$, U ma wymiar $N \times M$ a kolumny tej macierzy nazywane są lewymi wektorami własnymi, S ma wymiar $M \times M$ i jest macierzą diagonalną, której wartości nazywane są wartościami własnymi, V ma wymiar $M \times M$ i jej kolumny nazywane są prawymi wektorami własnymi. Macierz S zawiera wagi dla poszczególnych wektorów własnych macierzy U. Jej wartości mają własność, że $s_{11} > s_{22} > s_{33} \dots > s_{MM}$. Wektory własne U z mniejszymi indeksami dają większy wkład do rozkładu macierzy A. Za pomocą metody SVD jesteśmy w stanie zredukować serię danych do mniejszej liczby komponentów tj. zaprezentować je jako kombinację

liniową pewnych wektorów własnych. Macierz U ma ten sam wymiar jak wyjściowa macierz A , lecz w wielu przypadkach tylko kilka pierwszych wektorów własnych U jest w stanie zrekonstruować wyjściową macierz A z zadany błądem. Także można zapisać:

$$A \approx A' = U' S' V'^T \quad (2.88)$$

gdzie nowe macierze: U' , S' , V' mają zredukowany wymiar M do liczby M' nie zerowych wartości własnych macierzy U . Za pomocą testów statystycznych można określić, które wektory własne macierzy U są statystycznie nie istotne (np. test Wilcozona w programie PRIMUS z pakietu ATSAS). Inną metodą jest obliczenie funkcji autokorelacji dla wektorów własnych macierzy U i wybranie jedynie tych dla, których funkcja autokorelacji ma niezerową wartość (funkcja autokorelacji dla szumu ma taki sam charakter jak sam szum). Metoda SVD znalazła zastosowanie w problemach naukowych ponieważ pozwalała na oszacowanie za pomocą kombinacji ilu komponentów da się zapisać otrzymaną serię danych (Feng i wsp., 2013, de Juan i wsp., 2014). W zastosowaniu do metody SAXS, metoda SVD pozwala oszacować ile różnych różnych cząsteczek (np. oligomerów, różnych konformacji danego białka) daje wkład do rozpraszania w serii danych np. eksperymentów czasowo-rozdzielczych czy też zmianie pH, siły jonowej, stężenia cząsteczek w roztworze temperaturze itp. (Williamson i wsp., 2008, Minh i wsp., 2013). Nie mniej jednak wektory własne U nie mają żadnego sensu fizycznego tak jak macierz V^T , która opisuje jak zmienia się udział poszczególnych wektorów własnych w pojedynczych zestawach danych. Nie mniej jednak okazuje się, że da się zrekonstruować krzywe rozpraszania i/lub profil stężenia komponentów za pomocą metody SVD (Williamson i wsp., 2008, Galantini i wsp., 2010).

Niech macierz A będzie teraz macierzą danych, gdzie każda kolumna jest jedną serią danych (N krzywych) o q punktach tj. macierz A ma wymiar $q \times N$. Jeżeli układ da się opisać za pomocą K komponentów tj. w poszczególnej serii danych w eksperymencie każda krzywa jest kombinacją liniową krzywych dla K komponentów a zmienia się jedynie ich ułamek wagowy (objętościowy) możemy zapisać:

$$A \approx OF \quad (2.89)$$

gdzie: O jest macierzą o wymiarze $q \times K$, a macierz F opisuje ułamki wagowe dla poszczególnych komponentów w i -tej serii danych i ma wymiar $K \times N$. Łącząc równanie 2.88 i równanie 2.89 otrzymamy:

$$A \approx OF = U'(S'V'^T) \quad (2.90)$$

możemy również założyć, że za pomocą kombinacji wektorów własnych U' możemy, również zapisać macierz O . Ten, krok jest uzasadniony ponieważ macierz U opisuje serie danych, która jest kombinacją liniową kolumn macierzy O .

$$O = U'H \quad (2.91)$$

łącząc równania 2.90 i 2.91 możemy zapisać:

$$U'HF = U'(S'V'^T) \quad (2.92)$$

$$HF = S'V'^T \quad (2.93)$$

W celu zrekonstruowania macierzy O należy w jakiś sposób określić współczynniki macierzy H i F . W pracy (Williamson i wsp., 2008) autorzy pokazali, że optymalne wartości H i F w przypadku oligomeryzacji białka można znaleźć poprzez przeszukanie wartości stałych dysocjacji K_{dys} , za pomocą których da się wyliczyć macierz F (za pomocą prawa działania mas), a następnie wyznaczyć macierz H i zrekonstruować krzywe rozpraszania. W przypadku danych czasowo-rozdzielczych Kim i współpracownicy zastosowali podobne podejście lecz zamiast prawa działania mas zastosowali model kinetyczny (Kim i wsp., 2012).

W ich przypadku macierz V'^T można zapisać w postaci równania:

$$V'^T = (CP)^T \quad (2.94)$$

gdzie: C jest macierzą o wymiarach $n_t \times K$, gdzie n_t jest ilością punktów czasowych a K ilością zakładanych komponentów. Macierz C określa jak zmienia się stężenie komponentów K w poszczególnych punktach czasowych. Macierz P bezpośrednio wiąże macierz V z macierzą C . Macierz obliczana jest za pomocą pewnego modelu kinetycznego, dla którego należy zdefiniować reakcje (stałe czasowe reakcji k_{ij}) prowadzące do wzajemnych przemian komponentów K tj.:



Jest to analogiczne jak skonstruowanie macierzy F za pomocą stałych dysocjacji. Dla takiego modelu należy znaleźć kombinacje wartości macierzy P i C (określonej modelem kinetycznym), dla których różnica pomiędzy oryginalną macierzą danych A a macierzą zrekonstruowaną A' będzie jak najmniejsza. Mając macierz P można zrekonstruować krzywe dla K komponentów za pomocą równania:

$$K = U'S'P^T \quad (2.96)$$

W dwóch poprzednich przykładach zastosowania metody SVD potrzebny był model pozwalający na otrzymanie krzywych (np. model kinetyczny, stałe dysocjacji). Nie mniej jednak w wielu przypadkach zdefiniowanie takiego modelu jest trudne. W taki przypadku można wykorzystać metodę MCR-ALS (ang. multicurve resolution - alternating last squares) (Blobel i wsp., 2009, Minh i wsp., 2013, de Juan i wsp., 2014). Metoda ta służy w ogólności do modelowania danych chemometrycznych dotyczących np. profili reakcji chemicznych, w celu określenia liczby komponentów reakcji i ich profilu stężenia. Podobne podejście stosuje się też do danych spektroskopowych. Najważniejszą zaletą tej metody jest możliwość jej zastosowania bez uprzedniej wiedzy na temat badanego systemu (np. modelu kinetycznego). W metodzie tej można zapisać macierz danych D w postaci:

$$D = D^* + E \quad (2.97)$$

$$D^* = CS^T \quad (2.98)$$

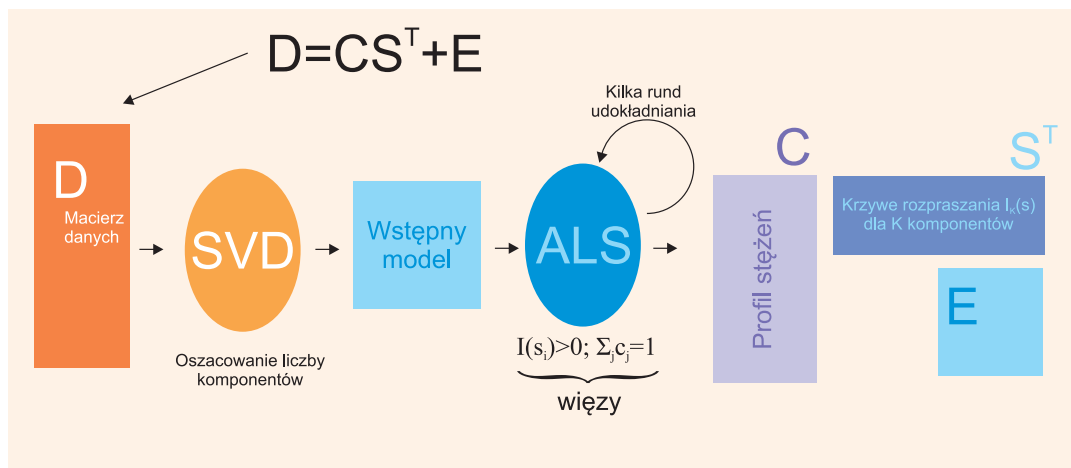
macierz E jest macierzą opisującą błąd metody, a D^* jest zrekonstruowaną macierzą danych. Macierze C i S odpowiadają odpowiednio: profilowi stężenia i krzywymi komponentów (np. dane ze eksperymentów SAXS, spektroskopii UV czy dichroizmu kołowego). W metodzie MCR-ALS za pomocą analizy SVD oszacowuje się liczbę komponentów, a później przybliżone rozwiązanie tj, macierze C i S , które następnie udokładnia się za pomocą algorytmu ALS (ang. alternating last-squares). Przybliżone rozwiązanie można otrzymać za pomocą kilku metod np. wybierając z zestawu krzywych “najczystsza” tj. tą która składa się tylko z jednego komponentu. Dla wstępnie określonej macierzy C można otrzymać wtedy macierz S za pomocą równania:

$$S^T = C^\dagger D_{SVD} \quad (2.99)$$

gdzie C^\dagger jest macierzą pseudo-odwrotną do C , D_{SVD} - jest odfiltrowaną macierzą D (macierz A' w równaniu 2.88). Do analizy można użyć także więzy np. unimodalny kształt krzywej czy też pozytywne wartości macierzy S^T (np. krzywa absorpcji UV czy krzywa SAXS) czy też, założyć że suma ułamków wagowych musi być równa 1. W kolejnym kroku nowa macierz C jest obliczana za pomocą równania:

$$C = (S^T)^\dagger D_{SVD} \quad (2.100)$$

Kroki algorytmu powtarza się aż otrzyma się zgodność macierzy D^* z macierzą D_{SVD} tj. macierz E jest takiego samego rzędu jak zakładany poziom błędu. Metodę MCR -ALS można podsumować za pomocą schematu przedstawionego na rysunku 2.19.



Rysunek 2.19 Schemat analizy MCR-ALS (na podstawie: Tauler 2005) .

Blobel i współpracownicy (2009) wykazali, że metoda ta może być skutecznie wykorzystana do modelowania danych SAXS dla mieszaniny kilku form oligomerycznych białka lmwPTP (ang. low molecular weight protein tyrosine phosphatase). Autorom udało się uzyskać krzywe rozpraszania dla

dimeru i monomeru białka lmwPTP i wygenerować niskorozdzielcze modele dla obu form oligomerycznych.

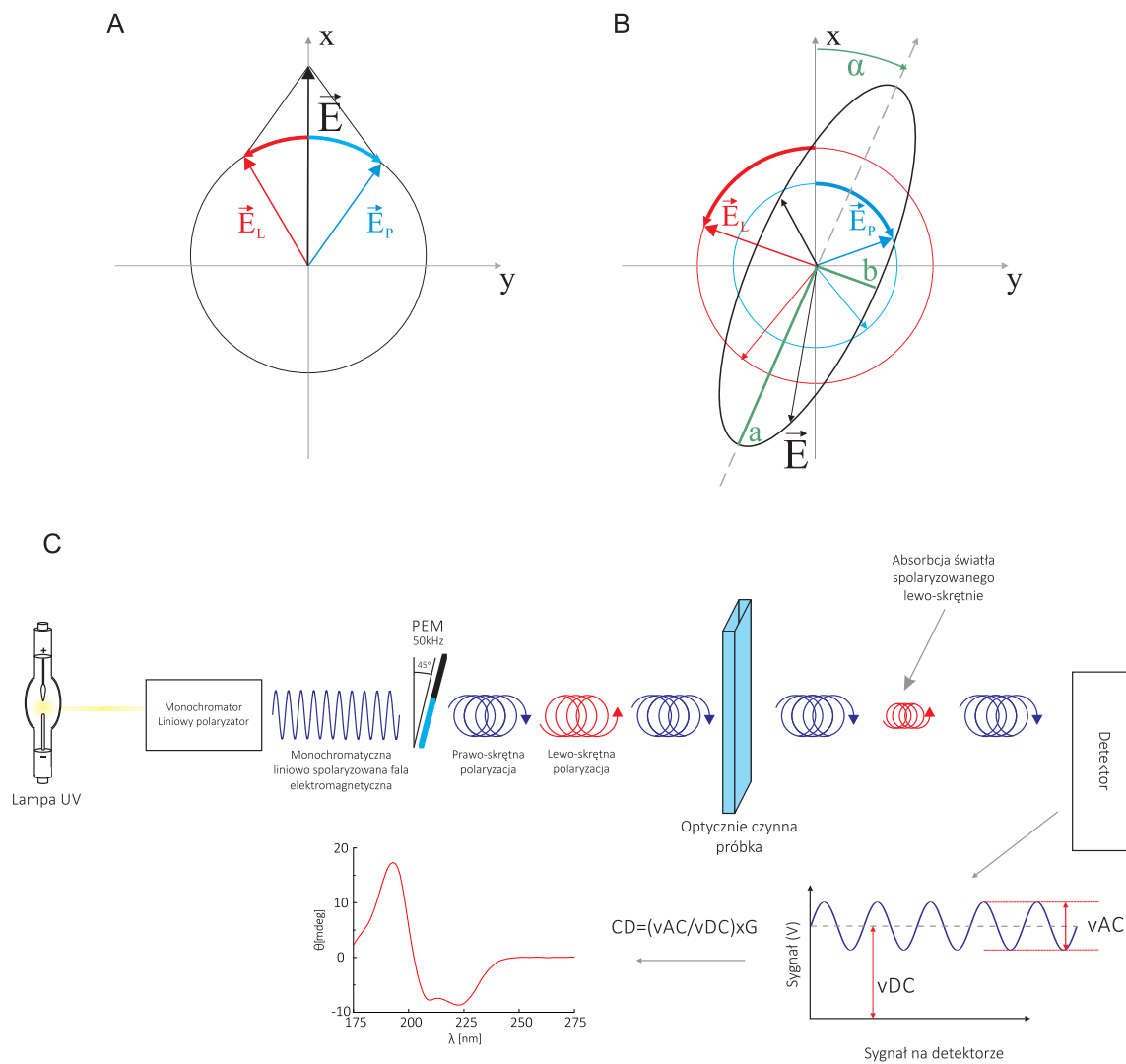
Podsumowując metody SVD i MCR-ALS stanowią doskonałe narzędzie analizy danych małokątowego rozpraszania promieniowania rentgenowskiego dla mieszanin w celu określenia ilości komponentów i stężenia a co najważniejsze także krzywych rozpraszania. Należy jednak mieć na uwadze, że założenia jakie stosujemy do modelowania (np. liczba komponentów, model kinetyczny itp.) mogą mieć istotny wpływ na otrzymane wyniki. To samo dotyczy także jakości danych eksperymentalnych, ponieważ szum w danych wprowadza błąd do zrekonstruowanych krzywych rozpraszania, a co najmniej redukuje zakres krzywej SAXS możliwej do rekonstrukcji.

2.7 Dichroizm kołowy

2.7.1 Zjawisko dichroizmu kołowego

Zjawisko dichroizmu kołowego jest jednym z przykładów istnienia aktywności optycznej. Aktywność optyczna ma miejsce gdy, cząsteczka zawierająca chromofor jest chiralna lub gdy niechiralny chromofor znajdzie się w chiralnym otoczeniu. Aktywność optyczna manifestuje się w postaci: zmiany polaryzacji z liniowej na eliptyczną podczas przechodzenia światła przez próbkę, dichroizmu kołowego - różny stopień absorpcji światła lewoskrętnie i prawoskrętnie spolaryzowanego, zjawiska skręcania polaryzacji światła, różnego współczynnika załamania dla różnej skrętności padającego światła. Schematycznie efekt dichroizmu kołowego oraz samego eksperymentu przedstawiony jest na rysunku 2.20.

Spektrometr dichroizmu kołowego przedstawiony jest schematycznie na rysunku 2.20 C. Promieniowanie generowane jest przez lampę UV jest kierowane na monochromator i polaryzator. W celu otrzymania promieniowania o polaryzacji kołowej, na drodze optycznej ustawiony jest modulator fotoelastyczny (ang. photoelastic modulator - PEM). Najczęściej w spektrometrach dichroizmu kołowego wykorzystuje się element piezoelektryczny drgający z częstotliwością ok. 50 kHz połączony z blokiem stopionego kwarcu. Po przyłożeniu naprężenia układ staje się dwójłomny. PEM działa jak ćwierćfalówka powodując przesunięcie w fazie składowej horyzontalnej względem wertykalnej pola elektrycznego o $\pi/4$ i na odwrót, zgodnie z kierunkiem przyłożonego naprężenia. Także, po przejściu przez PEM, światło ma polaryzację kołową lewo- lub prawo-skrętną. Skrętność zmienia się wraz z częstotliwością drgań elementu piezoelektrycznego. Następnie wiązka kierowana jest na próbkę. Jeżeli próbka absorbuje światło niezależnie od rodzaju skręcenia polaryzacji to sygnał obserwowany na detektorze jest stały w czasie. W przypadku, gdy współczynniki ekstynkcji różnią się w zależności od rodzaju skręcenia polaryzacji, na detektorze sygnał będzie oscylował z częstością 50 kHz. vAC jest różnicą w absorpcji pomiędzy rodzajami polaryzacji kołowej. Sygnał na detektorze jest skalowany do średniej intensywności padającego światła (vDC) i przemnożony przez czynniki G , zapewniający odpowiednie jednostki (np. ΔA lub θ) sygnału dichroizmu kołowego.



Rysunek 2.20 Zjawisko dichroizmu kołowego. **A.** Przed przejściem przez ośrodek optycznie czynny falę elektromagnetyczną liniowo spolaryzowaną można rozpatrywać jako złożenie dwóch fal: o polaryzacji kołowej lewo- i prawo-skrętnej: $\vec{E} = \vec{E}_L + \vec{E}_P$. **B.** W wyniku różnicy współczynników ekstynkcji dla składowych kołowych, fala elektromagnetyczna po przejściu przez próbkę ma polaryzację eliptyczną. Na rycinie **C.** kolorem zielonym zaznaczone są półosie elipsy oraz kąt skręcenia α . **C.** Schemat eksperymentu dichroizmu kołowego (opis w tekście poniżej) (rysunek zaadaptowano ze strony internetowej <http://www.photophysics.com/tutorials/circular-dichroism-cd-spectroscopy/4-cd-spectrometer-operating-principles>) .

Jak wspomniano wyżej, w eksperymencie dichroizmu kołowego na próbkę na zmianę kierowane jest światło spolaryzowane kołowo lewoskrętnie lub prawoskrętnie i obliczana jest różnica we współczynnikach ekstynkcji:

$$\Delta\epsilon = \epsilon_L - \epsilon_P \quad (2.101)$$

gdzie: ϵ_L , ϵ_P są współczynnikami ekstynkcji dla światła lewo - prawo-skrętnego. Inną miarą efektu dichroizmu kołowego jest eliptyczność zdefiniowana jako:

$$\theta = \arctan \frac{a}{b} \quad (2.102)$$

gdzie: a i b są odpowiednio długością wielkiej i małej półosi elipsy światła spolaryzowanego eliptycznie (patrz rysunek 2.20 B). Obie te wielkości można wyrazić za pomocą jednego wzoru:

$$\theta[^\circ] = 2.303 (A_L - A_R) \frac{180}{4\pi} \quad (2.103)$$

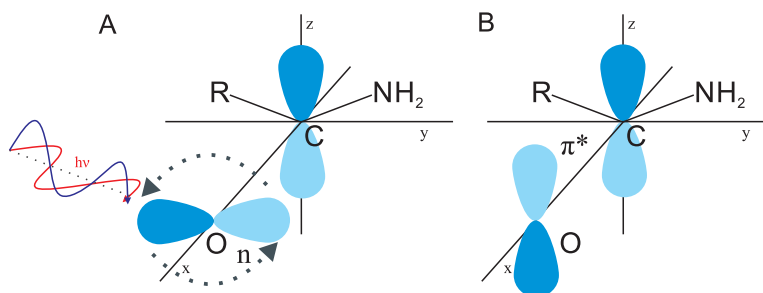
gdzie: A_L , A_P są wartościami składowej lewo- i prawo-skrętnej absorpcji związanymi ze wzorem 2.101 za pomocą prawa Lamberta-Beera.

Zjawisko dichroizmu kołowego ma miejsce, gdy oprócz dipolowego momentu przejścia będzie występował magnetyczny moment przejścia (Cantor and Schimmel 1998, Parson 2009). Poprzez kolejną analogię z fizyką klasyczną moment dipolowy będzie indukowany, gdy nastąpi rotacja ładunku elektrycznego podczas przejścia ze stanu podstawowego do stanu wzbudzonego. Rotacja może być lewo lub prawo-skrętna. Samo zjawisko polega na różnej absorpcji światła spolaryzowanego lewo i prawoskrętnie. Światło liniowo spolaryzowane możemy rozważać jako superpozycję dwóch składowych o polaryzacji kołowej i tej samej amplitudzie ale przeciwnej skrętności. W wyniku zajścia efektu dichroizmu kołowego jedna ze składowych jest pochłaniana silniej i światło przechodząc przez próbkę zmienia swoją polaryzację z liniowej na eliptyczną. Miarą dichroizmu kołowego jest więc eliptyczność. Z teoretycznego punktu widzenia wielkością opisującą efekt dichroizmu kołowego jest siła skręcenia \mathfrak{R} , którą można obliczyć za pomocą wzoru wyprowadzonego przez Rosenfelda (Rosenfeld 1928):

$$\mathfrak{R} = Im \langle \psi_a | \vec{\mu} | \psi_0 \rangle \langle \psi_0 | \vec{m} | \psi_a \rangle \quad (2.104)$$

gdzie \vec{m} jest operatorem momentu magnetycznego a wielkość $\langle \psi_0 | \vec{m} | \psi_a \rangle$ jest nazywana magnetycznym momentem przejścia i jest odpowiednikiem dipolowego momentu przejścia. Żeby wystąpiło zjawisko dichroizmu kołowego podczas przejścia muszą wystąpić niezerowe momenty przejścia oraz nie mogą być one do siebie prostopadłe. Dichroizm kołowy powstaje, gdy w hamiltonianie oddziaływania układu z falą elektromagnetyczną uwzględni się także wkład od oddziaływania z polem magnetycznym. W przypadku przejść elektronowych ich absorpcja może być mała natomiast efekt dichroizmu kołowego bardzo silny. Przykładem przejścia elektronowego mogącego dać efekt dichroizmu kołowego jest przejście $n \rightarrow \pi^*$ wolnej pary elektronowej atomu tlenu (orbital $2p_y$) w wiązaniu karbonylowym. W tym wypadku następuje obrót orbitalu okupowanego przez wolną parę elektronową o kąt $\frac{\pi}{2}$ ponieważ orbital molekularny π^* jest asymetryczną superpozycją orbitali $2p_z$ atomów węgla i tlenu: $\frac{1}{\sqrt{2}} (2p_z^C - 2p_z^O)$. Magnetyczny moment przejścia ma niezerową wartość w przeciwieństwie do dipolowego momentu przejścia. Przejście to jest dozwolone jeżeli wiązanie to znajdzie się w otoczeniu zaburzającym symetrię orbitali elektronowych. Zaburzenie to jest źródłem niezerowego dipolowego momentu przejścia. Przejście to schematycznie jest przedstawione na rysunku 2.21.

Dla przejścia $n \rightarrow \pi^*$ w wiązaniu karbonylowym dipolowy moment przejścia $\langle \psi_{\pi^*} | \vec{\mu} | \psi_n \rangle$ jest równy zero. Tak więc w prostym przybliżeniu takie przejście jest zabronione, jednak w eksperymencie często da się je zarejestrować ponieważ w obliczeniach nie uwzględnia się otoczenia oraz drgań molekularnych. Nie mniej jednak absorpcja promieniowania dla tego przejścia jest dużo słabsza niż w przypadku przejścia $\pi \rightarrow \pi^*$. Natomiast magnetyczny moment przejścia jest różny od zera i dlatego też to przejście jest dużo bardziej widoczne na widmie dichroizmu kołowego.



Rysunek 2.21 Przejście elektronowe $n \rightarrow \pi^*$ w wiązaniu karbonylowym. Wiązanie karbonylowe znajduje się w płaszczyźnie xy . **A.** Orbitale n atomu tlenu i $2p_z$ atomu węgla są do siebie symetryczne przed absorpcją kwantu promieniowania. W wyniku absorpcji następuje rotacja orbitalu n . **B.** Po rotacji orbital $2p_z$ atomu tlenu tworzy antysymetryczną kombinację liniową z orbitalem $2p_z$ atomu węgla. Rysunek wykonano na podstawie (William W. Parson, Modern Optical Spectroscopy Springer-Verlag 2009, pp 312.)

Eksperymentalnie mierzoną eliptyczność dla przejścia elektronowego $0 \rightarrow a$ można związać ze wzorem na siłę skręcenia za pomocą wzoru:

$$\mathfrak{R}_{a0} = \left(\frac{3hc}{8\pi^3 N_0} \right) \int \left\{ \frac{\theta(\lambda)}{\lambda} \right\} d\lambda \quad (2.105)$$

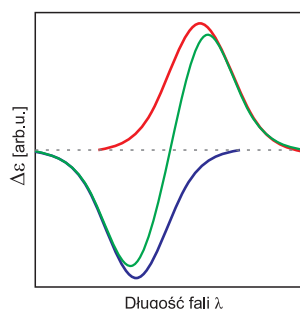
Wzór ten stanowi więc podstawę do porównywania danych eksperymetalnych z modelem teoretycznym. Dla nieoddziałujących dipoli sygnał dichroizmu kołowego ma taki sam kształt jak widmo absorpcji ale może przyjmować zarówno dodatni jak i ujemny znak. Dla cząsteczek będących stereoizomerami sygnał dichroizmu będzie miał taki sam kształt ale przeciwny znak. W przypadku gdy mamy do czynienia z dwoma oddziałującymi dipolami Hamiltonian dla takiego układu będzie następującej postaci:

$$\hat{H}'_{12} = \hat{H}'_1 + \hat{H}'_2 + \hat{V}_{12} \quad (2.106)$$

$$\hat{V}_{12} = \frac{\vec{\mu}_1 \vec{\mu}_2}{r_{12}^3} - 3 \frac{(\vec{\mu}_1 \vec{r}_{12})(\vec{r}_{12} \vec{\mu}_2)}{r_{12}^5} \quad (2.107)$$

gdzie: \hat{V}_{12} - jest operatorem oddziaływania pomiędzy dwoma dipolami odległymi o wektor \vec{r}_{12} , $\vec{\mu}_1$ i $\vec{\mu}_2$ są operatorami momentu dipolowego jednego i drugiego monomeru. Dodatkowy człon w Hamiltonianie będzie prowadził do rozszczepienia poziomu energetycznego: $E_{0a} \pm V_{12}$ i obserwacji dwóch

zamiast jednego pasma absorpcji dla przejścia $0 \rightarrow a$. Rozszczepienie to nazywamy rozszczepieniem ekscytonowym. W przypadku dichroizmu kołowego występuje jeszcze dodatkowe sprzężenie magnetyczne wywołane przez oddziaływanie momentów magnetycznych przejścia. Także siła rotacji będzie składała się z trzech komponentów: członu jedno-elektronowego, który jest sumą \mathfrak{R}_{0a} dla monomerów, członu sprzężenia elektromagnetycznego, który pochodzi od oddziaływania pomiędzy momentem magnetycznym jednego monomeru a momentem dipolowym drugiego oraz trzeciego członu nazywanego członem sprzężenia ekscytonowego. Trzeci człon jest związany z oddziaływaniem momentów dipolowych pomiędzy monomerami. Człon ten bardzo często jest członem dominującym. Przykładowy sygnał dichroizmu kołowego dla oddziałujących monomerów ma postać widoczną na rysunku 2.22. Na wykresie można zaobserwować dwa pasma o przeciwnym znaku położone symetrycznie wokół pasma absorpcji dla monomeru.



Rysunek 2.22 Schemat przykładowego widma dichroizmu kołowego dla dimeru cząsteczek. Rysunek wykonano na podstawie (William W. Parson, *Modern Optical Spectroscopy* Springer-Verlag 2009, pp 326). Kolorem czerwonym i niebieskim zaznaczone są widma dichroizmu kołowego dla indywidualnych przejść. Kolorem zielonym zaznaczony jest całkowity sygnał dichroizmu kołowego.

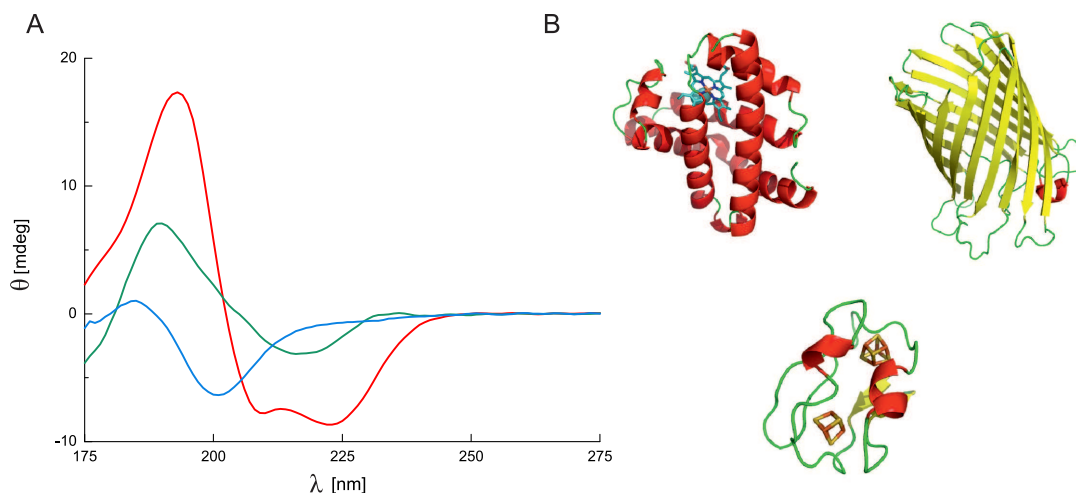
2.7.2 Wykorzystanie spektroskopii dichroizmu kołowego do wyznaczania struktury drugorzędowej białek

W przypadku białek dwa przejścia elektronowe dają wkład do obserwowanego widma dichroizmu kołowego. Przejściami tymi są przejścia $\pi \rightarrow \pi^*$ oraz $n \rightarrow \pi$ w wiązaniu peptydowym. Białka są polimerami składającymi się z wielu monomerów (aminokwasów) i w przypadku tych przejść dochodzi do sprzężenia ekscytonowego i występowania charakterystycznego sygnału dichroizmu kołowego dla różnych konformacji łańcucha polipeptydowego. W przypadku białek, widma dichroizmu kołowego dla poszczególnych struktur drugorzędowych można wyliczyć teoretycznie (Tinoco i wsp., 1963, Woody i wsp., 1967, Sreerama i wsp., 2004). Dla helisy alfa przejścia elektronowe $\pi \rightarrow \pi^*$ dają wkład w postaci minimum przy ok. 205 nm, oraz dwa pasma o pozytywnym sygnale poniżej pasma 205 nm (przeważnie w pobliżu 190 nm). Przejście przy 205 nm jest spolaryzowane równoległe do osi helisy a dwa przejścia wysokoenergetyczne są spolaryzowane prostopadle do osi helisy. Przejście $n \rightarrow \pi$ jest zabronione, ponieważ moment dipolowy jest równy 0. Przejście to może zostać wzbudzone poprzez sprzężenie ekscytonowe z wysokoenergetycznymi przejściami $\pi \rightarrow \pi^*$ dając w rezultacie dodatkowe pasmo o negatywnym sygnale przy długości fali ok. 220 nm. W rezultacie spektrum składa się z

trzech pików dodatniego przy długości fali ok. 190 nm oraz dwóch negatywnych przy długości fali 205 nm i 220 nm. Model ten wyjątkowo dobrze zgadza się z wynikami eksperymentalnymi otrzymanymi dla modelowego peptydu o konformacji helisy alfa.

Dla struktury arkusza beta spektrum dichroizmu kołowego występują pasma o dodatnim sygnale ok 200 nm oraz negatywnym sygnale dla krótszych długości fali pochodzące od sprzężenia ekscytonowego przejścia $\pi \rightarrow \pi^*$ (Tinoco i wsp., 1963, Woody i wsp., 1967, Sreerama i wsp., 2004). Sprzężenie przejścia $\pi \rightarrow \pi^*$ z przejściem $n \rightarrow \pi^*$ jest rezultatem pasma o negatywnym sygnale przy ok. 215 nm. W odróżnieniu jednak od wyników otrzymanych dla helisy alfa przewidywania dla arkuszy beta mniej zgadzają się z eksperymentalnie otrzymanymi widmami dla modelowych układów. Najprawdopodobniej jest to związane z wrażliwością widma na skręcenie β -arkusza.

Z eksperymentów nad zwijaniem się białek wynika, że konformacja kłęбка statystycznego ma inny sygnał dichroizmu kołowego. Dodatni sygnał dichroizmu kołowego występuje poniżej 210 nm a negatywny poniżej 210 nm. Teoretycznie za pomocą kombinacji liniowej widma dla helisy alfa, arkusza beta oraz struktury niepofałdowanej (kłęбка statystycznego) można określić ich procentowy wkład do ogólnej struktury drugorzędowej białka. Przykładowe widma dichroizmu kołowego dla wybranych białek przedstawione są na rysunku 2.23.



Rysunek 2.23 Widmo dichroizmu kołowego dla białek. **A.** Eksperymentalne widma dichroizmu kołowego dla trzech białek o różnej zawartości elementów struktury drugorzędowej: mioglobiny z kaszalota (kolor czerwony) - białko bogate w helisy alfa, poryny OmpG z bakterii *E.coli* (kolor zielony) - białko bogate w arkusze beta, ferredoksyny z bakterii *Clostridium acidurici* - białka z licznymi fragmentami o nieregularnej strukturze i pętłach (kolor niebieski). **B.** Struktury przestrzenne wymienionych białek, w kolejności: mioglobiny, poryny OmpG oraz ferredoksyny. Dane CD pochodzą z bazy danych Protein Circular Dichroism Data Bank (<http://pcddb.cryst.bbk.ac.uk/>).

W przypadku białek spektroskopia dichroizmu kołowego znalazła zastosowanie w określaniu struktury drugorzędowej. W tym celu zarejestrowane widmo dla badanego białka rozkłada się w postaci

kombinacji liniowej krzywych bazowych, które w najprostszym przypadku są widmami dla modelowych białek o strukturze helisy alfa, arkusza beta i kłębka statystycznego. Nowsze wersje algorytmów uwzględniają także inne struktury drugorzędowe jak np. helisę 3_{10} czy też skręty beta. W założeniu dla danej długości fali λ eliptyczność θ_λ jest równa kombinacji liniowej eliptyczności dla poszczególnych struktur drugorzędowych:

$$\theta_\lambda = \sum_i \varepsilon_i S_{i\lambda} + szum \quad (2.108)$$

$$\sum_i \varepsilon_i = 1 \quad (2.109)$$

$$\varepsilon_i \geq 0 \quad (2.110)$$

Nie mniej jednak takie podejście ma swoje wady (Compton i wsp., 1986). Po pierwsze, najpierw należy określić jak wyglądają widma dla poszczególnych struktur drugorzędowych. Jak wyżej wspomniano teoretyczne przewidywanie widm CD dla białek nie zawsze jest dokładne. Widma dla modelowych układów, takich jak peptydy o określonej strukturze drugorzędowej, też mają ograniczone zastosowanie. Wykazano, że nawet dla modelowych układów różne parametry takie jak np. długość helisy alfa czy też skręcenia w arkuszach beta mają wpływ na kształt widma. Dlatego też zaproponowano inne rozwiązanie. Hennessey i współpracownicy (Hennessey i wsp., 1981) wykazali, że dla zbioru widm CD dla 16 modelowych białek, dla których znana była struktura przestrzenna, istnieje pewna ilość informacji wystarczająca do rekonstrukcji widma oraz do określenia wkładu poszczególnych struktur drugorzędowych. Autorzy posłużyli się analizą SVD wspomnianą wcześniej przy analizie danych SAXS. Wykazali, że dla macierzy wektorów własnych pięć pierwszych wektorów wystarczy, by zrekonstruować widma i określić strukturę drugorzędową dla zbioru 16 modelowych białek. Dla zbioru modelowych białek można zdefiniować macierz F , która opisuje ilość struktur drugorzędowych w widmie tj:

$$C = BF \quad (2.111)$$

$$F = XC \quad (2.112)$$

gdzie: C jest macierzą zawierającą dane CD, B - macierzą zawierającą widma dla wyodrębnionych struktur drugorzędowych, macierz X - zawiera tzw. odwrotne widma CD dla struktur drugorzędowych. Mając zdefiniowane macierze: eksperymentalnie wyznaczoną macierz C oraz obliczoną na podstawie zdeponowanych w bazie danych struktur macierz F można za pomocą metody najmniejszych kwadratów wyznaczyć macierz X . Za pomocą macierzy X następnie można określić wkład poszczególnych struktur drugorzędowych dla innych białek tj. innej macierzy C' . Znając macierz F można także zrekonstruować macierz B widm dla wyodrębnionych struktur drugorzędowych za pomocą metody SVD rozkładając najpierw macierz F (Compton i wsp., 1986):

$$F = U_F S_F V_F^T \quad (2.113)$$

a następnie obliczyć wartość B ze wzoru:

$$B = C (V_F S_F^+ U_F^T) \quad (2.114)$$

Macierz V_F jest transponowaną macierzą V_F^T , macierz U_F^T jest transponowaną macierzą U_F , a macierz S_F^+ jest macierzą pseudo odwrotną (Moora-Penrosa) macierzy S_F . Jednak w przypadku metody najmniejszych kwadratów, gdy liczba widm jest większa niż liczba informacji dostępnych w danych, rozwiązanie może być niestabilne. Zaproponowano więc inne podjęcie z zastosowaniem metody SVD oraz metody samouzgodnionej (Sreerama i wsp., 1993, Johnson 1999). W metodzie tej tak jak poprzednio definiowana jest macierz F . Jednak w odróżnieniu od poprzedniej metody w skład danych macierzy widm bazowych C wchodzi również widmo dla białka o nieznannej strukturze drugorzędowej tworząc macierz C_2 . Analogicznie tworzona jest macierz F_2 , która zawiera także wkład od nieznanego białka f . Ponieważ f jest szukaną przez nas funkcją, zakłada się wartość początkową f i oblicza się macierz X za pomocą metody SVD:

$$X = F_2 V S^+ U^T \quad (2.115)$$

Macierze U , S , V^T oraz V , S^+ , U^T pochodzą z rozkładu macierzy C za pomocą metody SVD: $C = U S V^T$. W analizie SVD wykorzystuje się, tak jak wspomniano poprzednio, tylko 5 pierwszych wektorów i wartości własnych. Z pomocą tak otrzymanej macierz X oblicza się wartość f_1 :

$$f_1 = X C' \quad (2.116)$$

W następnym kroku obliczeń wartość f_1 zastępuje f w macierzy F_2 . Korki algorytmu powtarzane są aż poprawa w wartości f nie ulegnie zmianie tj. gdy różnica pomiędzy kolejnymi wartościami f będzie mniejsza niż 0,025. W trakcie pracy algorytmu dodatkowo zmieniane są widma w macierzy bazowej C tak, aby rozwiązanie było jak najlepsze. Zabieg ten pozwala również na uniknięcie problemu związanego z tym, że ponieważ ilość informacji w macierzy C jest ograniczona do pięciu, jedynie pięć równań (z analizy SVD) musi wystarczyć do opisanie struktury drugorzędowej. Niestety liczba różnych struktur drugorzędowych jakie można wyróżnić jest większa niż pięć jak również ich udział w poszczególnych białkach z macierzy C jest różny. Autorzy w pracy wykazali, że jeżeli zmniejszy się ilość białek z macierzy C do ośmiu otrzymane wyniki będą jeszcze bardziej dokładne. Oczywiście wybór ośmiu białek z 16 nie jest prosty i jednoznaczny. Dlatego w programie osiem składowych bazowych jest losowo dobierana i rozwiązania, dla których $0,952 \leq \sum_i f_i \leq 1,05$ oraz $f_i \geq -0,03$ są zatrzymywane. Dodatkowym, kryterium jest warunek, że średniokwadratowy błąd pomiędzy oryginalnym widmem a widmem zrekonstruowanym wynosi mniej niż $0,25 \Delta\epsilon$. Metoda opisana powyżej i zaimplementowana w programie CDSSTR jest w stanie oszacować wartość struktury drugorzędowej w widmie z błędem mniejszym niż 4 %.

Istnieje kilka programów do analizy zawartości struktur drugorzędowych białek na podstawie danych ze spektroskopii dichroizmu kołowego. Różnią się one między sobą stosowaną metodą obliczeń, choć programy VARSLC i CDSSTR wykorzystują podobną technikę obliczeń (dobieranie krzywych referencyjnych). Również CDSSTR oraz SELCON3 dzielą część metod obliczeniowych (metoda samouzgodniona). Program K2D wykorzystuje metodę opartą na sieciach neuronowych. Natomiast CONTINLL wykorzystuje połączenie metody najmniejszych kwadratów z dodatkowymi więzami statystycznymi zapewniającymi stabilność rozwiązania. Każdy z tych programów wymaga określonego minimalnego zakresu długości fali w celu przeprowadzenia analizy (tabela 2.1). Program K2D potrzebuje do analizy najmniejszy zakres długość fali. Istnieje też możliwość wyboru zestawu referencyjnych

widm. Opis każdego z zestawów dostępny jest na stronie internetowej (http://dichroweb.cryst.bbk.ac.uk/html/userguide_datasets.shtml). Wszystkie wymienione w tabeli 2.1 programy są dostępne za pomocą serwera Dichroweb (Whitmore i wsp., 2004).

Tabela 2.1 Wybrane programy do estymacji struktur drugorzędowych na podstawie danych ze spektroskopii dichroizmu kołowego.

Program	Minimalny zakres długości fali	Referencje
K2D	200-241 nm	Andrade i wsp., 1993
CDSSTR	190-260 nm	Johnson 1999
SELCON3	190-240 nm	Sreerema i wsp., 1993
CONTINLL	178-240 nm	Van Stokkum i wsp., 1990
VARSLC	178-260 nm	Compton i wsp., 1986, Manavalan i wsp., 1987

2.8 Źródła promieniowania rentgenowskiego

W rozdziale tym chciałbym przedstawić źródła promieniowania rentgenowskiego jakie stosowane są w badaniach SAXS. Informacje te dopełniają wcześniej opisaną metodykę analizy danych SAXS od strony eksperymentalnej. Historycznie pierwszym źródłem promieniowania rentgenowskiego była wykorzystana w pionierskich eksperymentach przez Wilhelma Konrada Röntgena tzw. lampa wyładowcza Crookes'a. Zmodyfikowana wersja lampy wyładowczej Crookes'a ale opierająca się na tych samych zasadach używana jest do dzisiaj jako laboratoryjne źródło promieniowania rentgenowskiego. Obecnie standardowym źródłem promieniowania w zakresie rentgenowskim jest synchrotron. Poniżej, zostanie krótko opisane wytwarzanie promieniowania w synchrotronach.

2.8.1 Promieniowanie synchrotronowe

Z fizyki klasycznej wiemy, że ładunek mający przyspieszenie będzie źródłem fali elektromagnetycznej. W przypadku cząsteczki poruszającej się po okręgu w wyniku działania np. siły Lorentza cząsteczka będzie miała przyspieszenie radialne. Dla cząstki naładowanej poruszającej się po okręgu zależność kątowa wypromieniowanej energii w układzie odniesienia związanym z poruszającym się ładunkiem (*) ma postać (Wiedemann 2003):

$$\frac{dP}{d\Omega} = \frac{r_c m c}{4\pi} a_*^2 \sin^2 \Theta^* \quad (2.117)$$

gdzie: Θ^* - jest kątem pomiędzy kierunkiem obserwacji a wektorem przyspieszenia \vec{a}^* , $r_c = \frac{1}{4\pi\epsilon_0} \frac{e^2}{m c^2}$ jest tak zwanym klasycznym promieniem elektronu. W przypadku cząsteczki poruszającej się z prędkością zbliżoną do prędkości światła przechodząc do układu laboratoryjnego związanego z obserwatorem musimy zastosować transformację Lorentza. Transformacja ta powoduje zniekształcenie pola elektrycznego związane ze skończoną prędkością światła. W wyniku tego zniekształcenia promieniowanie emitowane przez cząstkę jest silnie skolimowane w kierunku stycznym do toru ruchu cząstki. Rozkład kątowy promieniowania można opisać za pomocą kąta Θ :

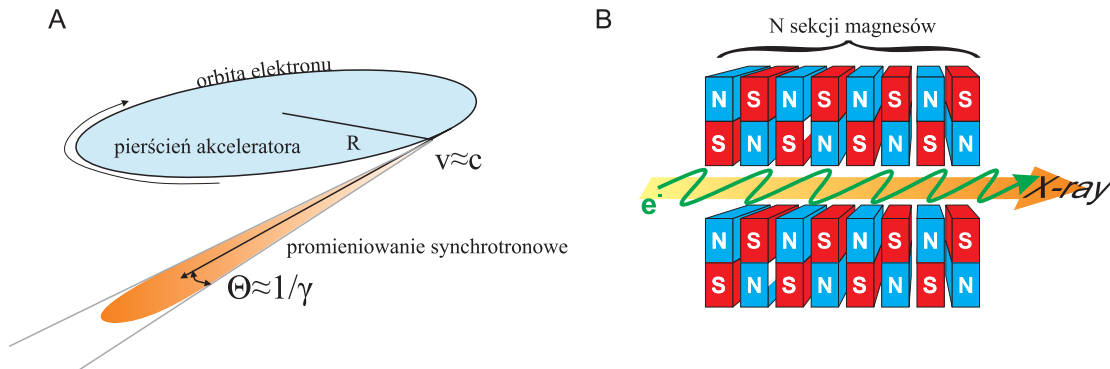
$$\Theta \approx \frac{\sin \Theta^*}{\gamma(1 + \beta \cos \Theta^*)} \approx \frac{1}{\gamma} \quad (2.118)$$

Gdzie: $\gamma = \sqrt{1 - \beta^2}$, $\beta = \frac{v}{c}$. Dla relatywistycznych elektronów poruszających się w akceleratorach cząstek $\gamma \sim 10^3 - 10^4$ a Θ jest rzędu miliradianów. Innymi słowy promieniowanie emitowane przez cząstkę ma kąt rozwarcia Θ . Promieniowanie synchrotronowe jest spolaryzowane i ma bardzo dużą intensywność w porównaniu z konwencjonalnymi źródłami promieniowania. Można wykazać, że liczba fotonów wypromieniowania przez elektrony jest proporcjonalna do stałej struktury subtelnej α :

$$N_{rad} \sim \frac{1}{4\pi} \alpha \quad (2.119)$$

Dla prądu elektronów rzędu 1A otrzymamy wartość rzędu 10^{17} (na jedno przejście przez pierścień akumulacyjny synchrotronu).

Jeśli chodzi o wykorzystanie promieniowania synchrotronowego (nazwa pochodzi od typu akceleratora cząstek elementarnych, w którym zarejestrowano po raz pierwszy eksperymentalnie promieniowanie emitowane przez naładowane cząstki w ruchu po okręgu) to w pierwszych latach po odkryciu traktowano je jedynie jako promieniowanie pasożytnicze, które ogranicza możliwości akceleratorów cząstek. Jednak bardzo szybko zreflektowano się, że intensywne promieniowanie otrzymywane w akceleratorach cząstek, może być wykorzystane w wielu dziedzinach. Z biegiem czasu zaczęto budować akceleratory cząstek, które służyły tylko do otrzymywania promieniowania synchrotronowego a obecnie na świecie pracuje kilkadziesiąt synchrotronów (pełna lista na stronie internetowej: <http://www.lightsources.org/>).

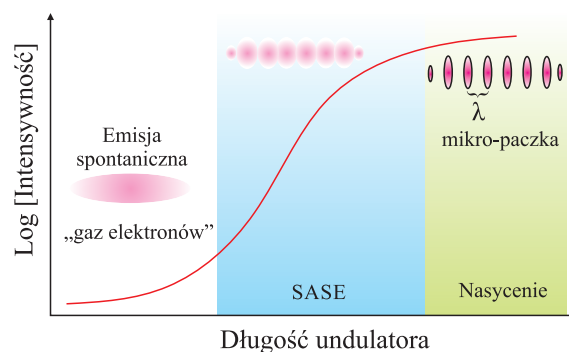


Rysunek 2.24 A. Schemat powstawania promieniowania synchrotronowego w pierścieniu akumulacyjnym dla relatywistycznych elektronów. **B.** Schemat powstawania promieniowania synchrotronowego w urządzeniach typu *insertion device*.

Na rysunku 2.24 przedstawiony jest schematycznie proces powstawania promieniowania synchrotronowego. W nowoczesnych synchrotronach pierścień akumulacyjny składa się z sekcji prostych oraz sekcji zakrzywiających tor ruchu. Sekcje zakrzywiające tor ruchu są to tzw. magnesy zakrzywiające (ang. bending magnet). Magnesy zakrzywiające generują promieniowanie synchrotronowe na zasadzie opisanej powyżej. Efektywniejszym sposobem na otrzymywanie promieniowania synchrotronowego jest wykorzystanie urządzeń wstawkowych: tzw. wigglerów i undulatorów (ang. wiggler, undulator). Oba te urządzenia składają się z kilkunastu dipoli magnetycznych ułożonych jeden za

drugim ale naprzemiennie ułożonych biegunami. Takie ułożenie powoduje, że elektron porusza się w płaszczyźnie horyzontalnej ruchem oscylującym. Zmiana toru ruchu elektronu związana jest z pojawieniem się przyspieszenia, które jest źródłem promieniowania co schematycznie pokazane jest na rysunku 2.24 B. W przypadku undulatorów magnesy są ułożone w taki sposób, że promieniowanie powstałe podczas jednego okresu oscylacji jest w fazie z promieniowaniem powstałym podczas kolejnych oscylacji. Teoretycznie tak powstałe promieniowanie jest monochromatyczne, jednak ze względu na skończoną liczbę magnesów jest ono jedynie quasi-monochromatyczne. W przypadku wigglera oscylacje elektronu są dużo większe a emitowane promieniowanie nie jest już koherentne. Tor ruchu w wigglerze może zostać rozłożony na $2N$ (N liczba sekcji magnesów) okresów, gdzie elektron porusza się po torze półkolistym. W rezultacie wiggler będzie generował promieniowanie $2N$ razy bardziej intensywne niż magnes zakrzywiający o tej samej wartości indukcji pola magnetycznego.

Undulatory stanowią techniczną podstawę do budowy lasera na swobodnych elektronach. Gdy elektrony poruszają się w undulatorze każdy elektron będzie emitował promieniowanie niezależnie od innych (elektrony będą zachowywały się jak gaz elektronowy). Jednak im dłużej paczka elektronów (o odpowiedniej gęstości) będzie podlegała oscylacjom w undulatorze tym promieniowanie będzie bardziej koherentne (rysunek 2.25). Jeśli jednak uda się w jakiś sposób zmusić elektrony do uporządkowania się w taki sposób, żeby paczka falowa podzielona była na fragmenty odległe od siebie o długość generowanego promieniowania wtedy każdy fragment będzie można traktować (w skali długości fali generowanego promieniowania) jak ładunek punktowy o ładunku równym eN (N liczba elektronów). Promieniowanie z tak podzielonej paczki falowej byłoby koherentne. Elektrony w undulatorze po pewnym czasie zaczynają odczuwać wpływ generowanego promieniowania o pewnej długości fali. Oddziaływanie to powoduje grupowanie się elektronów w przestrzeni z periodem modulacji równym właśnie długości fali promieniowania. Gdy znajdzie już ten proces więcej promieniowania jest generowane i modulacja przestrzenna jest jeszcze bardziej efektywna. Proces ten nosi nazwę SASE (ang. self amplified stimulated emission). Wymagana szerokość paczki elektronów to $\sim 30 \mu\text{m}$ (szerokość połówkowa) i czasie 0,1 ps pomiędzy kolejnymi paczkami. Lasery na swobodnych elektronach dostarczają promieniowania o ogromnej intensywności i koherencji. Dzięki tym parametrom pozwalają na zarejestrowanie np. dyfrakcji na pojedynczych cząsteczkach jak również na badanie zjawisk fizycznych o bardzo szybkiej dynamice rzędu femtosekund.



Rysunek 2.25 Schematyczny rysunek przedstawiający zależność mocy promieniowania od długości drogi przebytej w undulatorze przez elektrony. Na wykresie zaznaczony jest też obszar gdzie

występuje efekt SASE jak również gdzie elektrony tworzą mikro-paczki oddalone od siebie o długość fali promieniowania. (Rysunek wykonano na podstawie Nielsen JA. Elements of Modern X-ray Physics, 2nd Edition Wiley 2011 pp.60).



3 Materiały i metody

W niniejszym rozdziale przedstawiono podstawowe protokoły wykonywanych w ramach prezentowanej pracy doktorskiej eksperymentów.

3.1 Lista wektorów używanych do nadekspresji białek

Nazwa wektora	Źródło/Producent	Tag/Białko fuzyjne	Miejsce cięcia proteazą	Konstrukty
pETM11	EMBL Heidelberg	6xHis	TEV	TaHSP90, HvHSP90 Δ C, HvTPR, AtTPR, ScTPR, HsTPR
pETM30	EMBL Heidelberg	6xHis-GST	TEV	HvSGT1-CS (SAXS), AtSGT1a-CS
pETM41	EMBL Heidelberg	6xHis-MBP	TEV	HvSGT1-CS (CD)
pnEA-vH	IGBMC Illkirch	6xHis	TEV	HvSGT1 Δ SGS, HvSGT1 Δ TPR
pET151-D-TOPO	Life Technologies	6xHis	TEV	HvSGT1
pDEST-HisMBP	Addgene	6xHis-MBP	TEV	HvRAR1

Opis skrótów: TEV - miejsce rozpoznawane przez proteazę z wirusa wżerkowej plamistości tytoniu (ang. Tobacco etch virus), GST (ang. glutathione S-transferase) - transferaza S-glutationowa, MBP (ang. maltose binding protein) - białko wiążące maltozę. Opis konstruktyw znajduje się w dalszej części tekstu.

3.2 Transformacja bakterii *E. coli* metodą szoku cieplnego

Dla trzech używanych w pracy szczepów bakterii *E. coli*: DH5 α (Invitrogen), BL21(DE3) Codon-Plus (Stratagene), BL21 Rosetta 2 (MerckMilipore), komórki kompetentne transformowano wektorem plazmidowym w ten sam sposób. Do rozmrożonych komórek dodawano 1-2 μ l plazmidu (\sim 100 ng) i inkubowano przez 30 minut na lodzie. Następnie bakterie inkubowano przez 45 sekund w 42 $^{\circ}$ C i przenoszono na lód w celu inkubacji przez dwie minuty. Po inkubacji do bakterii dodawano 300 μ l pożywki LB (o temperaturze 37 $^{\circ}$ C) i wytrząsano 300 rpm w 37 $^{\circ}$ C przez godzinę. Następnie wysiewano na szalki z pożywką LB i odpowiednim antybiotykiem selekcyjnym i inkubowano przez noc w 37 $^{\circ}$ C.

3.3 Przygotowanie konstruktyw plazmidowych do nadekspresji białek w systemie prokariotycznym

3.3.1 Przygotowanie wektora do klonowania

Wektor w pierw transformowano wg. procedury opisanej powyżej a następnie oczyszczano za pomocą zestawu do oczyszczania plazmidowego DNA z komórek *E. coli* GenElute Plasmid Miniprep Kit (Sigma Aldrich) według protokołu producenta. Do przygotowania wektorów wykorzystywanych w pracy używano komórek *E. coli* szczep DH5 α (Invitrogen). 25 ml hodowli bakterii dzielono na 4

próbówki 2 ml, zwirowując przez 1 minutę 14000 rpm na wirówce Hettich Mikro 200, jednorazowo 2 ml zawiesiny. Na jedną kolumnkę, ze złożem wiążącym DNA, nakładano supernatant otrzymany po lizie komórek z każdej z czterech probówek. Dalej, postępując zgodnie z protokołem zestawu GeneElute, otrzymywano czysty plazmid. Stężenie plazmidu mierzono za pomocą spektrometru UV-VIS Nanodrop 1000, mierząc absorbcję przy długości fali 260 nm.

3.3.2 Cięcie wektora za pomocą enzymów restrykcyjnych

Do reakcji cięcia enzymami restrykcyjnymi wykorzystywano enzymy z firmy Fermentas. Przykładowy skład mieszaniny reakcyjnej o objętości 50 μ l wyglądał następująco:

Reagent	objętość
Enzym 1 (10 j./ μ l)	1 μ l
Enzym 2 (10 j./ μ l)	1 μ l
plazmidowe DNA	30 μ l \sim 3 μ g
10x stężony bufor	5 μ l
H ₂ O	13 μ l

Po reakcji cięcia, trwającej około 16 godzin w 37 °C dodawano 1 μ l fosfatazy alkalicznej z cielęcych jelit (ang. calf intestinal alkaline phosphatase - CIAP) (Fermentas) i inkubowano przez 1 godzinę w celu usunięcia grupy fosforanowej z końca 5' DNA powstałego w wyniku cięcia enzymami, która powodowała by obniżenie wydajności reakcji ligacji insertu DNA do wektora. Po reakcji defosforylacji dezaktywowano obie reakcje przez inkubację przez 20 minut w 80 °C. Produkt cięcia oczyszczano za pomocą elektroforezy w żelu agarozowym. Wycięty z żelu prążek DNA oczyszczano następnie za pomocą zestawu GenElute Gel Extraction Kit (Sigma-Aldrich).

3.3.3 Przygotowanie insertu DNA do klonowania do wektora plazmidowego

Insert DNA przygotowywano poprzez reakcję łańcuchowej polimerazy (ang. polimerase chain reaction - PCR). Do przygotowania konstruktów zawierających pełną sekwencję białka wykorzystywano cDNA (ang. complementary DNA) z odpowiedniego organizmu. Do przygotowywania konstruktów zawierających tylko pojedyncze domeny lub fragmenty białek do reakcji PCR wykorzystywano gotowe konstrukty z wklonowaną pełną sekwencją kodującą białko. Skład przykładowej mieszaniny reakcyjnej PCR o objętości łącznej 25 μ l przedstawiony jest w tabelce poniżej:

Reagent	objętość
dNTP mix (4x2.5 mM każdy)	2 μ l
25mM MgSO ₄	2 μ l
Starter Forward (10 μ M)	2.5 μ l
Starter Reverse (10 μ M)	2.5 μ l
Matryca DNA	1 μ l
Polimeraza Pfu (Fermentas) (2.5j./ μ l)	0.5 μ l
10xBufor Pfu	2.5 μ l
H ₂ O	12 μ l

Produkt reakcji PCR oczyszczano za pomocą elektroforezy w żelu agarozowym za pomocą zestawu GenElute Gel Extraction Kit (Sigma-Aldrich). Następnie oczyszczony produkt reakcji PCR trawiono za pomocą odpowiednich enzymów restrykcyjnych podobnie jak plazmidowe DNA, jedynie opuszczając etap defosforylacji.

3.3.4 Ligacja insertu DNA do wektora plazmidowego

Reakcję ligacji przeprowadzano w objętość 10 μ l z użyciem ligazy DNA z faga T4 (Fermentas). Zazwyczaj do przygotowania jednego konstruktów wykonywano trzy reakcje zmieniając stosunek moli insertu do wektora: 3:1, 2:1, 1:1, zgodnie ze wzorem (gdzie $n = 3, 2$ lub 1):

$$\text{ilość insertu w ng} = n \frac{\text{wielkość insertu [pz]}}{\text{wielkość wektora [pz]}} \times \text{ilość wektora w ng} \quad (3.1)$$

a sama mieszanina reakcyjna wyglądała następująco:

Reagent	objętość
10xT4 ligase buffer	1 μ l
T4 ligaza (200 jednostek Weissa)	1 μ l
wektor	x μ l ~ 20-30 ng
insert	y μ l (zgodnie ze wzorem)
H ₂ O	do 10 μ l

Reakcję przeprowadzano w temperaturze pokojowej przez 3 godziny. Reakcję dezaktywano przez 10 minut w 70 °C. Całą objętością reakcji transformowano następnie komórki *E.coli* DH5 α zgodnie z protokołem.

3.3.5 Reakcja PCR “kolonijna”

Reakcję PCR “kolonijną” wykonywano w celu sprawdzenia reakcji ligacji. W tym celu wykonywano reakcję PCR ze starterami specyficznymi dla danego wektora, które dawały produkt PCR zawierający wklonowaną sekwencję. Identyfikację wielkość produktu (powinna być taka jak wielkość prowadzanego insertu) przeprowadzano za pomocą elektroforezy w żelu agarozowym. Pojedyncze kolonie z szalki zawieszano w 20 μ l pożywki LB i hodowano w 37 °C wytrząsając 300 rpm przez 1 godzinę. Następnie, bakterie używano jako matrycę w reakcji PCR z użyciem polimerazy DreamTaq (Fermentas). Składa mieszaniny reakcyjnej (10 μ l na jedną reakcję) wyglądał następująco:

Reagent	objętość
10xDreamTaq buffer	1 μ l
dNTP mix (4x2.5mM każdy)	0,8 μ l
Starter Forward (10 μ M)	0,5 μ l
Starter Reverse (10 μ M)	0,5 μ l
Matryca DNA	1 μ l
Polimeraza DreamTaq (5 j./ μ l)	0,1 μ l
H ₂ O	6,1 μ l

3.3.6 Namnażanie pozytywnych klonów

Kolonie bakteryjne zawierające wektor z wklonowanym insertem dodawano do 25 ml pożywki LB z antybiotykiem selekcyjnym i hodowano przez noc w 37 °C. Następnie wektor oczyszczano z komórek *E.coli* tak jak w procedurze (3.2.1). Wyizolowane plazmidy sprawdzano czy zawierają poprawnie wklonowaną sekwencję poprzez sekwencjonowanie w Wydziałowej Pracowni Techniki Biologii Molekularnych Wydziału Biologii UAM.

3.4 Nadekspresja białek na małą skalę w komórkach *E.coli*

W celu sprawdzenia poziomu ekspresji rekombinowanych białek przeprowadzono testy ekspresji w komórkach BL21 CodonPlus oraz Rosetta 2. W tym celu transformowano bakterie odpowiednim wektorem, zgodnie z protokołem 3.1. Następnego dnia po wysianiu, z szalki pobierano dwie-trzy kolonie i zawieszano je w 10 ml pożywki LB z dodatkiem odpowiednich antybiotyków. Bakterie hodowano przez noc (około 16 godzin) w 37 °C wytrząsając 200 rpm. Tak przygotowana kultura bakterii służyła następnie do zaszczepienia 25 ml świeżej pożywki LB. Do 25 ml pożywki LB dodawano 1 ml hodowli całonocnej i wytrząsano przez około 2 godziny w 37 °C do uzyskania $OD_{600\text{ nm}} = 0,6-0,8$ (pomiar absorpcji (rozpraszania) światła przy długości fali 600 nm przez kulturę bakterii). Następnie bakterie inkubowano przez około pół godziny w temperaturze pokojowej w celu jej schłodzenia. Do schłodzonej kultury bakterii dodawano IPTG (izopropyl β-D-1-tiogalaktopiranozyd), związku który indukuje ekspresję rekombinowanego białka, w ilości od 0,1 mM do 1 mM stężenia końcowego. Następnie bakterie inkubowano przez noc w temperaturze 18 °C przez 16 godzin wytrząsając 240 rpm. Dla niektórych białek nadekspresję prowadzono w 37 °C przez pięć godzin bez schładzania przed indukcją. Na sam koniec bakterie wirowano w 4 °C przy prędkości 14000 rpm 2 minuty i zamrażano w -20 °C do czasu dalszych analiz.

3.4.1 Nadekspresja białek na małą skalę metodą szoku cieplnego i osmotycznego

Metoda ta różni się od metody opisanej powyżej. W tej metodzie bakterie hoduje się najpierw na płynnej pożywce LBN zawierającej 0,5 M NaCl w celu wywołania szoku osmotycznego. Szok osmotyczny powoduje wytworzenie przez bakterie małych cząsteczkowych związków chemicznych zwanych osmolitami. Związki te stabilizują białka w komórce, w tym także rekombinowane białko, które chcemy otrzymać. W 10 ml pożywki LBN zawieszano 2-3 kolonie bakterii a następnie hodowano przez noc w 37 °C. Następnego do świeżej pożywki LBNB (pożywka LBN z dodatkiem 1mM betainy) dodawano 1 ml hodowli całonocnej i wytrząsano 240 rpm w 37 °C do uzyskania $OD_{600\text{ nm}} = 0,6-0,8$. Następnie bakterie przenoszono do łaźni wodnej o temperaturze 47 °C aż do ustalenia się tej temperatury. Następnie bakterie wytrząsano 240 rpm w 47 °C przez 20 minut. Zabieg ten miał na celu indukcję ekspresji białek szoku cieplnego, które pomagają w fałdowaniu się innych białek w tym rekombinowanego białka, które chcemy otrzymać. Następnie bakterie inkubowano w temperaturze 4 °C przez około 20-30 minut w celu schłodzenia i dodawano IPTG w celu indukcji nadekspresji. Bakterie inkubowano przez 16 godzin w 18 °C, wytrząsając 240 rpm. Po inkubacji bakterie wirowano w 4 °C przy prędkości 14000 rpm 2 minuty i zamrażano w -20 °C do czasu dalszych analiz.

3.5 Analiza poziomu ekspresji białek za pomocą elektroforezy w żelu poliakrylamidowym

Poziom ekspresji białek sprawdzano za pomocą elektroforezy w żelu poliakrylamidowym według systemu Laemliego. Do analizy poziomu ekspresji białek pobierano 2 ml kultury bakterii i wirowano przez 2 minuty z prędkością 14000 rpm. Odrzucano supernatant a osad zawieszano w 100 μ l buforu do lizy: 50 mM bufor fosforanowy pH 7,5, 500 mM NaCl, 30 mM imidazol, 10 mM β -merkaptopoetanol (dodawany tuż przed użyciem do buforu). Zawieszony osad bakteryjny sonikowano następnie z użyciem sonikatora Bioruptor firmy Diagenode, 15 cykli: 30 sekund sonikacja, 30 sekund inkubacja w temperaturze 4 °C. Po sonikacji bakterie wirowano 14000 rpm przez 10 minut w temperaturze 4 °C. Supernatant pobierano do nowej próbówki a osad zawieszano w 100 μ l buforu do lizy. Z próby supernatantu i osadu pobierano po 10 μ l i do każdej dodawano 10 μ l buforu 2xSSB. Próby następnie denaturowano w 95 °C przez 5 minut i nakładano na żel poliakrylamidowy przygotowany według następującej procedury. Najpierw przygotowywano żel separujący (10 ml) o składzie:

Składnik	Objętość
1M Tris pH 8,6	3,68 ml
30 % akrylamid/0,8 % bis-akrylamid	4 ml (14 %)
20 % dodecylosiarczan sodu	50 μ l
10 % nadsiarczan amonu	33 μ l
TEMED (N,N,N',N'-Tetrametyloetylenodiamina)	33 μ l
H ₂ O	2,232 ml

Czekano pół godziny aż żel spolimeryzuje. Następnie wierzchnią warstwą żelu pokryto izopropanolem (~300 μ l) w celu wyrównania jego powierzchni. Po pół godziny wylewano izopropanol a na żel separujący nakładano żel zagęszczający o składzie:

Składnik	Objętość
1M Tris pH 6,8	0,68 ml
30 % akrylamid/0,8 % bis-akrylamid	0,85 ml (5%)
20 % dodecylosiarczan sodu	25 μ l
10 % nadsiarczan amonu	25 μ l
TEMED (N,N,N',N'-Tetrametyloetylenodiamina)	12,5 μ l
H ₂ O	3,468 ml

Do roztworu niespolimeryzowanego żelu zagęszczającego dodawano również kilka kropeł wodnego roztworu błękitu bromofenolowego w celu lepszego uwidocznienia kieszonek do nakładania prób. Żel pozostawiano do spolimeryzowania przez pół godziny. Elektroforezę prowadzono w dwóch etapach. W pierwszym etapie, aż do momentu jak próby wnikały w żel separujący, używano napięcia 125 V. W drugim etapie, aż do momentu gdy barwnik zaczął wypływać z żelu, elektroforezę prowadzono przy napięciu 155 V. Po elektroforezie żel płukano w wodzie dwa razy po 10 minut a następnie barwiono za pomocą koloidalnego roztworu błękitu Coomassie: żel pokrywano roztworem barwnika i grzano w mikrofalówce aż do momentu pojawienia się pierwszych bąbelków wrzącej wody. Po pięciu minutach można było zaobserwować wyraźnie wybarwione prążki białka w żelu. Barwienie

kontynuowano przez noc. Żel odbarwiano poprzez ogrzewanie w mikrofalówce i płukanie w wodzie z włożonym kawałkiem ręcznika papierowego, który absorbował wypłukiwany barwnik. W celu archiwizacji wykonywano zdjęcie za pomocą systemu G:Box firmy Syngene. Do nadekspresji na dużą skalę wybierano warunki, w których poziom rekombinowanego białka był jak największy we frakcji rozpuszczalnej (supernatant).

3.6 Nadekspresja na dużą skalę rekombinowanych białek w bakteriiach *E. coli*

Po ustaleniu warunków nadekspresji na małą skalę prowadzono właściwą nadekspresję na dużą skalę w celu uzyskania preparatu białkowego do eksperymentów. Nadekspresję na dużą skalę prowadzono w objętość 1,6 litra (cztery kolby o objętości 1 l z 400 ml pożywki) lub 2 litry pożywki (dwie kolby o objętości 3 l z 1 l pożywki). W 100 ml pożywki zawieszano 2-3 kolonie z szalki i wytrząsano przez noc 200 rpm w 37 °C. Następnego dnia do pożywki o temperaturze 37 °C dodawano 20 ml (kolby 1 l) lub 30 ml (kolby 3 l) całonocnej kultury bakterii do każdej z kolb i wytrząsano w 37 °C 220-240 rpm aż do uzyskania $OD_{600\text{ nm}} = 0,6-0,8$. Następnie kolby inkubowano w temperaturze pokojowej przez pół godziny i dodawano IPTG do stężenia ustalonego w eksperymentach na małą skalę. Wytrząsano następnie przez 16 godzin w temperaturze 18 °C lub, bez uprzedniego chłodzenia, przez 5 godzin w 37 °C 220-240 rpm. Po zakończonej nadekspresji bakterie wirowano przez 15 minut 5000 g w 4 °C na wirówce Sorvall Evolution XL (rotor SLC-6000) lub Beckman Avanti J-26 XP (rotor JLA 9.1000). Osad zamrażano w temperaturze -20 °C do czasu dalszych analiz.

Procedurę nadekspresji na dużą skalę z zastosowaniem szoku osmotycznego i cieplnego przeprowadzano w takich samych warunkach jak nadekspresję na małą skalę oraz tak samo jak nadekspresję w zwykłych warunkach na dużą skalę z użyciem kolb o objętości 1 l.

3.7 Oczyszczanie białek z użyciem systemu do chromatografii cieczonej Äkta Explorer 10

Preparatywne oczyszczanie białek do analiz SAXS oraz dichroizmu kołowego przeprowadzano za pomocą systemu do chromatografii cieczonej Äkta Explorer 10 firmy Pharmacia. Poniżej zamieszczony jest standardowy protokół używany do oczyszczania białek wykorzystywanych w pracy.

3.7.1 Oczyszczanie za pomocą chromatografii powinowactwa

W pierwszym etapie białka oczyszczano za pomocą chromatografii powinowactwa do immobilizowanych jonów niklu. W technice tej, wprowadzone za pomocą inżynierii genetycznej, sześć reszt histydyny selektywnie wiąże się do złoża agarozowego zawierającego immobilizowane na jego powierzchni jony Ni^{2+} . Zamrożone w temperaturze -20 °C bakterie rozmrażano inkubując na lodzie. Do rozmrożonych komórek dodawano bufor IMAC A (skład buforu różnił się dla różnych białek, skład buforów dla różnych białek zebrany jest w tabeli w sekcji 3.7.5 z dodatkiem Triton X-100 0,5 % oraz inhibitorów proteaz bez EDTA (Roche). Do zawieszenia komórek w buforze używano 20 ml buforu na osad z 1,6 l hodowli bakterii. Bakterie zawieszano do momentu aż nie były obserwowane żadne agregaty osadu. Następnie za pomocą strzykawki z igłą rozbijano komórki obierając roztwór

komórek i przepuszczając szybko przez igłę. Tak przygotowane bakterie sonikowano za pomocą sonikatora Unipan ustawionego na poziom 4. Sonikacje przeprowadzano przez 6-7 cykli sonikacji trwającej 45 sekund z przerwą trwającą 1 minutę, podczas której inkubowano komórki na lodzie. Po sonikacji próbę wirowano 1 godzinę w 4 °C przy prędkości 8500 rpm z użyciem rotora kontowego i wirówki Heraeus (Thermo Scientific) . Zbierano supernatant i filtrowano go przez filtr Milex (Milipore) o średnicy porów 0,22 μm -0,45 μm . Następnie próbę nakładano na kolumnę HisTrap FF 1ml z prędkością 0,5 ml/min z użyciem buforu IMAC A. Próbę nakładano za pomocą pętli SuperLoop 10 ml (Pharmacia). Po nałożeniu próby, kolumnę płukano dziesięcioma objętościami kolumny (10 ml) buforem A z prędkością 1 ml/min w celu odmycia niespecyficznie związanych białek. Następnie związane białko eluowano za pomocą buforu IMAC B (zmieniając liniowo jego stężenie od 0 % do 100 % przez objętość elucji dwudziestu objętości kolumny (20 ml). Elucję przeprowadzano przy przepływie 0,5 ml/min. Eluowane białko zbierano w postaci frakcji o objętości 1 ml. Obecność białka we frakcjach oraz jego czystość sprawdzano za pomocą elektroforezy denaturującej w żelu poliakrylamidowym. Frakcje zawierające białko zbierano, dodawano proteazy TEV (mutant S219V) w ilość 60 μl -70 μl (stężenie 4-5 mg/ml) w celu odcięcia fragmentu sześciu histydyn na N -końcu białka i dializowano (membrana Spectra-por o średnicy porów przepuszczających cząsteczki mniejsze niż 8 kDa lub membrany Snake Skin o średnicy porów przepuszczających cząsteczki mniejsze niż 3 kDa) wobec buforu IMAC A przez noc w 4°C delikatnie mieszając roztwór za pomocą mieszadła magnetycznego. Następnego dnia próbę przefiltrowano przez filtr Milex w celu pozbycia się ewentualnego precipitatu powstałego podczas dializy. Następnie roztwór białka nakładano na kolumnę HisTrap FF 1 ml z użyciem buforu IMAC A. Zbierano frakcje nie związane ze złożem, które powinny zawierać interesujące nas białko bez etykiety sześciu histydyn.

3.7.2 Oczyszczanie białek za pomocą sączenia molekularnego

Roztwór białka po oczyszczeniu na kolumnie HisTrap FF 1 ml zagęszczano do objętości \approx 2 ml. W tym celu roztwór białka zagęszczano za pomocą filtrów Amicon 15 (Milipore) z membraną przepuszczającą cząsteczki o masie mniejszej niż 3 kDa. Wirowanie przeprowadzano przy prędkości 4000 g z użyciem rotora wychylnego i wirówki Heraeus. Jedno wirowanie trwało 10 minut. Po zagęszczeniu próbę filtrowano przez filtr Milex o średnicy porów 0,22 μm -0,45 μm i nakładano, przy pomocy pętli SuperLoop 10 ml, na kolumnę HiLoad 16/60 Superdex 200pg (GE Lifescience) z przepływem 0,2 ml/min przy pomocy 3 ml buforu SEC (skład buforu w sekcji 3.7.5). Elucję prowadzono przy przepływie 0,5 ml/min, kolekcjonując frakcje o objętości 1,5 ml/min. Zebrane frakcje analizowano za pomocą elektroforezy w warunkach denaturujących w żelu poliakrylamidowym. Frakcje zawierające białko, zbierano i zagęszczano za pomocą filtra Amicon 15 do objętości 0,5 ml - 1 ml. Stężenie białka mierzono za pomocą spektrofotometru UV-vis Nanodrop-1000 przy długości fali 280 nm korzystając z teoretycznie obliczonego współczynnika ekstynkcji na podstawie sekwencji białka. Roztwór białka dzielono na porcje o objętości 50 μl -100 μl , zamrażano w ciekłym azocie i przechowywano do czasu dalszych analiz w temperaturze -80 °C.

3.7.3 Oczyszczanie białek za pomocą chromatografii jonowymiennej

Dla niektórych białek, w celu otrzymania czystego preparatu, niezbędny był jeszcze jeden krok oczyszczania - chromatografia jonowymienna. W zależności od białka stosowano kolumnę z wymiennicem kationowym MonoS 5/5 HR (Pharmacia) lub anionowym MonoQ 5/5 HR (Pharmacia). Skład używanych buforów opisany jest w sekcji 3.7.5. Przed rozdziałem frakcje z kolumny Superdex dializowano przez noc wobec buforu IEX A w celu pozbycia się chlorku sodu. Próbkę po dializie filtrowano przez filtr Milex i nakładano na kolumnę z przepływem 0,5 ml/min używając buforu IEX A. Dla niektórych białek zbierany był przesącz zawierający właściwe, oczyszczone już białko. Następnie odmyto niespecyficznie związane białka przemywając kolumnę dziesięcioma objętościami kolumny (10 ml) buforem IEX A. Białko wymywano za pomocą buforu IEX B, liniowo zmieniając jego stężenie od 0 % do 50 % w trakcie elucji przez trzydzieści objętości kolumny a następnie zmieniając stężenie od 50 % do 100 % przez dziesięć objętości kolumny. Zbierano frakcje o objętości 0,5 ml/min. Frakcje zawierające białko zbierano i dializowano przez noc w 4 °C względem buforu docelowego. Po dializie roztwór białka zagęszczano za pomocą filtra Amicon 15 do objętości 0,5 ml-1 ml. Stężenie białka mierzono za pomocą spektrofotometru UV-vis Nanodrop-1000 przy długości fali 280 nm korzystając z teoretycznie obliczonego współczynnika ekstynkcji na podstawie sekwencji białka. Roztwór białka dzielono na porcje o objętości 50 μ l -100 μ l, zamrażano w ciekłym azocie i przechowywano do czasu dalszych analiz w temperaturze -80 °C.

3.7.4 Oczyszczanie białka SGT1 z komórek *E.coli*

W tym celu posłużono się strategią polegającą na ekspresji białka SGT1 z dołączoną etykietą sześciu histydyn do oczyszczania za pomocą chromatografii powinowactwa. Oprócz etykiety 6xHis wprowadzono również miejsce rozpoznawane przez proteazę TEV, za pomocą której możliwe było docięcie etykiety 6xHis i otrzymania pełnej długości białka SGT1 (z dodatkowymi dwoma aminokwasami na N-końcu, pozostałymi po odcięciu etykiety 6xHis). Białko SGT1 było produkowane w bakteriach *E.coli* BL21 CodonPlus na wysokim poziomie i znajdowało się głównie we frakcji rozpuszczalnej. Białko oczyszczano z komórek bakteryjnych zgodnie z protokołem. Najpierw posłużono się chromatografią powinowactwa w celu wstępnego oczyszczania białka. Podczyszczony materiał był dalej oczyszczany na kolumnie HiLoad Superdex 200pg. Białko SGT1 wypływało z kolumny w postaci jednego, asymetrycznego piku. Asymetria jest często związana z występowaniem równowagi pomiędzy formami oligomerycznymi lub degradacją. W tym przypadku degradacja białka SGT1 była odpowiedzialna za ten efekt. Analiza czystości zebranych frakcji pokazała, że jedynie dwie pierwsze frakcje zawierają niezdegradowane białko SGT1. Frakcje te były dalej cięte za pomocą proteazy TEV przez noc a następnie oddzielono proteazę TEV i niepocięte białko SGT1 za pomocą chromatografii powinowactwa. Zebrany roztwór białka SGT1 dializowano do ostatecznego buforu przez noc.

3.7.5 Skład buforów używanych podczas oczyszczania białek

Nazwa buforu	Skład	Konstrukty
IMAC A	50 mM bufor fosforanowy pH 7,5, 300 mM NaCl, 30 mM imidazol, 10 mM 2-merkaptioetanol	HvSGT1, HvSGT1-CS (CD), HvSGT1-CS (SAXS), AtSGT1a-CS, HvSGT1 Δ SGS, HvSGT1 Δ TPR, HvTPR, AtTPR, ScTPR, HsTPR,
IMAC B	50 mM bufor fosforanowy pH 7,5, 300 mM NaCl, 500 mM imidazol, 10 mM 2-merkaptioetanol	HvSGT1, HvSGT1-CS (CD), HvSGT1-CS (SAXS), AtSGT1a-CS, HvSGT1 Δ SGS, HvSGT1 Δ TPR, HvTPR, AtTPR, ScTPR, HsTPR,
IMAC A	50 mM Tris pH 7,5, 300 mM NaCl, 30 mM imidazol, 10 mM 2-merkaptioetanol	TaHSP90, HvHSP90 Δ C
IMAC B	50 mM Tris pH 7,5, 300 mM NaCl, 500 mM imidazol, 10 mM 2-merkaptioetanol	TaHSP90, HvHSP90 Δ C
IMAC A	50 mM Tris pH 7,5, 500 mM NaCl, 30 mM imidazol, 100 μ M ZnSO ₄ , 1 mM DTT	HvRAR1
IMAC B	50 mM Tris pH 7,5, 500 mM NaCl, 500 mM imidazol, 100 μ M ZnSO ₄ , 1 mM DTT	HvRAR1
Bufor SEC	20 mM Tris pH 7,5, 200 mM NaCl, 5 mM DTT	wszystkie konstrukty
Bufor IEX A	20 mM Tris pH 7,5, 1 mM DTT	TaHSP90, HvHSP90 Δ , HvRAR1
Bufor IEX B	20 mM Tris pH 7,5, 1 M NaCl, 1 mM DTT	TaHSP90, HvHSP90 Δ , HvRAR1

3.7.6 Oczyszczanie wariantów białek RAR1, SGT1 i HSP90

Białka używane w eksperymentach wykonanych w pracy zostały oczyszczone wg. protokołów 3.7.1-3.7.3. W poniższej tabeli zostały zebrane wszystkie oczyszczone i używane przeze mnie konstrukty białek RAR1, SGT1 i HSP90.

Nazwa konstruktu białka	aminokwasy	Organizm	Białko (nazwa z bazy danych Uniprot)	Numer w bazie danych Uniprot
HvSGT1	1-373	<i>Hordeum vulgare</i>	SGT1	Q8W516
HvSGT1 Δ SGS	1-261	<i>Hordeum vulgare</i>	SGT1	Q8W516
HvSGT1 Δ TPR	174-373	<i>Hordeum vulgare</i>	SGT1	Q8W516
HvSGT1-CS (SAXS)	167-276	<i>Hordeum vulgare</i>	SGT1	Q8W516
HvSGT1-CS (CD)	173-264	<i>Hordeum vulgare</i>	SGT1	Q8W516
AtSGT1a-CS	151-240	<i>Arabidopsis thaliana</i>	SGT1a	Q9SUR9
HvRAR1	1-232	<i>Hordeum vulgare</i>	RAR1	Q9SE34
TaHSP90	1-700	<i>Triticum aestivum</i>	HSP90.2-B1	F4Y593
HvHSP90 Δ C	1-531	<i>Hordeum vulgare</i>	HSP90	Q7XJ80
HvTPR	6-124	<i>Hordeum vulgare</i>	SGT1	Q8W516
AtTPR	1-120	<i>Arabidopsis thaliana</i>	SGT1b	Q9SUR9
ScTPR	10-138	<i>Saccharomyces cerevisiae</i>	SGT1	Q08446
HsTPR	6-129	<i>Homo sapiens</i>	SUGT1	Q9Y2Z0-2

Na rysunku 2.4 pokazany jest schemat struktury domenowej białek RAR1, SGT1 i HSP90 wraz z zaznaczonymi numerami aminokwasów na granicach domen.

3.8 Pomiary dichroizmu kołowego białka SGT1 oraz konstruktów zawierających domeny TPR-CS, CS-SGS, CS

Roztwory białka do badań dichroizmu kołowego dializowano przez noc w 4 °C względem buforu 10 mM fosforan sodu pH 7,5. Po dializie próbę wirowano przez 30 minut 14000 rpm w 4 °C oraz przefiltrowano przez filtr o średnicy porów 0,1 μm w celu pozbycia się agregatów, które mogły powstać podczas dializy. Próby mierzono w kuwecie kwarcowej o drodze optycznej 0,02 cm z użyciem spektrofotometru Jasco J-815. Dla pełnej długości białka SGT1 o stężeniu 12 μM pomiar wykonano w postaci sześciu skanów w zakresie długości fali od 178 nm do 280 nm. Domeny TPR-CS (stężenie 10,4 μM) oraz CS-SGS (stężenie 12,8 μM) mierzono w zakresie 185 nm-260 nm wykonując sześć skanów. Pomiary dla domeny CS z białka SGT1 z jęczmienia i *Arabidopsis* przeprowadzono w zakresie 205 nm-260 nm również w ilości 5 skanów. W celu sprawdzenia efektu zagęszczenia molekularnego na konformację domeny SGS, konstrukt domeny CS-SGS mierzono w buforze: 10 mM fosforan sodu pH 7,5, 30 % Ficoll 70K, w zakresie długości fali 205 nm- 260 nm, w ilości 5 skanów.

3.9 Pomiary małokątowego rozpraszania promieniowania synchrotronowego

Pomiary synchrotronowe wykonywano w ten sposób, że dla każdej próbki mierzono bufor, następnie roztwór białka i drugi raz bufor. Dwukrotny pomiar rozpraszania dla buforu przeprowadzano w celu otrzymania dobrego tła, które należało dociąć w celu otrzymania krzywej rozpraszania dla samego białka. Na każdej linii pomiarowej, za pomocą robota, wykonano pomiar rozpraszania małokątowego promieniowania synchrotronowego na roztworze białka w objętości od 30 μl do 100 μl w kapilarze szklanej w temperaturze 15°C. W zależności od linii pomiarowej, jednorazowy pomiar trwał od 2 sekund do 2 minut. W programie PRIMUS z pakietu ATSAS dokonywano odjęcia tła od krzywej rozpraszania dla badanego białka.

3.9.1 Pomiary SAXS białka HvSGT1

Białko HvSGT1 mierzono na linii pomiarowej X33 na synchrotronie Doris w Hamburgu. Białko mierzono w buforze 50 mM Tris HCl, 5 mM DTT oraz: NaCl 0 mM i stężeniu białka 3,46 mg/ml, 500 mM NaCl i stężeniu białka 3,41 mg/ml, 1 M NaCl i stężeniu białka 2,71 mg/ml.

3.9.2 Pomiary SAXS białka HvRAR1

Białko HvRAR1 mierzono na linii pomiarowej BM29 na synchrotronie ESRF. Białko mierzono w buforze 50 mM Tris HCl, 150 mM NaCl, 5 mM DTT, 50 μM ZnCl₂ w dwóch stężeniach 3,01 mg/ml i 6,02 mg/ml.

3.9.3 Pomiary SAXS białek HvTPR, AtTPR, ScTPR, HsTPR

Pomiary SAXS białka HvTPR przeprowadzono na linii pomiarowej BM29 na synchrotronie ESRF w Grenoble, Francja. Natomiast białka AtTPR, ScTPR, HsTPR mierzono na linii pomiarowej P12 na synchrotronie Petra III w Hamburgu. Wszystkie białka mierzono w buforze 50 mM Tris HCl pH 7,5, 5 mM DTT oraz stężeniu NaCl (białko HvTPR) lub KCl (białka AtTPR, ScTPR, HsTPR) od 0 mM do 800 mM.

3.9.4 Pomiary SAXS białka TaHSP90

Pomiary SAXS białka TaHSP90 w różnym stężeniu KCl i z różnymi nukleotydami przeprowadzono na linii pomiarowej BM29 na synchrotronie ESRF w Grenoble, Francja. Stężenie białka wynosiło 4 mg/ml. Pomiary przeprowadzono w buforze 50 mM Tris HCl, 1 mM DTT, 10 mM MgCl₂ oraz KCl w stężeniu od 0 mM do 800 mM z dodatkiem 10 mM nukleotydu lub bez. Pomiary temperaturowe białka TaHSP90 przeprowadzono na linii pomiarowej P12 na synchrotronie Petra III w Hamburgu. Próbkę zawierała białko TaHSP90 o stężeniu 9 mg/ml oraz bufor o składzie: 50 mM Tris HCl, 1 mM DTT, 150 mM NaCl. Temperaturę zmieniano co 5 °C stopni w zakresie od 10 °C do 50 °C, inkubując próbkę 1 minutę w zadanej temperaturze. Dla każdej temperatury badana była nowa próbka.

3.9.5 Pomiary SAXS kompleksów białek TaHSP90-HvSGT1ΔSGS

Pomiary synchrotronowe kompleksu białek TaHSP90-HvSGT1ΔSGS przeprowadzono na linii pomiarowej P12 na synchrotronie Petra III w Hamburgu, Niemcy. Białka mierzono w buforze 50 mM Tris pH 7,5, 100 mM NaCl, 1 mM DTT, 10 mM ADP i 10 mM MgCl₂. Białko HvSGT1ΔSGS mierzono w stężeniu 0,9 mg/ml a białko TaHSP90 przy stężeniu 4 mg/ml. Pomiary dla kompleksu białek prowadzono przy stężeniu białka 4 mg/ml i zmieniając stężenie białka HvSGT1ΔSGS by otrzymać stosunek molowy białka TaHSP90 do białka HvSGT1ΔSGS: 1,2:1; 1:1; 1:1,2; 1:1,5; 1:1,7, 1:2. Kompleksy białek przygotowano poprzez zmieszanie białek w odpowiednich stosunkach molowych i inkubacje na lodzie przynajmniej 30 minut.

3.9.6 Pomiary SAXS kompleksów białek HvHSP90ΔC- HvSGT1ΔSGS i HvHSP90ΔC-HvSGT1-CS

Pomiary synchrotronowe kompleksu białek HvHSP90ΔC-HvSGT1ΔSGS i HvHSP90ΔC-HvSGT1-CS oraz pojedynczych białek przeprowadzono na linii pomiarowej B21 na synchrotronie Diamond w Didcot, Wielka Brytania. Białka mierzono w buforze 50 mM Tris pH 7,5, 100 mM NaCl, 1 mM DTT, 10 mM DTT i 10 mM MgCl₂ oraz 10 mM ADP. Kompleks białek HvHSP90ΔC- HvSGT1ΔSGS mierzono przy stężeniu białka HvHSP90 4,55 mg/ml i zmieniając stężenie białka HvSGT1ΔSGS by otrzymać stosunek molowy białka HvHSP90ΔC do białka HvSGT1ΔSGS: 1,5:1; 1,2:1; 1:1; 1:1,2; 1:1,5. Kompleks białek HvHSP90ΔC- HvSGT1-CS mierzono przy stężeniu białka HvHSP90 4,42 mg/ml i zmieniając stężenie białka HvSGT1-CS by otrzymać stosunek molowy białka HvHSP90ΔC do białka HvSGT1-CS: 1,5:1; 1,2:1; 1:1; 1:1,2; 1:1,5. Kompleksy białek przygotowano poprzez zmieszanie białek w odpowiednich stosunkach molowych i inkubacje na lodzie przynajmniej 30 minut.

3.9.7 Lista programów używanych w analizie danych SAXS

W poniższej tabeli przedstawione są używane przeze mnie programy, którymi posługiwałem się w trakcie analizy danych małokątowego rozpraszania promieniowania rentgenowskiego.

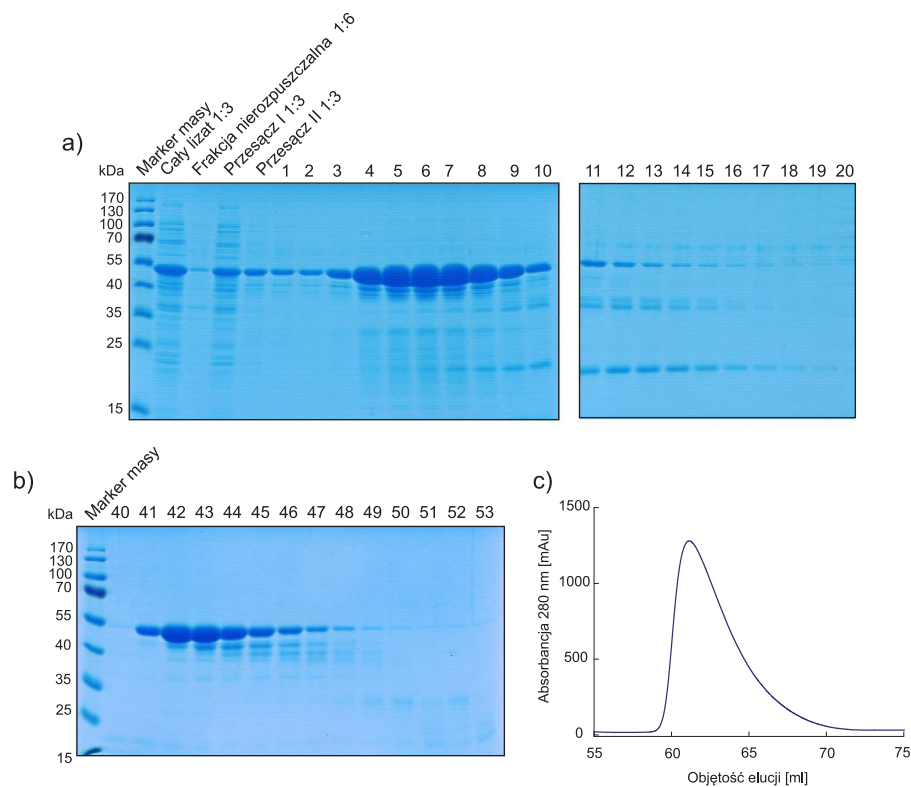
Program	Wydawca	Pakiet oprogramowania	Typ analizy danych
PRIMUS	EMBL, D.Svergun	ATSAS 2.3-2.6	Redukcja danych, analiza R_g
GNOM	EMBL, D.Svergun	ATSAS 2.3-2.6	Obliczanie krzywej $P(r)$, analiza D_{max}
CRYSOL	EMBL, D.Svergun	ATSAS 2.3-2.6	Obliczanie krzywej rozpraszania dla teoretycznych modeli białek
DAMMIN	EMBL, D.Svergun	ATSAS 2.3-2.6	Modelowanie niskorozdzielczej struktury białek
DAMMIF	EMBL, D.Svergun	ATSAS 2.3-2.6	Modelowanie niskorozdzielczej struktury białek
DAMAVER	EMBL, D.Svergun	ATSAS 2.3-2.6	Uśrednianie modeli kulkowych
MONSA	EMBL, D.Svergun	ATSAS 2.3-2.6	Modelowanie niskorozdzielczej struktury białek
GASBOR	EMBL, D.Svergun	ATSAS 2.3-2.6	Modelowanie niskorozdzielczej struktury białek (łańcucho-podobnej)
BUNCH	EMBL, D.Svergun	ATSAS 2.3-2.6	Modelowanie struktury białek metodą bryły sztywnej (z dynamicznymi połączeniami pomiędzy domenami)
SASREF	EMBL, D.Svergun	ATSAS 2.3-2.6	Modelowanie struktury białek metodą bryły sztywnej
CORAL	EMBL, D.Svergun	ATSAS 2.3-2.6	Modelowanie struktury białek metodą bryły sztywnej (kilka łańcuchów polipeptydowych)
EOM	EMBL, D.Svergun	ATSAS 2.3-2.6	Modelowanie struktury białek za pomocą zespołu konformacji
Scatter	Diamond Light Source, R.Rambo	Scatter 2.3a	Redukcja danych, analiza R_g , określanie masy molekularnej
Matlab	MathWorks, Inc.	Matlab R2010, R2014	analiza SVD
MCR-ALS Toolbox	Romà Tauler, IDAEA-CSIC	Matlab R2010, R2014	analiza MCR-ALS
Quark	Yang Zhang, University of Michigan	http://zhanglab.cmb.med.umich.edu/QUARK/	Modelowanie <i>ab-initio</i> struktury białek
I-Tasser	Yang Zhang, University of Michigan	http://zhanglab.cmb.med.umich.edu/I-TASSER/	Modelowanie homologiczne struktury białek

Wartość dopasowania modeli otrzymanych przez programy do modelowania *ab-initio* (programy z pakietu ATSAS) do danych eksperymentalnych podawane jest w postaci wartości χ . Dla programu EOM wartość dopasowania podawana jest w postaci wartości χ^2 .

4 Wyniki

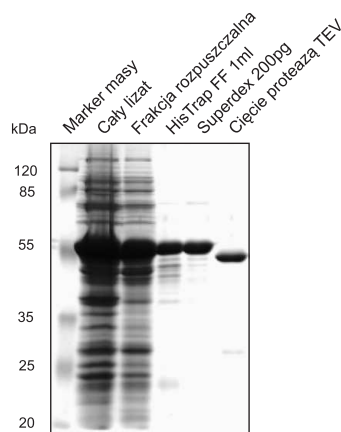
4.1 Otrzymywanie pełnej długości białka SGT1

Jednym z głównych celów pracy doktorskiej było określenie struktury białka SGT1 z jęczmienia w roztworze za pomocą małokątowego rozpraszania promieniowania synchrotronowego. W tym celu należało otrzymać czysty i homogeny roztwór białka SGT1. Z prac innych autorów wynikało, że białko SGT1 ma tendencję do ulegania częściowej proteolizie podczas nadekspresji w komórkach bakteryjnych i owadzych (Nyarko i wsp., 2007, Catlett i wsp., 2006). Pierwszym zadaniem było więc opracowanie metody otrzymania niezdegradowanego białka SGT1. W tym celu zastosowano procedurę oczyszczania nadproduktowanego w bakteriiach *E.coli* białka SGT1, na którą składają się: chromatografia powinowactwa na złożu z immobilizowanymi jonami niklu, sączenie molekularnego na złożu Superdex 200, cięcie proteazą TEV, ponownie chromatografia powinowactwa na złożu z immobilizowanymi jonami niklu. Na rysunku 4.1 zamieszczone są przykładowe wyniki oczyszczania za pomocą chromatografii powinowactwa oraz sączenia molekularnego.



Rysunek 4.1 Oczyszczanie białka SGT1. Rozdział elektroforetyczny w żelu poliakrylamidowym frakcji zebranych po wstępnym oczyszczeniu za pomocą chromatografii powinowactwa a), oraz po sączeniu molekularnym b). Na wykresie c) przedstawiony jest profil elucji białka SGT1 z kolumny Superdex 200pg.

Jak widać na rysunku 4.1 B jedynie dwie pierwsze frakcje otrzymane z oczyszczania na kolumnie ze złożem Superdex 200 zawierały białko SGT1 o odpowiedniej czystości. Na rysunku 4.2 przedstawiona jest czystość białka SGT1 podczas kolejnych kroków oczyszczania analizowana, na żelu poliakrylamidowym. Procedura ta pozwalała na otrzymanie 1 do 2 miligramów czystego białka SGT1 z 1,6 l hodowli komórek *E.coli*.

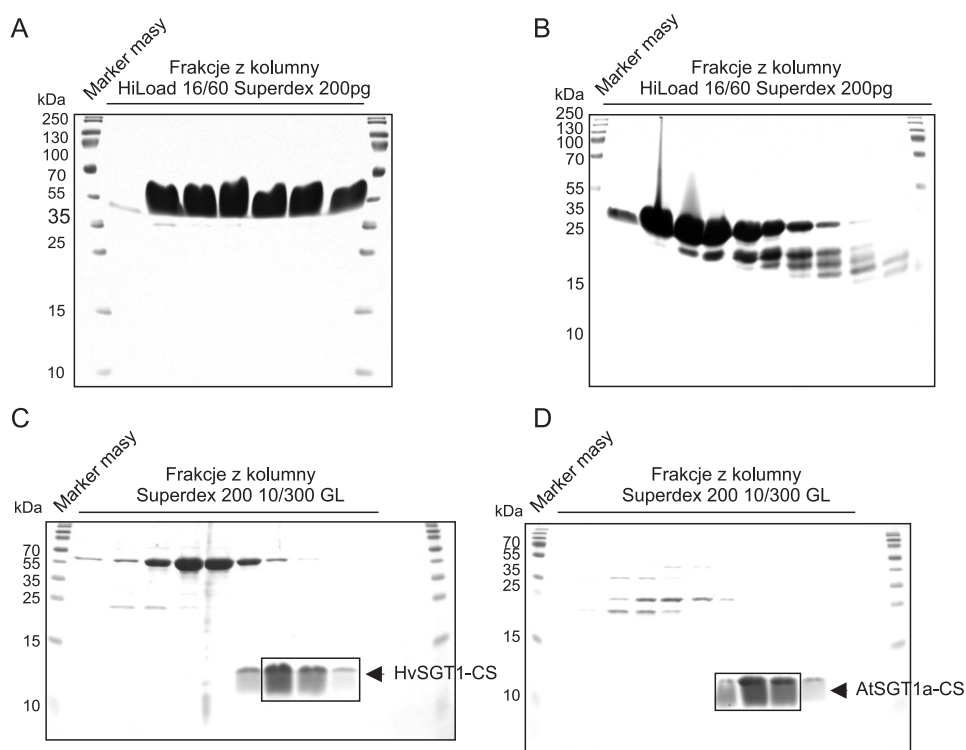


Rysunek 4.2 Elektroforeza w żelu poliakrylamidowym (14 %) podsumowująca poszczególne etapy oczyszczania pełnej długości białka SGT1.

4.2 Pomiary spektroskopii dichroizmu kołowego pełnej długości białka SGT1 oraz konstruktyw zawierających domeny: TPR-CS, CS-SGS oraz CS

4.2.1 Oczyszczanie białek do pomiarów dichroizmu kołowego

Konstrukty białka SGT1 zawierające domeny: TPR-CS (HvSGT1 Δ SGS), CS-SGS (HvSGT1 Δ TPR), CS (HvSGT1-CS (CD) oraz AtSGT1a-CS) (dokładny opis w tabeli w paragrafie 3.7.5) oczyszczane były według protokołów opisanych w sekcji 3.7.1-3.7.3, przy użyciu chromatografii powinowactwa, cięcia proteazą TEV, ponownym oczyszczaniu za pomocą chromatografii powinowactwa oraz sączenia molekularnego. Na rysunku 4.3 przedstawiona jest czystość konstruktyw białek po oczyszczaniu za pomocą sączenia molekularnego. W przypadku konstruktyw domen CS, rekombinowane białka oczyszczano z 200 ml bakterii *E.coli*, a do sączenia molekularnego posłużono się kolumną Superdex 200 10/300 GL o mniejszej objętości złoża niż kolumna HiLoad 16/60 Superdex 200pg.



Rysunek 4.3 Oczyszczanie konstruktów białka SGT1 do pomiarów dichroizmu kołowego. **A.** Oczyszczanie konstruktu domen TPR-CS na kolumnie Superdex 200pg. **B.** Oczyszczanie konstruktu domen CS-SGS na kolumnie Superdex 200pg. **C.** Oczyszczanie konstruktu domeny CS białka HvSGT1 na kolumnie Superdex 10/300 GL. Czarną ramką zaznaczone są frakcje użyte do pomiarów dichroizmu kołowego. **D.** Oczyszczanie konstruktu domeny CS białka AtSGT1a na kolumnie Superdex 10/300 GL. Czarną ramką zaznaczone są frakcje użyte do pomiarów dichroizmu kołowego.

4.2.2 Pomiary dichroizmu kołowego

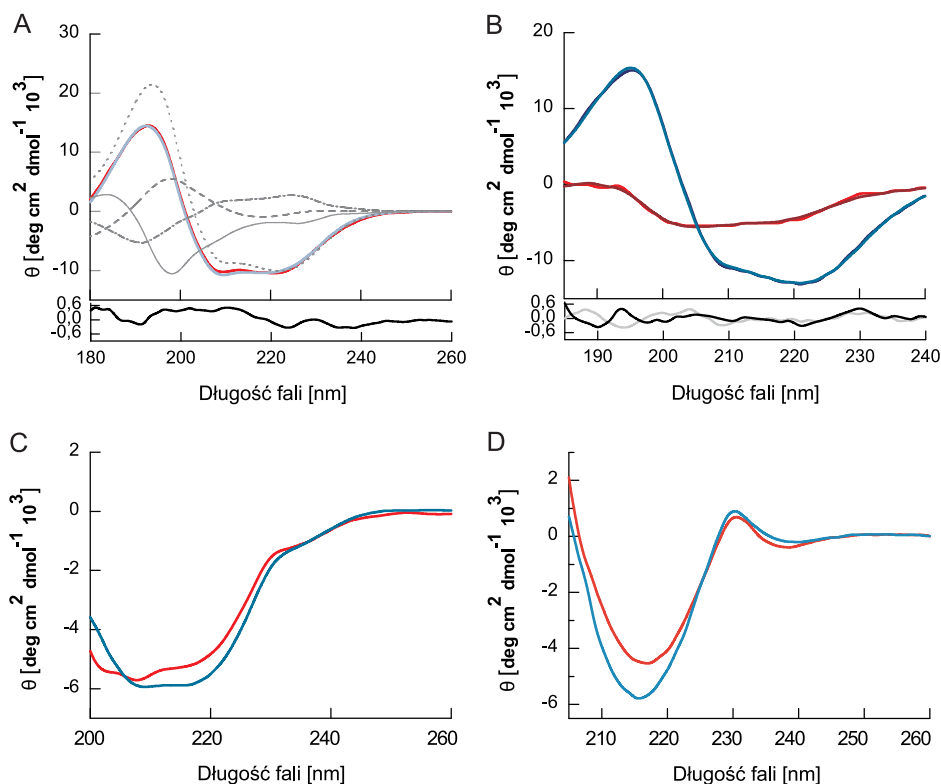
Pomiary dichroizmu kołowego przeprowadzono w celu zbadania stopnia uporządkowania (struktury drugorzędowej) białka SGT1 oraz jego domen. Widmo dichroizmu kołowego dla pełnej długości białka SGT1 przedstawione jest na rysunku 4.4 A. Z analizy kształtu widma CD można wywnioskować, że pełnej długości białko posiada fragmenty zawierające helisy α , o czym świadczą dwa minima: przy 208 nm oraz przy 222 nm. Drugie minimum na widmie CD ma mniejszą intensywność co świadczy o wkładzie innych elementów struktury drugorzędowej do widma dichroizmu kołowego. W celu procentowego określenia poszczególnych komponentów struktury drugorzędowej dokonano dekonwolucji widma za pomocą programu CDSSTR. Niestety dla izolowanej domeny CS udało się zarejestrować widma jedynie od 205 nm, co uniemożliwiło wykonanie dekonwolucji w programie CDSSTR oraz innych dostępnych programach, ponieważ do dekonwolucji wymagane są dane CD zarejestrowane co najmniej od 200 nm (program K2D). Wyniki zebrane są w tabeli 4.1. Z widma CD zebranego dla domeny CS (rysunek 4.4 D, krzywa czerwona) wynika, że domena CS uporząd-

kowana jest w przeważającej mierze w postaci struktury arkusza β , ponieważ można zaobserwować charakterystyczne minimum dla długości fali ≈ 215 nm.

Tabela 4.1 Analiza zawartość elementów struktury drugorzędowej na podstawie pomiarów dichroizmu kołowego.

Konstrukt	helisy α	arkusz β	zwrot β	regiony nieuporządkowane
SGT1 pełnej długości	33,9 %	21,9 %	11,8 %	33,5 %
TPR-CS	36 %	17 %	16 %	30 %
CS-SGS	5 %	29 %	18 %	48 %
CS	-	-	-	-
CS-SGS + Ficoll	-	-	-	-

Dla porównania, na tej samej rycinie, przedstawione jest także widmo dichroizmu kołowego dla domeny CS z białka SGT1 z *Arabidopsis thaliana* (rysunek 4.4 D), dla którego znana jest struktura przestrzenna. Oba widma mają ten sam kształt jedynie różnią się wartością sygnału, co może być spowodowane błędem w określeniu stężenia białka, nie można wykluczyć również drobnych zmian strukturalnych pomiędzy domenami CS z różnych organizmów.



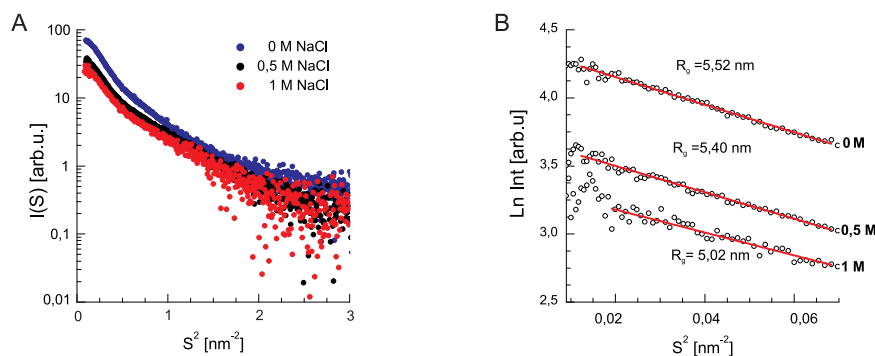
Rysunek 4.4 Widma dichroizmu kołowego dla białka SGT1 i jego izolowanych domen. **A.** Widmo dichroizmu kołowego dla pełnej długości białka SGT1. Krzywa czerwona dane eksperymentalne, dopasowana krzywa jasnoniebieska otrzymana w procesie dekonwolucji. W kolorze szarym widoczne są

widma dla poszczególnych struktur drugorzędowych: (-) struktury nieuporządkowane, (—) arkusz β , (.....) helisy α , (-.-.) zwrot β . **B.** Widma dichroizmu kołowego dla izolowanych białek TPR-CS oraz CS-SGS, krzywa ciemnoniebieska oraz jasnoczerwona odpowiednio wraz z dopasowaniem z programu CDSSTR, dla izolowanych białek TPR-CS, krzywa jasnoniebieska oraz CS-SGS, krzywa bordowa. **C.** Wpływ zagęszczenia środowiska na strukturę drugorzędową białka SGT1 CS-SGS. Kolor czerwony konstrukt CS-SGS w buforze bez czynnika zagęszczającego Ficoll 70k, kolor niebieski widmo dichroizmu kołowego dla domeny CS-SGS w buforze zawierającym 30 % czynnika zagęszczającego Ficoll 70k. **D.** Widmo dichroizmu kołowego dla domeny CS z białka SGT1 z jęczmienia (kolor czerwony) i *Arabidopsis thaliana* (kolor niebieski).

Na podstawie analizy zawartości struktur drugorzędowych dla pełnej długości białka można wywnioskować, że białko ma największy udział helisy α w strukturze drugorzędowej. Wykład od helisy α można przypisać strukturze drugorzędowej domeny TPR, ponieważ domena TPR składa się głównie z helisy α . Wkład procentowy jest proporcjonalny do długości domeny TPR, która ma długość 120 aminokwasów w porównaniu do 373 aminokwasów dla całego białka. Jak już wspomniano wyżej widmo dla domeny CS zawiera cechy charakterystyczne dla białek o budowie arkusza β . Wkład arkuszy β wynosi 21,9 % dla pełnej długości białka. Również i w tym przypadku jest to zgodne z porównaniem długości domeny CS do długości całego białka SGT1. Wyniki otrzymane dla krótszych fragmentów białka SGT1: TPR-CS oraz CS-SGS również potwierdzają, że domena TPR oraz CS mają odpowiednio strukturę α -helikalną i arkusza β . Analiza zawartości struktur nieuporządkowanych w widmach dla pełnej długości białka SGT1 oraz konstruktów TPR-CS oraz CS-SGS (widma widoczne na rysunku 4.4 B) pozwalają wywnioskować, że regiony pomiędzy domenami TPR i CS (VR1) oraz CS i SGS (VR2) są nieuporządkowane. Regiony VR1 i VR2 mają długość łącznie 67 aminokwasów (18 % aminokwasów w białku SGT1). Także można wywnioskować, że ok 18 % struktury drugorzędowej pochodzi od regionów VR1 i VR2. Natomiast 33,5 % struktury drugorzędowej stanowią struktury nieuporządkowane. Jeżeli przyjąć, że struktury nieuporządkowane pochodzą tylko od regionów VR1 i VR2 to prawie 12 % musi jednak pochodzić od reszty białka. W konstrukcie TPR-CS nieuporządkowane struktury stanowią 30 %, a ponieważ domeny TPR i CS są strukturalnie uporządkowane, fragment VR1 musi być nieuporządkowany. Uważa się, że domena SGS białka SGT1 jest nieuporządkowana (Lee i wsp., 2004, Botër i wsp., 2007). Domena ta ma długość 84 aminokwasów co stanowi około 22 % długości białka. Przyjmując, że pozostałe 12 % struktur drugorzędowych pochodzi właśnie od domeny SGS, sama domena musi mieć także jakieś fragmenty uporządkowane. Ponieważ dla fragmentu CS-SGS wykazano 5 % wkład helisy α , można wysnuć przypuszczenie, że te 5 % pochodzi właśnie od domeny SGS. W celu sprawdzenia czy, zagęszczenie molekularne może mieć wpływ na fałdowanie się domeny SGS przeprowadzono pomiar dichroizmu kołowego w obecności czynnika zagęszczającego Ficoll 70k, polisacharydu o średniej masie molekularnej 70 kDa powszechnie używanego do naśladowania efektów komórkowego zagęszczenia molekularnego. Niestety ilościowa analiza nie była możliwa do przeprowadzenia, ze względu na wąski zakres zebranych danych CD. Jednak ze zmiany kształtu widma CD widocznej na rysunku 4.4 C, można wywnioskować, że w zagęszczonym środowisku domena SGS ma tendencję do porządkowania się.

4.3 Badania małokątowego rozpraszania promieniowania rentgenowskiego białka SGT1 z jęczmienia

Do tej pory niewiele było wiadomo na temat struktury pełnej długości białka SGT1 oraz jak wygląda dimer tego białka. W celu poznania niskorozdzielczej struktury białka SGT1 z jęczmienia w roztworze w ramach niniejszej pracy wykonane zostały pomiary małokątowego rozpraszania promieniowania synchrotronowego. Pomiary te przeprowadzono na synchrotronie DESY w Hamburgu na linii pomiarowej X33. Zarejestrowano krzywe rozpraszania dla trzech różnych stężeń NaCl 0 M, 0,5 M oraz 1 M. Trzy stężenia soli wybrano w celu zbadania wpływu siły jonowej na dimeryzację białka SGT1. Zarejestrowane krzywe SAXS dla białka po odjęciu rozpraszania przez bufor zaprezentowane są na rysunku 4.5 A. Dla trzech krzywych rozpraszania wyznaczono promień żyrcji z wykresu Guiniera (rysunek 4.5 B). Wraz ze zwiększeniem stężenia soli można zaobserwować spadek wartości rozpraszania dla wartość wektora $S \rightarrow 0$. Może to oznaczać, że białko SGT1 zmienia swój stan oligomeryczny. Z pracy innych autorów wynikało, że białko SGT1 tworzy dimery w roztworach o niskiej sile jonowej (Nyarko i wsp., 2007). Także spadek wartości $I(0)$ można wiązać z przesunięciem się równowagi w stronę tworzenia monomeru. Promień żyrcji obliczony dla białka SGT1 dla trzech stężeń soli wyniósł odpowiednio: 5,52 nm dla 0 M NaCl, 5,40 nm dla 0,5 M NaCl oraz 5,02 nm dla 1 M NaCl. Nie zaobserwowano na krzywych SAXS agregacji oraz elektrostatycznych oddziaływań międzycząsteczkowych, o czym świadczy bardzo dobre dopasowanie danych eksperymentalnych do linii prostej wyznaczonej na podstawie równania Guiniera. Również z analizy zmian wartości promienia żyrcji możemy wysnuć te same wnioski jak z analizy zachowania się wartości $I(0)$. Wraz ze wzrostem siły jonowej promień żyrcji zmienia się od 5,52 nm do 5,02 nm.

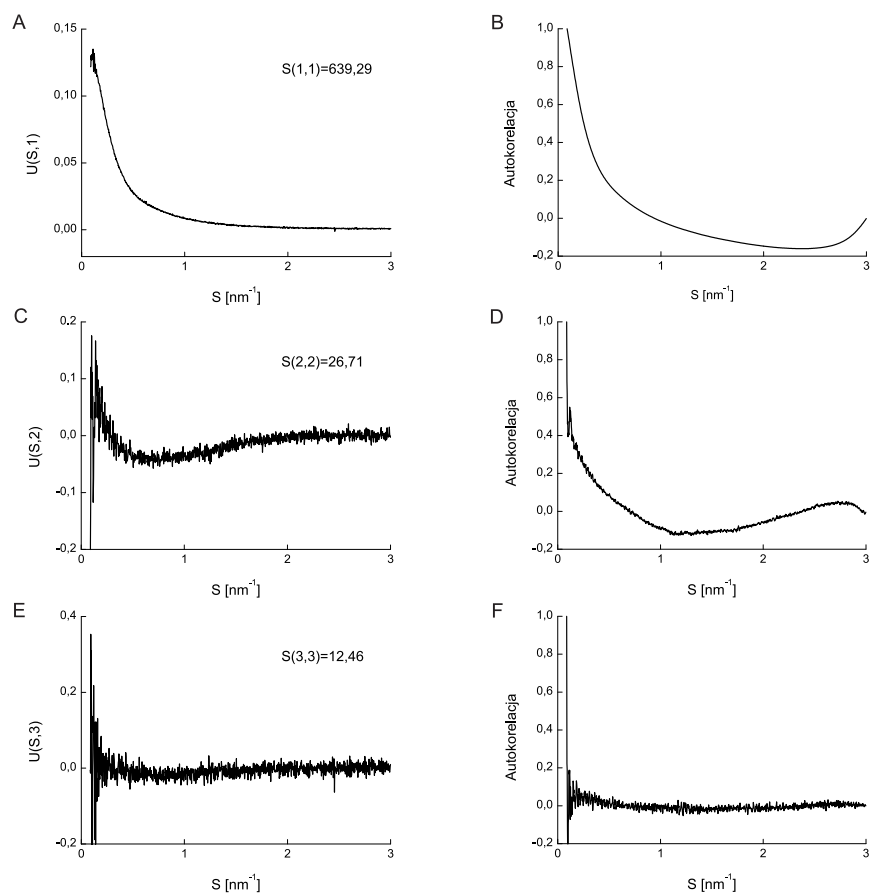


Rysunek 4.5 A. Krzywe rozpraszania dla białka SGT1 z jęczmienia dla trzech stężeń soli 0 M, 0.5 M, 1 M. **B.** Wykres Guiniera i wyznaczony promień żyrcji R_G dla białka SGT1.

4.3.1 Analiza SVD danych z małokątowego rozpraszania promieniowania synchrotronowego

Ponieważ z powyższych analiz jasno wynika, że białko SGT1 w roztworze występuje w równowadze pomiędzy kilkoma stanami oligomerycznymi przeprowadzono analizę SVD w celu sprawdzenia liczby komponentów (stanów oligomerycznych) dających wkład do każdej z krzywych rozpraszania. Analizę SVD oraz funkcje autokorelacji obliczono za pomocą pakietu Matlab R2010. Analiza SVD pozwoliła

oszacować, że zestaw danych eksperymentalnych (trzy krzywe rozpraszania) można przedstawić za pomocą kombinacji liniowej dwóch krzywych - dwóch pierwszych wektorów własnych macierzy U z analizy SVD. Wektory własne macierzy U oraz ich krzywe autokorelacji przedstawione są na rysunku 4.6. Wektory własne U_1 , U_2 oraz U_3 , cechują się wartościami własnymi równymi 639,29, 26,71 oraz 12,46. Ponieważ wartości własne dla wektorów U_2 i U_3 mają podobny rząd wielkości w celu sprawdzenia czy oba wektory dają istotny wkład do macierzy danych obliczono dla nich, jak również dla wektora U_1 , funkcje autokorelacji. Postać wektorów własnych U_1 oraz U_2 oraz ich funkcji autokorelacji wskazuje, że są one różne od szumu w przeciwieństwie do wektora U_3 . Wektor U_3 oraz jego funkcja autokorelacji są płaskie i przypominają losowy szum.

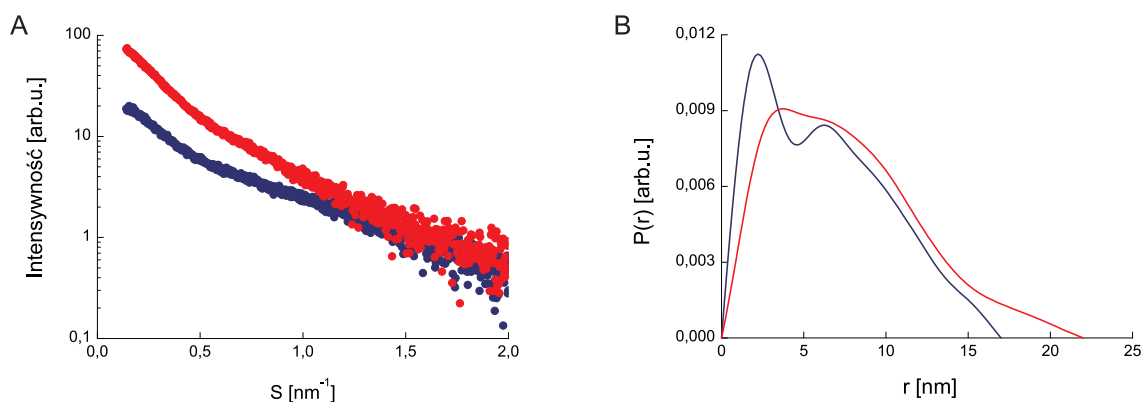


Rysunek 4.6 Analiza SVD krzywych rozpraszania dla białka SGT1 z jęczmienia. **A, C, E** - Wektory własne U_1 , U_2 i U_3 oraz odpowiadające im funkcje autokorelacji **B, D, F** - Na krzywych przedstawiających wektory własne zaznaczone są również wartości własne dla każdego wektora.

Wiadomo również, że funkcja autokorelacji szumu białego jest szumem. Z analizy funkcji autokorelacji możemy wywnioskować, że w serii krzywych SAXS dla pełnej długości białka SGT1 można wyodrębnić dwa komponenty, reprezentowane przez wektory własne U_1 i U_2 , których liniowa kombinacja da w rezultacie krzywe eksperymentalne. Istnienie dwóch komponentów zostało potwierdzone również w programie PRIMUS, który wykonuje analizę SVD i ocenia statystycznie otrzymane wyniki.

4.3.2 Analiza krzywych rozpraszania dla monomeru i dimeru białka SGT1 za pomocą procedury MCR-ALS

Wektory własne U1 oraz U2 nie są jednak prawdziwymi krzywymi rozpraszania. W celu otrzymania krzywych SAXS dla dwóch komponentów posłużono się analizą MCR-ALS. Za pomocą tej metody, otrzymano krzywe dla monomeru oraz dimeru białka SGT1. Analizę MCR-ALS ograniczono do danych w zakresie do 2 nm^{-1} w celu uniknięcia błędów związanych z niskim stosunkiem sygnału do szumu dla $s > 2 \text{ nm}^{-1}$. Kombinacja liniowa tych krzywych daje w rezultacie eksperymentalne krzywe rozpraszania. Na podstawie krzywych rozpraszania w programie GNOM obliczono krzywe rozkładu odległości wewnątrzcząsteczkowych dla monomeru i dimeru białka SGT1. Krzywe rozpraszania oraz znormalizowane do wartości $I(0)$ krzywe $P(r)$ przedstawione są na rysunku 4.7.

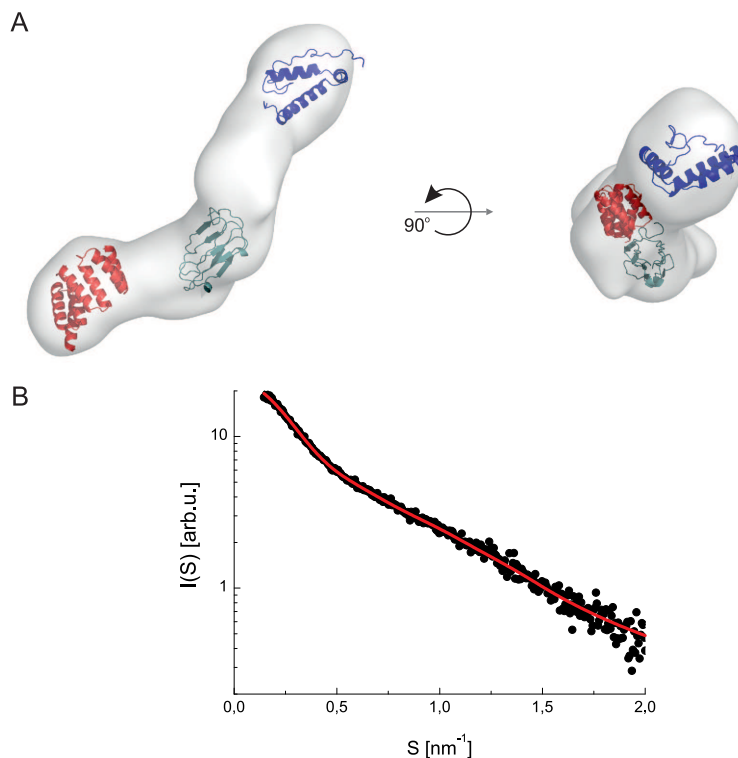


Rysunek 4.7 A. Krzywe rozpraszania dla dimeru (kolor czerwony) i monomeru (kolor granatowy) białka SGT1 z jęczmienia otrzymane z analizy MCR-ALS. **B.** Znnormalizowane krzywe $P(r)$ dla monomeru (kolor granatowy) i dimeru (kolor czerwony) pełnej długości białka SGT1 z jęczmienia.

Kształt krzywych rozpraszania zarówno dla dimeru i monomeru pozbawiony jest wyraźnych minimum oraz maksimum jest charakterystyczny dla cząsteczek o wydłużonym kształcie i/lub mających dynamiczną strukturę. Natomiast na krzywej $P(r)$ dla monomeru (rysunek 4.7 B, granatowa krzywa) można wyróżnić dwa wyraźne maksima oraz zarys trzeciego dla większych wartości R . Maksima te można interpretować jako odległości pomiędzy uporządkowanymi domenami białka SGT1. W przypadku dimeru białka SGT1 maksima te są zarysowane dużo mniej wyraźnie choć nadal są widoczne. Ich położenie nieznacznie tylko różni się od położenia w krzywej $P(r)$ dla monomeru. Świadczy to o tym, że prawdopodobnie, przestrzenne ułożenie domen w monomerze jak i dimerze białka SGT1 jest bardzo podobne. Za pomocą programu GNOM wyznaczono także, maksymalny rozmiar cząsteczki (D_{max}). Wynosi on odpowiednio 17 nm dla monomeru oraz 22 nm dla dimeru. Dla kuli o masie białka SGT1 (40997 Da) i średniej gęstości białka $1,37 \text{ g/cm}^3$, maksymalny rozmiar takiej cząsteczki jest równy $\approx 7 \text{ nm}$. Tak więc dla białka SGT1 o masie 41 kDa maksymalny rozmiar cząsteczki 17 nm wskazuje, że białko to ma wydłużony kształt.

4.3.3 Modelowanie *ab-initio* monomeru białka SGT1

W celu poznania struktur białka SGT1 w roztworze posłużono się modelowaniem *ab initio* za pomocą program DAMMIN i GASBOR. Program DAMMIN generuje modele kulkowe, które stanowią niskorozdzielczy model badanego białka w roztworze. Wyniki modelowania przedstawione są na rysunku 4.8.

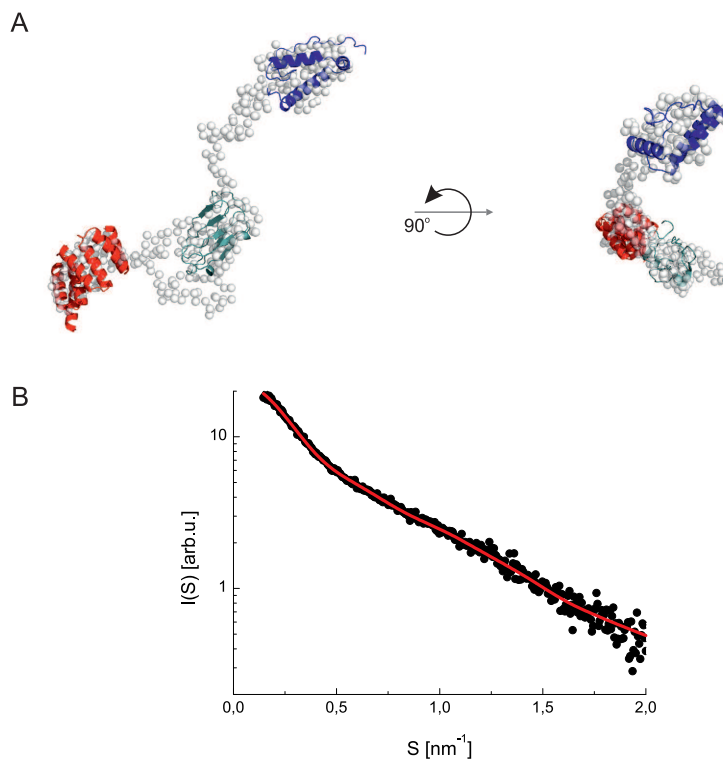


Rysunek 4.8 Modelowanie struktury niskorozdzielczej monomeru białka SGT1 w programie DAMMIN. **A.** Uśredniony model niskorozdzielczej struktury monomeru białka SGT1 w roztworze. **B.** Dopasowanie teoretycznej krzywej rozpraszania dla modelu do krzywej eksperymentalnej. Model przedstawiony jest w postaci powierzchni, na którą nałożone są teoretyczne modele domen: TPR (kolor czerwony), CS (kolor jasno niebieski), SGS (kolor niebieski).

Otrzymany model monomeru białka SGT1, zgodnie z przewidywaniami, ma wydłużony kształt. Na model zostały nałożone teoretyczne struktury domen białka SGT1 otrzymane w programie Quark (Xu i wsp., 2012). Z rysunku 4.8 A wynika, że rozmiary i kształt domen pasują do modelu otrzymanego z programu DAMMIN. W roztworze białko SGT1 ma rozciągniętą konformację, a każda z domen stanowi indywidualny moduł. Na końcach białka znajdują się domeny TPR, na N-końcu oraz SGS na C-końcu. Domena CS stanowi centralną domenę białka SGT1. Pomiędzy domenami jest miejsce, gdzie mogą znajdować się fragmenty VR1 i VR2. Odległości pomiędzy domenami TPR, CS, SGS są na tyle duże, że fragmenty zmienne muszą być rozciągnięte. Nawet jeśli, umiejscowienie domen w modelu jest arbitralne, to i tak jasno potwierdza tezę, że fragmenty VR1 i VR2 są nieuporządkowane

i rozciągnięte. Również przegięcie w środkowej części modelu, każe przypuszczać, że jest to miejsce gdzie najprawdopodobniej znajduje się domena CS.

W celu potwierdzenia przypuszczeń co do struktury fragmentów VR1 i VR2 posłużono się drugim programem do modelowania *ab initio*: GASBOR. Modele otrzymywane w programie GASBOR mają strukturę “łańcucho-podobną”, gdzie odległość pomiędzy pseudo-atomami jest równa odległości pomiędzy aminokwasami w białku. Wyniki modelowania przedstawione są na rysunku 4.9.

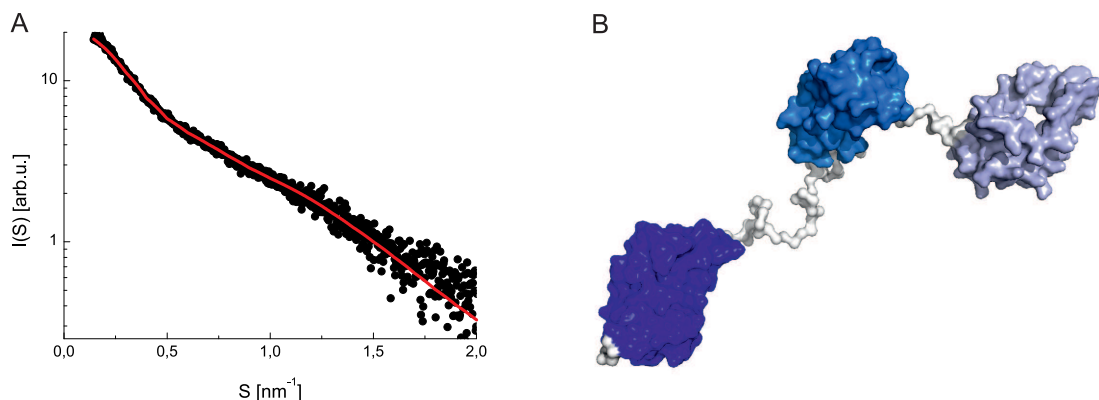


Rysunek 4.9 Modelowanie struktury niskorozdzielczej monomeru białka SGT1 w programie GASBOR. **A.** Reprezentatywny model niskorozdzielczej struktury monomeru białka SGT1. **B.** Dopasowanie teoretycznej krzywej rozpraszania dla modelu do krzywej eksperymentalnej. Na model kulkowy nałożone są teoretyczne modele domen: TPR (kolor czerwony), CS (kolor jasno niebieski), SGS (kolor niebieski).

W modelu “kulkowym” widocznym na rysunku 4.9 C można zauważyć trzy, wyraźnie zarysowane globularne struktury, na które nałożone są modele dla trzech domen białka SGT1. Tak, więc również w modelowaniu za pomocą programu GASBOR, fragmenty VR1 i VR2 mają rozciągniętą konformację. Co więcej, kształt modelu “kulkowego” jest bardzo podobny do modelu otrzymanego z programu DAMMIN, z wyraźnym przegięciem w środkowej części, gdzie znajdują się domena CS. Podobnie, tak jak dla modelu z programu DAMMIN, dopasowanie teoretycznej krzywej rozpraszania dla modelu z programu GASBOR bardzo dobrze pasuje do krzywej eksperymentalnej.

Modele otrzymane z programów DAMMIN i GASBOR są gruboziarniste. Mając do dyspozycji modele teoretyczne dla domen SGT1 przeprowadzono modelowanie bryły sztywnej w programie

BUNCH (Petoukhov i wsp., 2005) by otrzymać dokładniejszy model monomeru białka SGT1. Otrzymany model wraz z dopasowaniem do danych eksperymentalnych przedstawiony jest na rysunku 4.10.

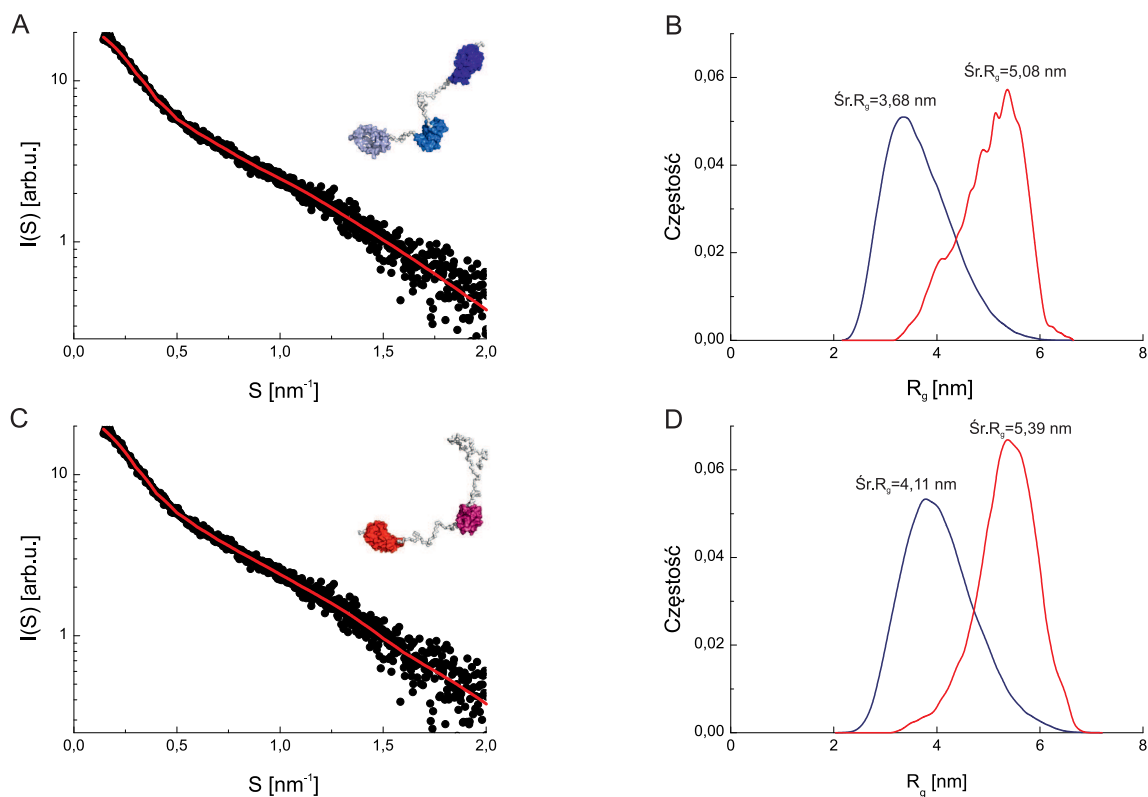


Rysunek 4.10 Modelowanie metodą bryły sztywnej monomeru białka SGT1. **A.** Dopasowanie krzywej teoretycznej do danych eksperymentalnych. **B.** Model otrzymany w programie BUNCH. Kolorem niebieskim oznaczona jest doman TPR, kolorem morskim domena CS a kolorem jasno fioletowym domena SGS.

Otrzymany model jest bardzo podobny do modeli niskorozdzielczych uzyskanych za pomocą programów DAMMIN i GASBOR i tak samo jak modele otrzymane za ich pomocą, model otrzymany w programie BUNCH posiada charakterystyczne przegięcie w obrębie domeny CS. Jednak w dopasowaniu krzywej teoretycznej do danych eksperymentalnych można zauważyć, że dopasowanie nie jest zbyt dobre dla danych w zakresie od $1,5 \text{ nm}^{-1}$ do 2 nm^{-1} . Może się to wiązać z tym że, białko SGT1 występuje w roztworze w kilku konformacjach o zbliżonym kształcie. Nie mniej jednak jasno widać, że wyniki modelowania w programach GASBOR, DAMMIN i BUNCH są zbieżne.

Ponieważ fragmenty VR1 oraz VR2 mają rozciągniętą konformację istnieje duże prawdopodobieństwo, że konformacja monomeru białka SGT1 otrzymana w programach GASBOR, DAMMIN i BUNCH jest jedynie uśrednioną konformacją a samo białko wykazuje dynamiczne zachowanie w roztworze co tłumaczyłoby też kształt krzywej rozpraszania dla modelu z programu BUNCH. W celu sprawdzenia powyższej hipotezy posłużono się analizą EOM. Wyniki analizy EOM przedstawione są na rysunku 4.11. Analizę EOM przeprowadzono dla dwóch konfiguracji molekularnych. W pierwszej, do analizy użyto modeli teoretycznych dla wszystkich trzech domen białka SGT1, przykładowy model przedstawiony jest w prawym, górnym rogu rysunku 4.11 A. W drugiej konfiguracji domenę SGS modelowano za pomocą modelu łańcucho-podobnego tak jak fragmenty VR1 i VR2 w celu sprawdzenia czy konformacja domeny SGS ma istotny wpływ na kształt krzywej rozpraszania i rozkład promieni żyracji. Dla obu konfiguracji krzywa rozpraszania jest bardzo podobna i pasuje bardzo dobrze do danych eksperymentalnych. Również rozkład promienia żyracji w obu przypadkach jest podobny z tą różnicą, że dla drugiej konfiguracji rozkład jest "gładszy". Kształt jest natomiast ten sam i zasadniczo różni się od rozkładu dla 10000 losowych konformacji. Maksimum przypada na

około 5,5 nm - 5,8 nm i szybko maleje wraz ze wzrostem R_g . Poniżej 5,5 nm rozkład maleje wolniej i jest szerszy. Z rozkładu wynika, że białko SGT1 ma dynamiczny charakter w roztworze i występuje głównie w rozciągniętej konformacji.

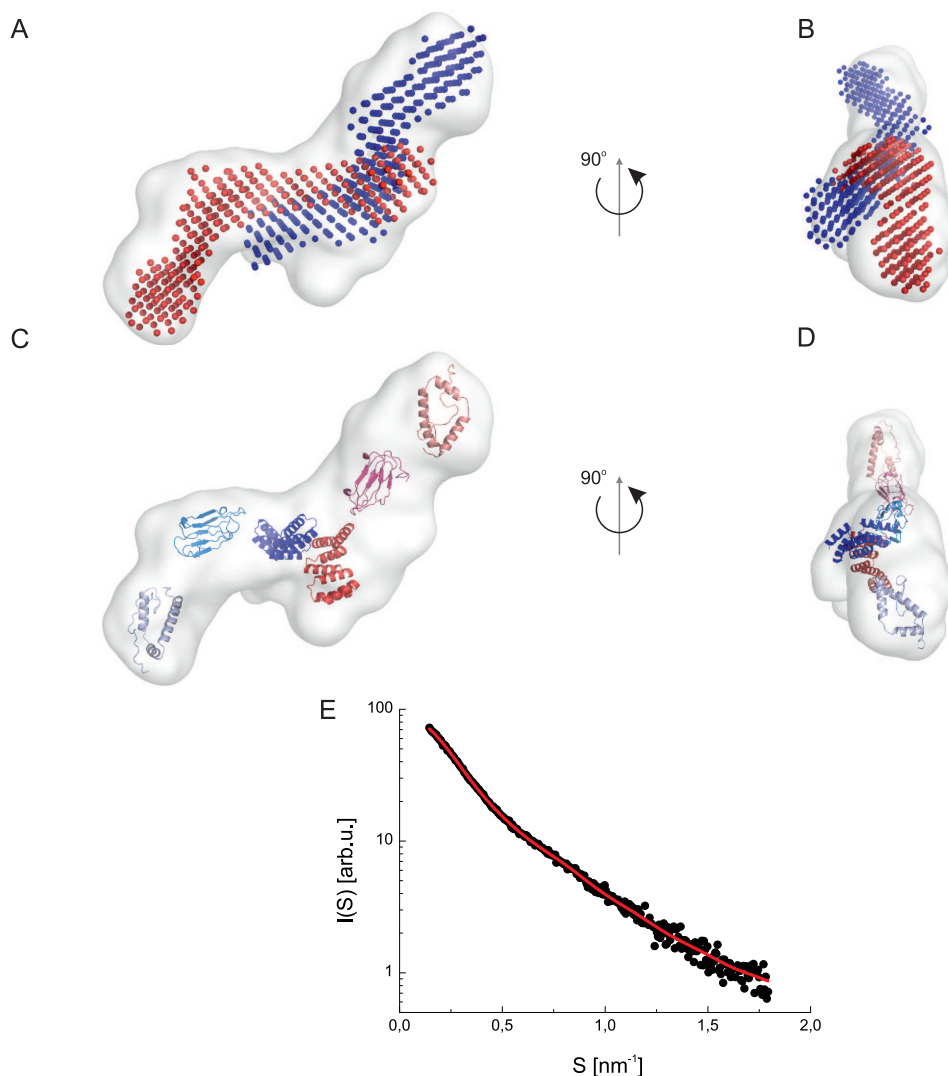


Rysunek 4.11 Analiza EOM monomeru białka SGT1. **A, C** - Dopasowanie teoretycznej krzywej rozpraszania dla zespołu otrzymanej z analizy EOM. **B, D** - Rozkład promienia żyracji R_g dla losowej puli (kolor niebieski) oraz dla wyselekcjonowanego zespołu (kolor czerwony). Na wykresach **A** i **B** przedstawione są wyniki dla analizy EOM przeprowadzonej dla modelu białka SGT1, gdzie domena SGS przedstawiona jest za pomocą modelu teoretycznego otrzymanego w programie Quark. Na wykresach **C** i **D** przedstawione są wyniki dla analizy EOM przeprowadzonej dla modelu białka SGT1, gdzie domena SGS była modelowana w takim samym sposób jak fragmenty VR1 i VR2.

Średni promień żyracji dla konfiguracji z modelem domeny SGS, otrzymany z wybranego zestawu struktur ma wartość 5,08 nm i jest bardzo podobny do wartości otrzymanej dla białka SGT1 w 1 M NaCl. Natomiast dla drugiej konfiguracji promień żyracji wyniósł 5,38 nm. Wynik ten, może wskazywać, że domena SGS jest jednak w pewnym stopniu uporządkowana.

4.4 Badanie struktury dimeru białka SGT1 w roztworze za pomocą małokątowego rozpraszania promieniowania synchrotronowego

W celu zbadania struktury dimeru białka SGT1 również posłużono się modelowaniem *ab initio* w programie DAMMIN. Do modelowania wykorzystano krzywą SAXS zrekonstruowaną za pomocą analizy MCR-ALS. Wynik modelowania przedstawiony jest na rysunku 4.12. Model dimeru białka SGT1 ma wydłużony kształt. Na model dimeru białka SGT1 otrzymanego z programu DAMMIN można nałożyć dwa modele monomeru białka SGT1 otrzymane za pomocą programu GASBOR (rysunek 4.12 A, B). Z nałożenia modeli monomeru białka SGT1 widać, że w dimerze białko SGT1 ma bardzo podobną konformację do monomerycznego białka SGT1.



Rysunek 4.12 Modelowanie struktury niskorozdzielczej dimeru białka SGT1 przy pomocy programu DAMMIN. **A, B** - Model białka SGT1 (szara powierzchnia) z nałożonymi uśrednionymi kulkowymi modelami monomeru białka SGT1 z programu GASBOR (kolor niebieski i czerwony).

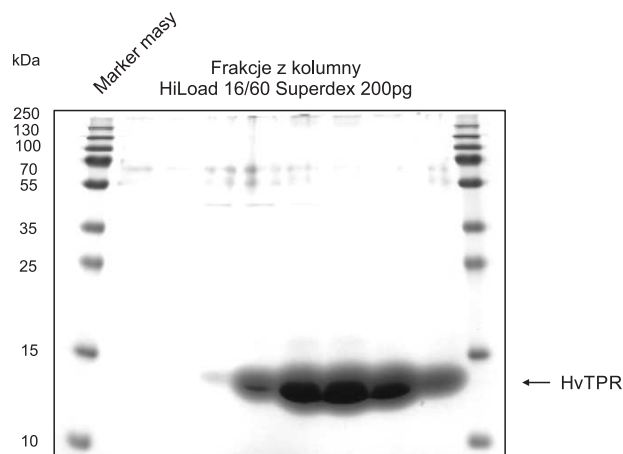
C, D - Model białka SGT1 (szara powierzchnia) z nałożonymi modelami dla domen białka SGT1 (modele dla jednego monomeru mają kolor w odcieniach niebieskiego, dla drugiego monomeru mają kolor w odcieniach czerwonego). **E** - Dopasowanie teoretycznej krzywej rozpraszania dla modelu dimeru białka SGT1 do danych eksperymentalnych.

W strukturze dimeru widoczne jest także, charakterystyczne zgięcie w miejscu gdzie powinna znajdować się domena CS. Wynik ten potwierdza, że model struktury niskorozdzielczej otrzymany dla dimeru białka SGT1 jest poprawny. Na rysunku 4.12 C, D nałożone są modele strukturalne dla domen białka SGT1. Dimer domen TPR stanowi centralny punkt w strukturze białka SGT1, z którego rozchodzi się reszta białka, odpowiedzialna za oddziaływanie z innymi białkami.

4.4.1 Badanie dimeryzacji domeny TPR białka SGT1 z jęczmienia

4.4.1.1 Oczyszczanie domeny TPR białka HvSGT1

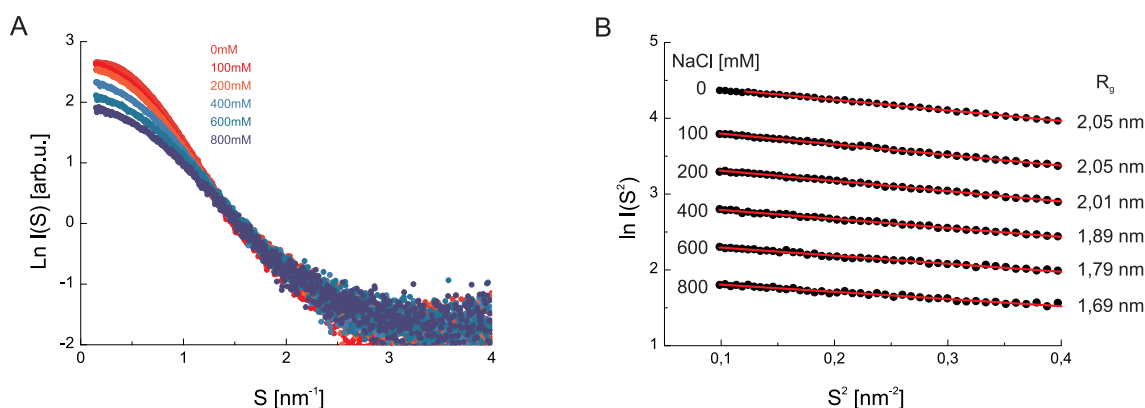
Do badań dimeryzacji domeny TPR białka HvSGT1 wybrano fragment, który jest najbardziej zakonserwowany w sekwencji pomiędzy białkami SGT1 z różnych organizmów (Azevedo i wsp., 2002). Fragment ten zawierał aminokwasy 6-124 z pełnej długości białka SGT1 z jęczmienia. Białko HvTPR oczyszczano z 1,6 l bakterii *E. coli* BL21 CodonPlus wg. protokołów 3.7.1-3.7.2 za pomocą chromatografii powinowactwa na złożu ze związanymi jonami niklu Ni^{2+} , cięcia proteazą TEV, ponownym oczyszczaniu za pomocą chromatografii powinowactwa oraz sączenia molekularnego. Po ostatnim etapie oczyszczania otrzymano białko HvTPR o czystości odpowiedniej do badań strukturalnych. Na rysunku 4.13 pokazana jest czystość frakcji białka HvTPR po sączeniu molekularnym na kolumnie ze złożem Superdex 200 w postaci zdjęcia elektroforezy w żelu poliakrylamidowym.



Rysunek 4.13 Oczyszczanie domeny TPR białka HvSGT1 na kolumnie Hiload 16/60 Superdex 200pg.

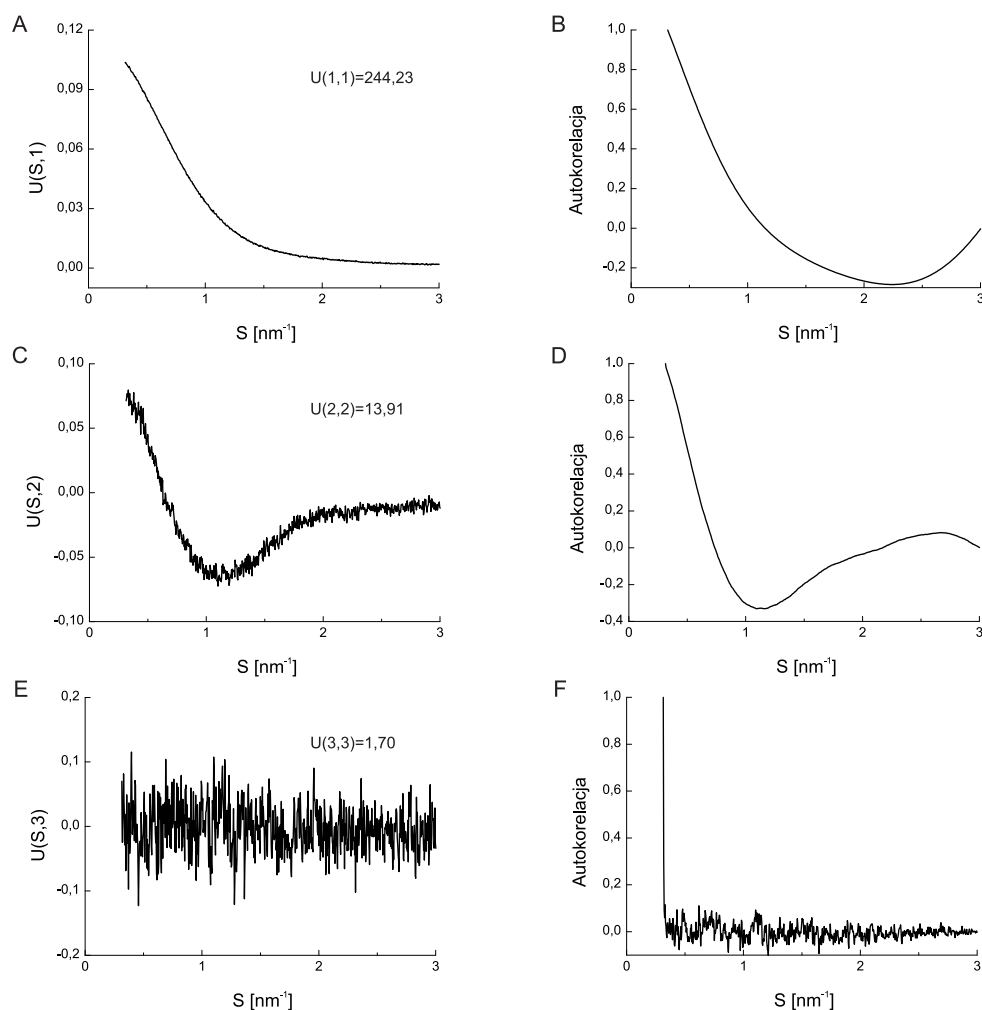
4.4.1.2 Badania małokątowego rozpraszania promieniowania rentgenowskiego białka HvTPR

Model dimeru białka SGT1 otrzymany w programie DAMMIN jest modelem niskorozdzielczym. Brak struktury lub modelu dimeru domeny TPR uniemożliwia modelowanie za pomocą metody bryły sztywnej. W celu zbadania jaką strukturę ma dimer domeny TPR w roztworze przeprowadzono eksperymenty małokątowego rozpraszania promieniowania synchrotronowego dla izolowanej domeny TPR z białka SGT1 z jęczmienia. Dimeryzacja jest zależna od siły jonowej dlatego też przeprowadzono eksperymenty w zakresie stężenia NaCl od 0 mM do 800 mM. Krzywe SAXS dla domeny TPR oraz odpowiadające im wykresy Guiniera wraz z wyznaczonym promieniem żyracji R_g w różnych warunkach siły jonowej przedstawione są na rysunku 4.14. Promień żyracji zmniejsza się od wartości 2,05 nm dla 0 mM NaCl do wartości 1,69 nm dla 800 mM NaCl.



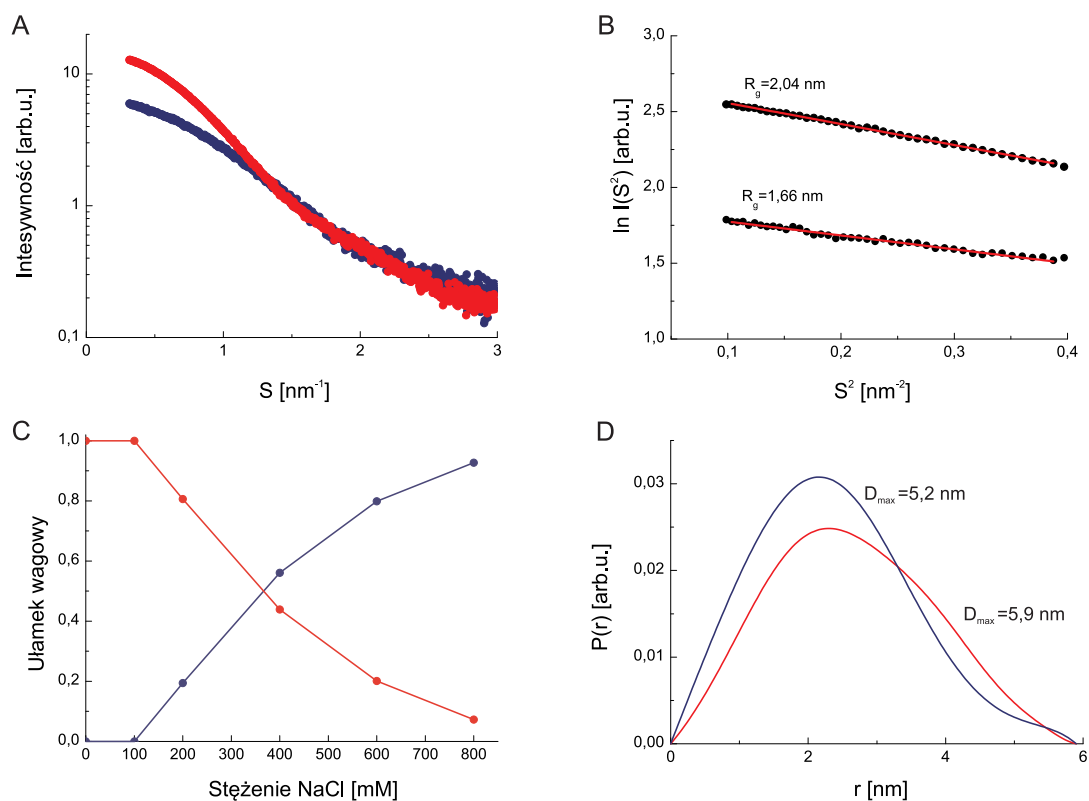
Rysunek 4.14 **A.** Krzywe SAXS dla domeny TPR białka SGT1 z jęczmienia zarejestrowane w różnych stężeniach NaCl. **B.** Wykresy Guiniera i wyznaczony promień żyracji R_g dla domeny TPR białka SGT1.

Dla zwiększającej się siły jonowej zarówno $I(0)$ jak i promień żyracji maleją. Ponieważ pełnej długości białko SGT1 występuje w równowadze monomer-dimer, zachowanie $I(0)$ oraz R_g jest wynikiem zmiany stanu oligomerycznego domeny TPR. Zachowanie się domeny TPR jest więc analogiczne do zachowania się pełnej długości białka SGT1. W celu sprawdzenia, czy domena TPR może występować w kilku stanach oligomerycznych, przeprowadzono także analizę SVD, której wynik przedstawiony jest na rysunku 4.15. Na rysunku 4.15 A, C, E widoczne są trzy pierwsze wektory własne. Wektory U_1 i U_2 są znacząco różne od szumu w przeciwieństwie do wektora U_3 . Również funkcje autokorelacji dla wektorów U_1 i U_2 są różne od szumu w przeciwieństwie do funkcji autokorelacji dla wektora U_3 . Kształt funkcji autokorelacji jak również wartości własne wektorów (dla U_1 : 244,23; dla U_2 : 13,91; dla U_3 : 1,70) wskazują, że tylko dwa pierwsze wektory są istotne. Można więc przypuszczać, że podobnie jak pełnej długości białko, domena TPR występuje w równowadze monomer-dimer.



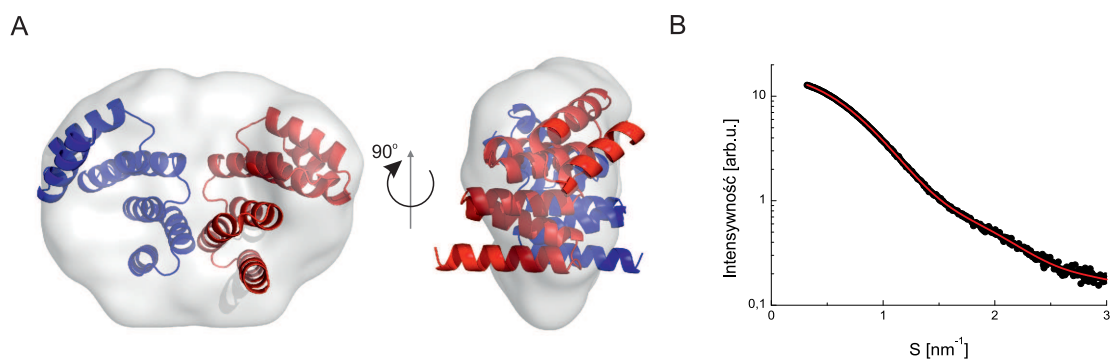
Rysunek 4.15 Analiza SVD krzywych SAXS dla domeny TPR białka SGT1 z jęczmienia. **A**, **C**, **E** - Wektory własne U_1 , U_2 i U_3 oraz odpowiadające im funkcje autokorelacji **B**, **D**, **F** - Na krzywych przedstawiających wektory własne zaznaczone są również wartości własne dla każdego wektora.

W celu otrzymania krzywych rozpraszania dla monomeru i dimeru domeny TPR białka SGT1 przeprowadzono analizę MCR-ALS. Otrzymane krzywe rozpraszania wraz z obliczonymi krzywymi rozkładu odległości wewnątrzcząsteczkowych przedstawione są na rysunku 4.16 A, D. Dla otrzymanych krzywych, z wykresu Guiniera (rysunek 4.16 B) wyznaczono promień żyracji R_g . Wartość R_g dla monomeru wynosiła 1,66 nm a dla dimeru 2,04 nm. Obie wartości są bardzo podobne do skrajnych wartości R_g otrzymanych dla krzywych rozpraszania w różnych stężeniach soli. Dla 0 mM oraz 100 mM NaCl, jak wynika z wykresu na rysunku 4.16 C, domena TPR tworzy jedynie dimery. Wraz ze wzrostem stężenia NaCl równowaga przesuwa się w stronę monomeru. Dla 800 mM NaCl 93% (wagowo) stanowi monomer domeny TPR.

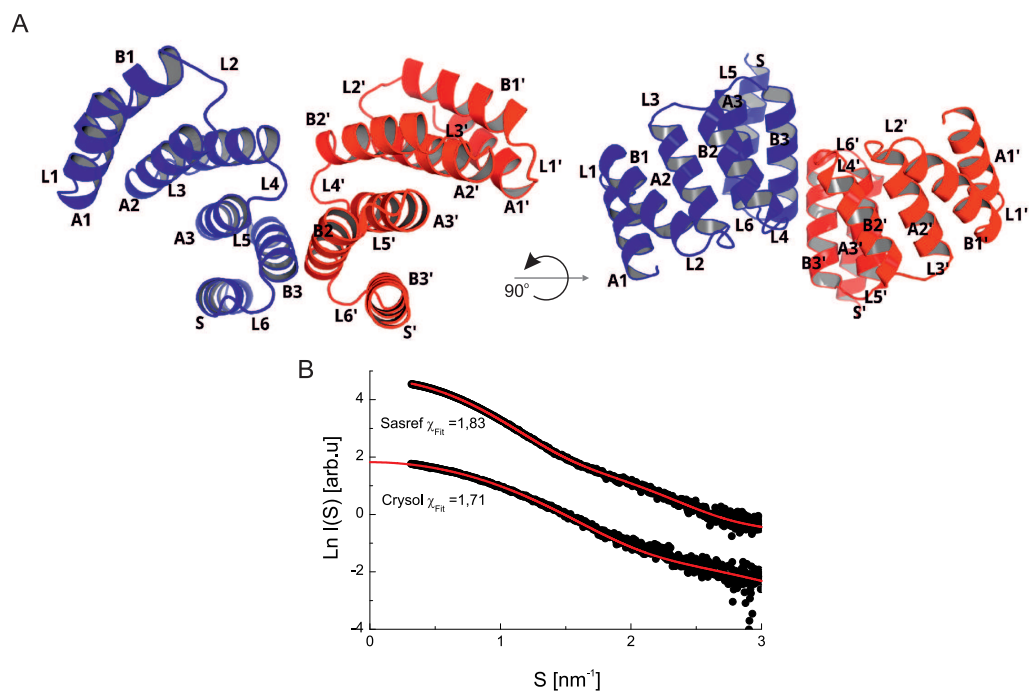


Rysunek 4.16 Analiza MCR-ALS krzywych SAXS domeny TPR białka SGT1 z jęczmienia. **A.** Krzywe rozpraszania dla monomeru (kolor niebieski) i dimeru (kolor czerwony) domeny TPR. **B.** Wykres Guiniera i wyznaczony promień żyrcacji R_g dla dimeru (górną krzywą) i monomeru (dolną krzywą) domeny TPR. **C.** Zależność frakcji monomer-dimer domeny TPR w roztworze o różnym stężeniu NaCl (kolor czerwony-dimer, kolor niebieski monomer). **D.** Krzywe $P(r)$ dla monomeru (kolor niebieski) i dimeru (kolor czerwony) domeny TPR.

Otrzymaną krzywą rozpraszania dla dimeru domeny TPR wykorzystano do modelowania za pomocą programu DAMMIN i SASREF. W programie DAMMIN otrzymano uśredniony model kulkowy, który przedstawiony jest wraz z dopasowaniem do krzywej rozpraszania dla dimeru domeny TPR białka SGT1 na rysunku 4.17. Uśredniony model dimeru domeny TPR ma kształt spłaszczonego dysku. Na model otrzymany w programie DAMMIN można nałożyć model dimeru domeny TPR otrzymany w programie SASREF.



Rysunek 4.17 Modelowanie dimeru domeny TPR w programie DAMMIN. **A.** Model dimeru domeny TPR (szara powierzchnia) z nałożonym na niego modelem z programu SASREF (kolor czerwony i niebieski). **B.** Dopasowanie do krzywej eksperymentalnej.

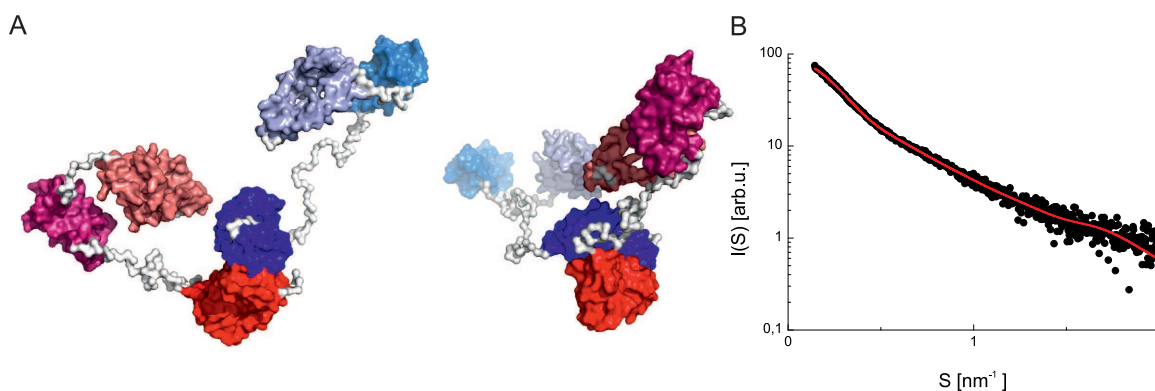


Rysunek 4.18 Modelowanie dimeru domeny TPR w programie SASREF. **A.** Model dimeru domeny TPR białka SGT1 otrzymany w programie SASREF. **B.** Dopasowanie dla modelu otrzymanego w programie SASREF do krzywej dla dimeru domeny TPR oraz dopasowanie dla modelu monomeru domeny TPR wymodelowanego w programie Quark do krzywej dla monomeru domeny TPR białka SGT1. Helisy jednego powtórzenia TPR (helisa-pętla-helisa) w domenie TPR, oznaczone są kolejno jako A i B oraz numerem tego powtórzenia, pętle pomiędzy helisami kolejnymi numerami a helisa solwatująca jako S.

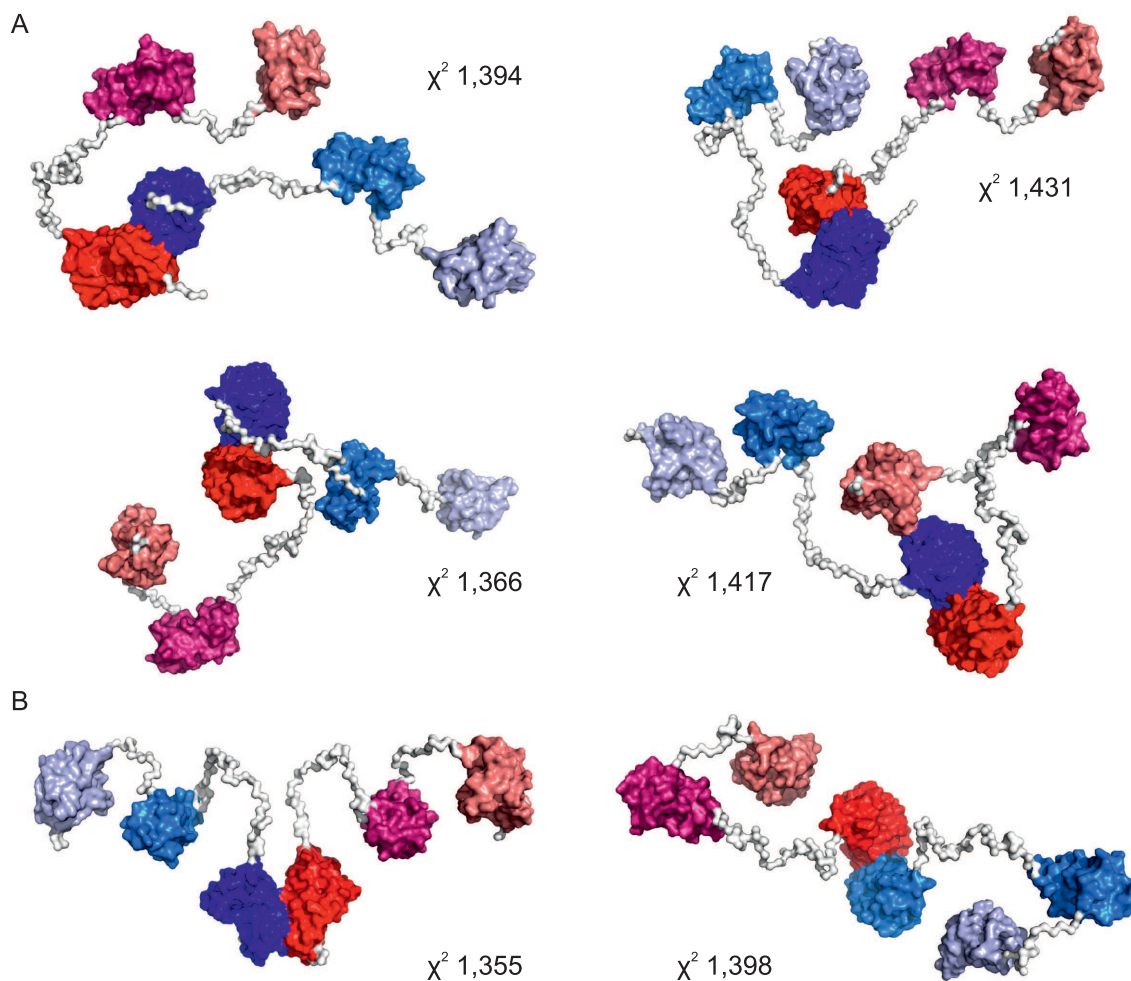
Wymiarami oba modele pasują do siebie, aczkolwiek model z programu DAMMIN jest bardziej płaski. Może to wynikać z uśrednienia dziesięciu otrzymanych indywidualnych modeli. Również kształt modelu z programu DAMMIN jest mniej szczegółowy, nie mniej jednak można np. zauważyć zwięźanie się modelu kulkowego w miejscu gdzie w helisy B3 oddziałują ze sobą. Ponieważ domena TPR jest małym białkiem, rozdzielczość danych może być niewystarczająca do dokładniejszego modelowania. Wewnętrzna, wklęsła strona domeny TPR jest skierowana przeciwnie w obu monomerach dimeru domeny TPR. Taka konformacja pozwala na jednoczesną dimeryzację i wiązanie potencjalnych ligandów przez wklęsłą stronę domeny TPR. Na rysunku 4.18 przedstawiony jest model otrzymany metodą bryły sztywnej w programie SASREF. W modelu tym domeny TPR oddziałują ze sobą głównie poprzez helisy B3 należące do trzeciego powtórzenia TPR. Zarówno model otrzymany w programie SASREF, jak również model monomeru domeny TPR bardzo dobrze pasują do danych eksperymentalnych, jak wynika z dopasowania krzywych dla modeli do danych eksperymentalnych (rysunek 4.18 B).

4.4.2 Modelowanie dimeru białka SGT1 pełnej długości

Mając do dyspozycji model dimeru domeny TPR, można było wymodelować strukturę pełnej długości białka SGT1 w formie dimerycznej za pomocą metody modelowania bryły sztywnej. Podczas modelowania struktura dimeru domen TPR była “zablokowana”, a reszta domen była modelowana jako bryły sztywne. Regiony VR1 i VR2 modelowane były za pomocą modelu pseudo-łańcucha. Wynik modelowania przedstawiony jest na rysunku 4.19.



Rysunek 4.19 Modelowanie dimeru pełnej długości białka SGT1. **A.** Model dimeru białka SGT1 otrzymany w programie CORAL (Petoukhov i wsp., 2005). Przedstawione są dwie różne orientacje modelu. **B.** Dopasowanie krzywej rozpraszania dla modelu do danych eksperymentalnych. Na rysunku A domeny poszczególnych podjednostek zaprezentowane są w odcieniach koloru czerwonego lub niebieskiego. Fragmenty VR1 oraz VR2 mają szary kolor.



Rysunek 4.20 Modelowanie dimeru białka SGT1 w programie CORAL. **A.** Cztery przykładowe modele otrzymane w programie CORAL przy braku założonej symetrii lokalnej modelu. **B.** Dwa przykładowe modele otrzymane w programie CORAL przy założonej symetrii P2 modelu. Na każdym rysunku podana jest wartość dopasowania χ^2 krzywej dla modelu do danych eksperymentalnych.

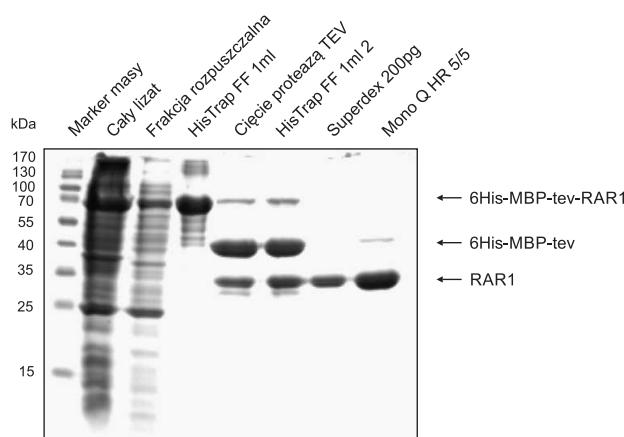
Model dimeru białka SGT1 ma rozciągnięty kształt, a regiony VR1 i VR2 są nieuporządkowane tak samo jak w przypadku monomeru. W roztworze, również dimer białka SGT1 nie ma jednej, określonej struktury i może przyjmować wiele konformacji. Na rysunku 4.20 przedstawione są modele otrzymane w programie CORAL. Modele te różnią się między sobą ogólną konformacją, jednak we wszystkich modelach fragmenty VR1 i VR2 są nieuporządkowane. Konformacja ta nie jest zależna od tego czy podczas modelowania narzucona była symetria lokalna P2 czy nie. Dla wszystkich modeli dopasowanie do danych eksperymentalnych było w przedziale wartości χ^2 od 1,355 do 1,431. Niestety ze względów na brak możliwości symulacji struktury niesymetrycznych oligomerów w programie

EOM nie można było przeprowadzić takiej samej analizy jak dla monomeru białka SGT1. Niemniej jednak, podobnie jak w przypadku monomeru dopasowanie do krzywej eksperymentalnej najbardziej odstaje dla danych SAXS w okolicy $S \approx 2 \text{ nm}^{-1}$, gdzie krzywa eksperymentalna jest bardziej płaska. Otrzymane dane i ich analiza wskazują na to, że zarówno monomer i dimer białka SGT1 w roztworze ma rozciągniętą konformację, a nieuporządkowane regiony VR1 i VR2 powodują, że białko SGT1 ma dynamiczną konformacyjną w roztworze.

4.5 Badania struktury i konformacji w roztworze białka RAR1 z jęczmienia

4.5.1 Oczyszczanie białka RAR1 z bakterii *E. coli*

Rekombinowane białko RAR1 z jęczmienia do pomiarów małokątowego rozpraszania promieniowania synchrotronowego oczyszczano z bakterii *E. coli* zgodnie z następującą procedurą: oczyszczanie za pomocą chromatografii powinowactwa, cięcie proteazą TEV, kolejny raz oczyszczanie chromatografią powinowactwa w celu usunięcia proteazy TEV, sączenie molekularne, chromatografia jonowymienna. Czystość rekombinowanego białka RAR1, podczas kolejnych kroków oczyszczania przedstawiona jest na rysunku 4.21.



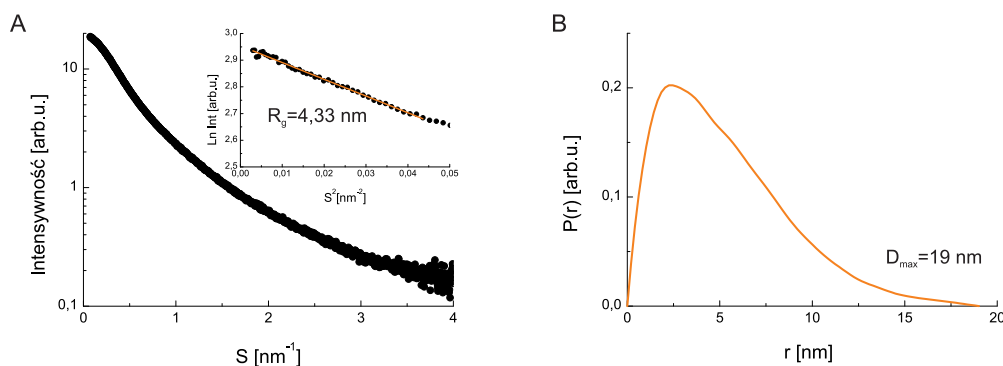
Rysunek 4.21 Etapy oczyszczania białka RAR1 z jęczmienia. Żel SDS-PAGE przedstawiający czystość białka RAR1 podczas kolejnych kroków oczyszczania.

Jak wynika z rysunku 4.21 czystość białka RAR1 po końcowym etapie jest bardzo dobra z jedynie marginalnym zanieczyszczeniem w postaci białka MBP.

4.5.2 Badania małokątowego rozpraszania promieniowania synchrotronowego białka RAR1

W celu zbadania struktury i konformacji białka RAR1 z jęczmienia w roztworze posłużono się tak jak w przypadku białka SGT1 metodą małokątowego rozpraszania promieniowania synchrotronowego. Zarejestrowano dwie krzywe rozpraszania dla stężenia białka RAR1 3,01 mg/ml oraz 6,03 mg/ml. Ponieważ obie krzywe nie różnią się od siebie i nie obserwowano żadnych efektów związanych ze stężeniem białka do analizy wykorzystano krzywą dla większego stężenia, która wykazywała większy stosunek sygnału do szumu. Krzywa SAXS dla próbki białka RAR1 z jęczmienia o stężeniu 6,03 mg/ml zaprezentowana jest na rysunku 4.22 A. Podobnie jak dla białka SGT1 krzywa rozpraszania pozbawiona jest wyraźnych maksimów i minimów i posiada charakterystyczny kształt dla białka o wydłużonym kształcie. Wyznaczony z wykresu Guiniera (rysunek 4.22 A, mniejszy wykres) promień żyrcacji wynosi 4,33 nm. Promień żyrcacji RAR1 jest nie wiele mniejszy niż R_g dla monomeru białka

SGT1. Białko RAR1 jak wykazano wcześniej jest monomerem o masie molekularnej 28 kDa w roztworze i nie wykazuje tendencji do tworzenia oligomerów. Wartość R_g sugeruje, że białko RAR1 ma rozciągniętą konformację w roztworze. Wyznaczony z funkcji rozkładu odległości wewnątrzcząsteczkowych $P(r)$ maksymalny rozmiar cząsteczki $D_{max} = 19$ nm, jest bardzo duży jak dla białka o masie 28 kDa. Funkcja $P(r)$ ma maksimum dla wartości R równego około 2,5 nm - 3 nm. Powyżej 3 nm funkcja $P(r)$ maleje aż do $D_{max} = 19$ nm. Wartość funkcji $P(r)$ dla r w przedziale pomiędzy 12,5 nm a 19 nm jest bardzo mała co wskazuje, że białka RAR1 może występować w kilku konformacjach i jedynie niewielka część populacji osiąga D_{max} równy 19 nm.

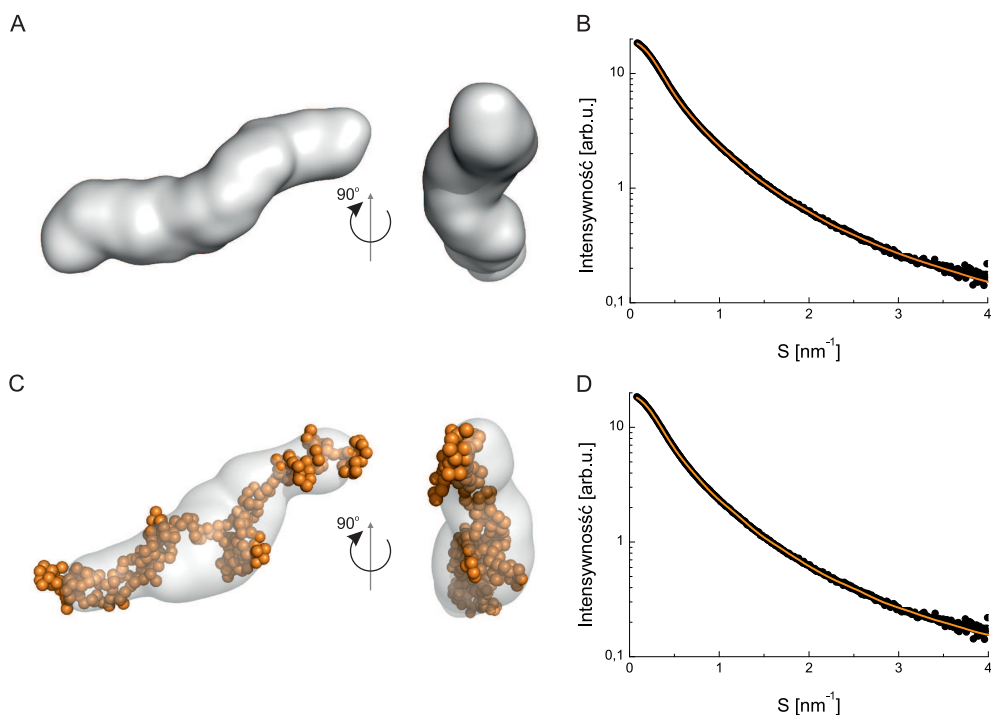


Rysunek 4.22 Eksperyment małokątowego rozpraszania promieniowania synchrotronowego dla białka RAR1. **A.** Krzywa rozpraszania dla białka RAR1 z jęzmiem i wykresem Guiniera i wyznaczonym promieniem żyrcacji R_g . **B.** Rozkład odległości wewnątrzcząsteczkowych dla białka RAR1 wraz z wyznaczonym maksymalnym rozmiarem cząsteczki D_{max} .

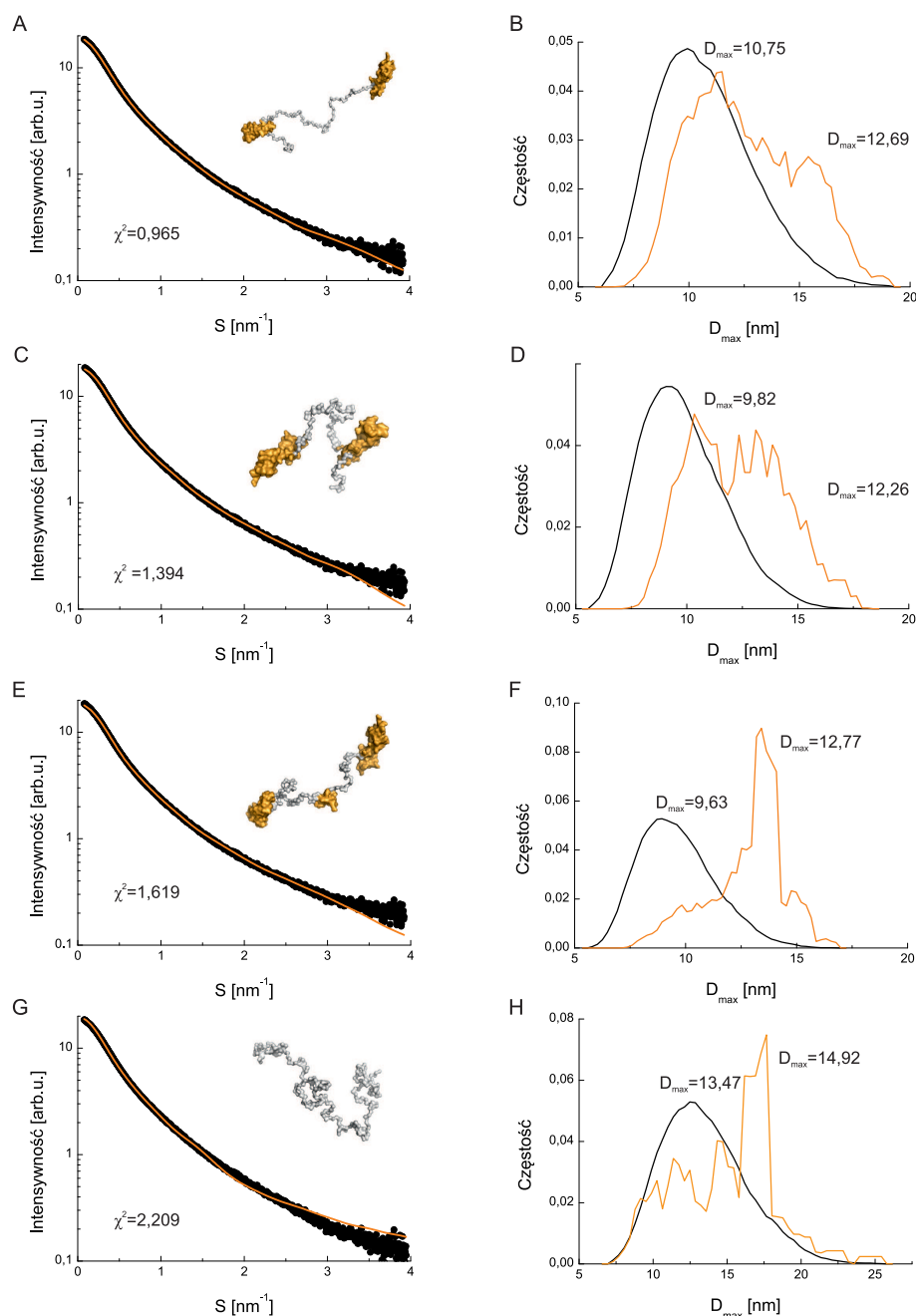
4.5.3 Modelowanie *ab initio* struktury białka RAR1 w roztworze

W celu zbadania struktury białka RAR1 w roztworze, posłużono się modelowaniem *ab initio* w programach DAMMIN i GASBOR. Wyniki modelowania zaprezentowane są na rysunku 4.23. Zarówno dla uśrednionego modelu z programu DAMMIN jak i GASBOR, białko RAR1 ma rozciągniętą konformację. Również pojedyncze modele o najlepszym dopasowaniu do danych eksperymentalnych mają rozciągniętą konformację. W roztworze białko RAR1 występuje więc głównie w postaci rozciągniętej. Nie mniej jednak wyniki te nie dają jednoznacznej odpowiedzi czy białko RAR1 jest pofałdowane czy też nie. W celu zbadania czy białko RAR1 jest w roztworze niepofałdowane i występuje w kilku konformacjach przeprowadzono analizę EOM w różnych konfiguracjach programu. W programie EOM wygenerowano losowe konformacje dla modelu białka RAR1 z domenami CHORD I oraz CHORD II wymodelowanymi za pomocą serwera I-TASSER, domenami CHORD I oraz CHORD II w postaci modelu pseudo-łańcucha oraz zakładając “podobny do natywnego” (naśladujący białka uporządkowane) charakter modelu lub pseudo-niepofałdowany (naśladujący białka nieuporządkowane). Wyniki modelowania przedstawione są na rysunku 4.24. Najlepsze wyniki otrzymano dla białka RAR1 z modelami domen CHORD (otrzymanymi w programie I-TASSER) oraz resztą białka mogącą przyjmować losową konformację (rysunek 4.24 A, B). Wartość dopasowania χ^2 dla wybranej puli modeli wynosiła 0,965. Średni promień żyrcacji wyniósł 4,18 nm (dla losowej puli 3,52 nm), co

jest wartością bardzo zbliżoną do tej wyznaczonej za pomocą wykresu Guiniera. Dla tej konfiguracji programu EOM otrzymano rozkład wartości D_{max} , w którym maksymalna wartość wynosiła około 19 nm, czyli tak jak w przypadku krzywej $P(r)$. Co ciekawe przy innych konfiguracjach programu EOM, maksymalna wartość D_{max} była różna niż 19 nm. Trochę gorszą wartość dopasowania χ^2 , 1,394, otrzymano dla konfiguracji z wykorzystaniem modeli domen CHORD z programu I-TASSER, w której losowa pula miała “podobny do natywnego” charakter (kąty torsyjne dla łańcucha białkowego mają wartości charakterystyczne dla białek uporządkowanych). Wskazuje to na to, że domeny CHORD są połańdowane a reszta białka ma charakter nieuporządkowany. Jeżeli w programie EOM użyjemy modelu dla fragmentu CCCH (ze środkowej części białka) wartość χ^2 wzrośnie do wartości 1,619 w porównaniu do 1,394 dla modelu bez struktury domeny CCCH. Dla tego modelu również rozkład wartości D_{max} ma inny kształt niż dla poprzednich dwóch przypadków z wyraźnym maksimum dla około 13 nm. Najgorzej analiza EOM wypadła dla modelu białka RAR1 w postaci pseudo-łańcucha. Dla konfiguracji natywnej wartość dopasowania χ^2 wynosiła 2,209 (rysunek 4.24 H, G). W przypadku gdy ten sam model białka RAR1 miał charakter nieuporządkowany wartości χ^2 wynosiła 3,118.



Rysunek 4.23 Modelowanie *ab initio* struktury niskorozdzielczej białka RAR1 w roztworze. **A.** Uśredniony model białka RAR1 otrzymany w programie DAMMIN. Model kulkowy zaprezentowany jest w postaci powierzchni. **B.** Dopasowanie krzywej dla najlepszego modelu z programu DAMMIN do danych eksperymentalnych. **C.** Uśredniony model białka RAR1 otrzymany w programie GASBOR (szara pół-przeźroczysta powierzchnia) wraz z nałożonym najlepszym modelem otrzymanym z programu GASBOR (pomarańczowy łańcuch). **D.** Dopasowanie krzywej dla najlepszego modelu do danych eksperymentalnych.



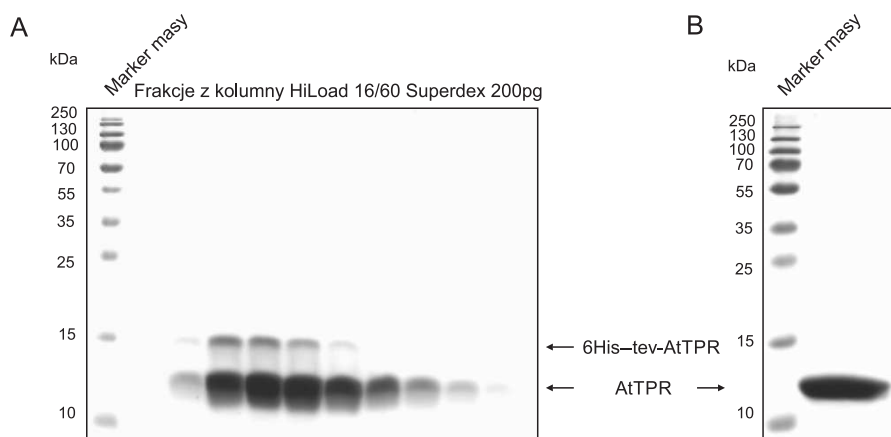
Rysunek 4.24 Analiza EOM modeli białka RAR1 z jęczmienia w roztworze. **A, C, E, G** - Dopasowanie krzywej otrzymanej w wyniku selekcji modeli w programie EOM. Na każdym wykresie podana jest wartość χ^2 określająca jakość dopasowania oraz przykładowy model białka RAR1 (modele dla domen, traktowane jak bryły sztywne, oznaczone są kolorem pomarańczowym) . **B, D, F, H** - Rozkład maksymalnego rozmiaru cząsteczki D_{\max} dla losowej puli (kolor czarny) i wybranego zestawu konformacji (kolor pomarańczowy). Przy każdym rozkładzie podany jest średni D_{\max} dla każdej puli.

4.6 Badanie dimeryzacji domeny TPR białka SGT1a z *Arabidopsis thaliana*

4.6.1 Oczyszczanie białka AtTPR z bakterii *E.coli*

Dimeryzacja białka SGT1 została wykazana eksperymentalnie dla rekombinowanych białek z jęczmienia, *Arabidopsis thaliana* oraz drożdży piekarskich. W odróżnieniu od tych białek ludzkie białko SGT1 nie tworzy dimerów w roztworze. Nie wiadomo czy mechanizm dimeryzacji domeny TPR z różnych organizmów jest jednakowy i dlaczego ludzkie białko SGT1 nie ulega dimeryzacji. W celu zrozumienia procesu dimeryzacji białka SGT1, w ramach niniejszej pracy zbadano dodatkowo zachowanie się izolowanych domen TPR z trzech organizmów: *Arabidopsis thaliana*, drożdży piekarskich oraz człowieka w różnych stężeniach siły jonowej: 0 mM - 800mM KCl, za pomocą małokątowego rozpraszania promieniowania synchrotronowego.

Do badań dimeryzacji został wybrany fragment domeny TPR zawierający aminokwasy 1-120 z sekwencji białka SGT1a (AtTPR) (w roślinie *Arabidopsis thaliana* występują dwie izoformy białka SGT1: SGT1a i SGT1b). Sekwencja DNA kodująca fragment białka AtSGT1a została wklonowana do wektora ekspresyjnego pETM11. Do nadekspresji domeny AtTPR użyto szczep bakterii *E.coli* BL21 Codon Plus RIL. Białko AtTPR zostało oczyszczone za pomocą protokołów opisanych w sekcji 3.7.1-3.7.2, przy użyciu chromatografii powinowactwa, cięcia proteazą TEV, ponownym oczyszczaniu za pomocą chromatografii powinowactwa oraz sączenia molekularnego.

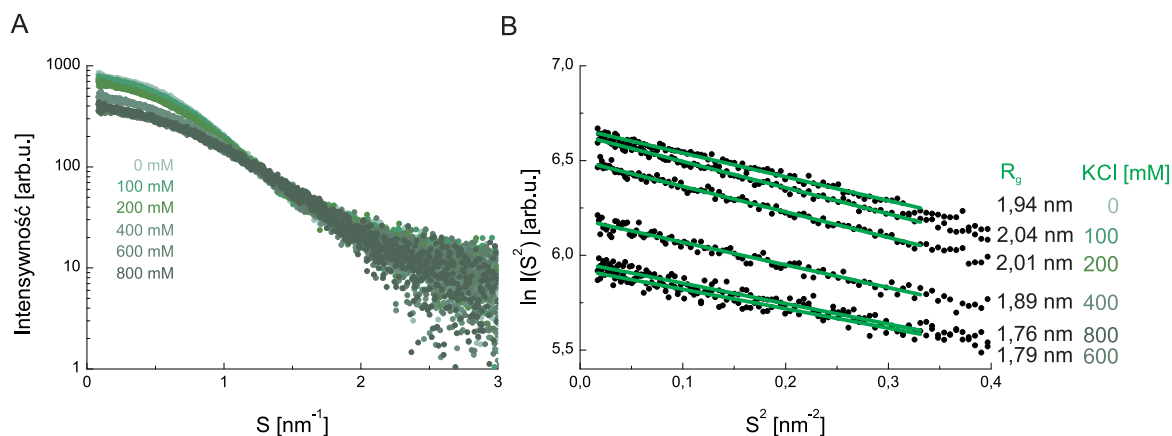


Rysunek 4.25 Oczyszczanie domeny TPR białka AtSGT1a. **A.** Oczyszczanie domeny TPR na kolumnie Superdex 200pg **B.** Czystość końcowej frakcji domeny TPR po chromatografii powinowactwa.

Po sączeniu molekularnym frakcje domeny TPR z nieodciętym fragmentem zawierającym dodatkowo histydyny oddzielono od domeny TPR za pomocą chromatografii powinowactwa. Końcowe oczyszczanie za pomocą chromatografii powinowactwa pozwoliło na otrzymanie bardzo czystego preparatu nadającego się do badań strukturalnych (rysunek 4.25).

4.6.2 Badania małokątowego rozpraszania promieniowania synchrotronowego białka AtTPR

Dla domeny AtTPR przeprowadzono pomiary SAXS przy stężeniu 4 mg/ml, gdyż wyższe stężenie powodowało pojawienie się oddziaływań repulsyjnych przy stężeniu 0 mM KCl. Zarejestrowane krzywe rozpraszania przedstawione są na rysunku 4.26.

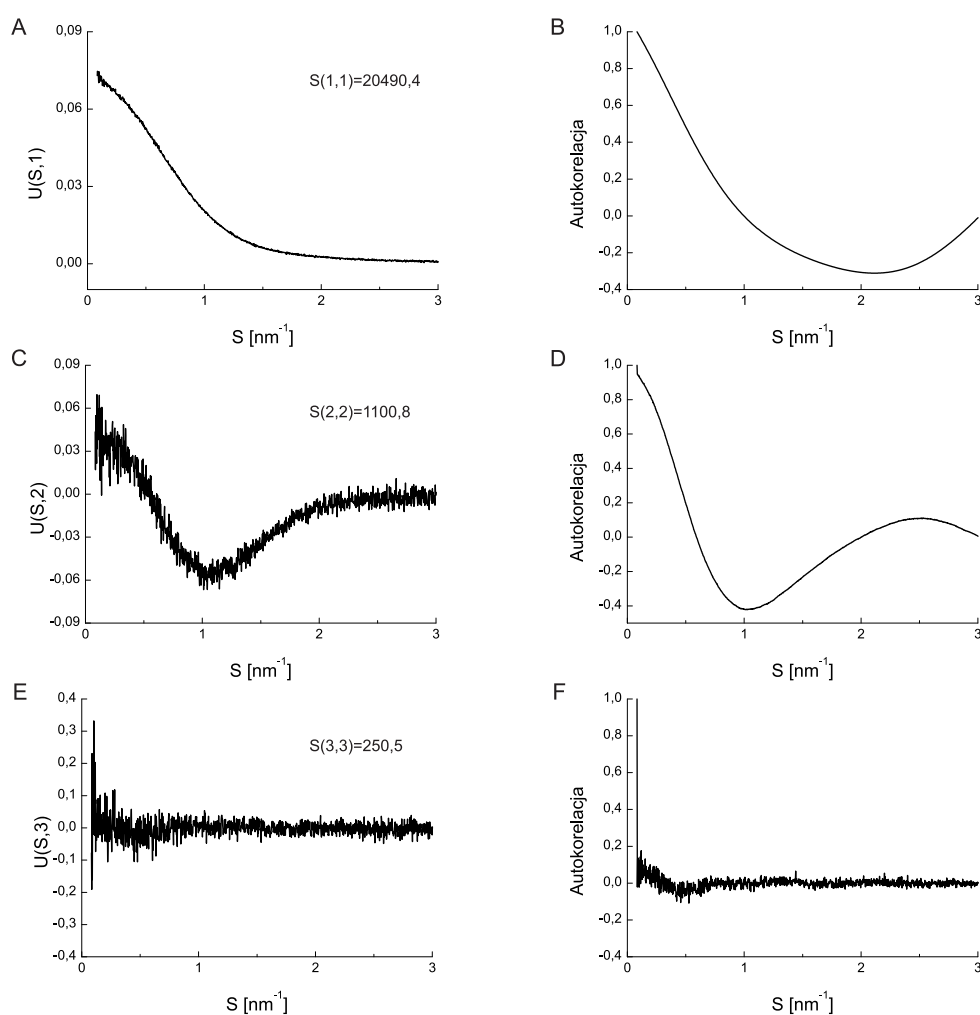


Rysunek 4.26 Eksperyment małokątowego rozpraszania promieniowania synchrotronowego domeny AtTPR. **A.** Krzywe rozpraszania dla domeny AtTPR zebrane dla stężenia KCl od 0 mM do 800 mM. **B.** Wykresy Guiniera oraz wyznaczone promienie żyracji R_g dla domeny AtTPR w różnych stężeniach soli.

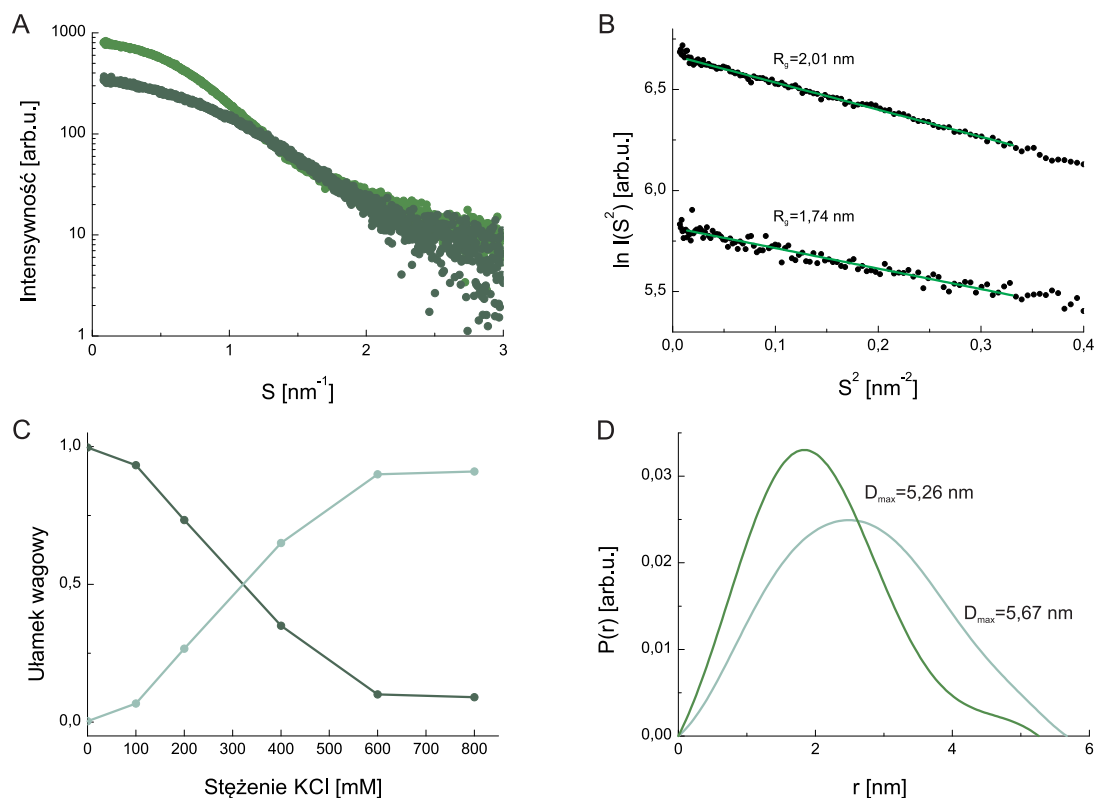
Jak przedstawiono na rysunku 4.26 A, wraz ze wzrostem siły jonowej intensywność rozpraszania ekstrapolowana do $S = 0$ maleje, podobnie jak dla domeny TPR z jęczmiennego białka SGT1. Promień promień żyracji R_g również maleje wraz ze wzrostem siły jonowej. Dla 0 mM KCl wartość promienia żyracji jest równa 1,94 nm i jest mniejsza niż dla domeny TPR z jęczmienia oraz dla 100 mM KCl. Najprawdopodobniej jest to związane z występowaniem oddziaływań repulsyjnych pomiędzy cząsteczkami białka pomimo tego, że kształt krzywej rozpraszania nie wskazywał na ich występowanie. Dla 100 mM KCl promień żyracji wyniósł 2,04 nm czyli bardzo podobnie jak dla domeny TPR z jęczmienia ($R_g = 2,05$ nm) natomiast dla stężenia KCl 200 mM, 400 mM i 600 mM promień żyracji był identyczny. Jedynie dla 800 mM promień żyracji wyniósł 1,76 nm w porównaniu do 1,69 nm dla domeny jęczmiennej. Podobnie jak dla domeny TPR z jęczmiennego białka SGT1 spadek wartości $I(0)$ oraz R_g wskazuje na występowanie równowagi pomiędzy kilkoma formami oligomerycznymi a sama równowaga jest zależna od siły jonowej.

W celu weryfikacji czy domena AtTPR również występuje w równowadze monomer-dimer przeprowadzono analizę SVD wraz z analizą MCR-ALS w celu otrzymania krzywych SAXS dla różnych form oligomerycznych białka AtTPR. W wyniku analizy SVD stwierdzono, że białko AtTPR występuje w roztworze w równowadze monomer-dimer. Wyniki samej analizy przedstawione są na rysunku 4.27. Dwa pierwsze wektory własne, U_1 oraz U_2 , mają odpowiednio wartości własne równe 20490,4 oraz

1100,8. Same wektory własne oraz ich funkcje autokorelacji są różne od zera i wyraźnie różnią się od wektora własnego U3 oraz jego funkcji autokorelacji, które mają postać szumu. Funkcja autokorelacji dla wektora U3 ma słabe minimum dla $S = 0,5 \text{ nm}^{-1}$, niemniej jednak test statystyczny wykonany w programie PRIMUS wykazał, że jedynie dwa pierwsze wektory własne są istotne. Również wartości własne dla wektorów U4-U6 mają wartości od 203,1 do 233,5 a więc bardzo zbliżoną do wartości własnej dla wektora U3. Na podstawie analizy SVD przeprowadzono analizę MCR-ALS podobnie jak dla domeny TPR z jęczmienia. Z analizy MCR-ALS otrzymano krzywe rozpraszania dla monomeru i dimeru domeny AtTPR oraz profil zamian równowagi monomeru i dimeru w zależności od stężenia KCl. Na podstawie krzywych rozpraszania wyznaczono (z wykresu Guiniera) promień żyracji R_g oraz funkcje $P(r)$ za pomocą odwrotnej transformaty Fouriera w programie GNOM. Wyniki te przedstawione są na rysunku 4.28.



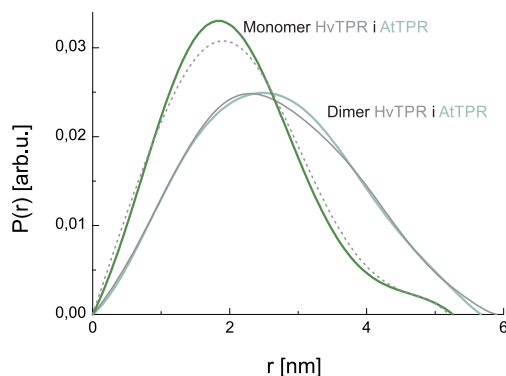
Rysunek 4.27 Analiza SVD krzywych rozpraszania dla domeny TPR białka SGT1 z *Arabidopsis thaliana*. **A, C, E** - Wektory własne U1, U2 i U3 oraz odpowiadające im funkcje autokorelacji **B, D, F** - Na krzywych przedstawiających wektory własne zaznaczone są również wartości własne dla każdego wektora.



Rysunek 4.28 Analiza MCR-ALS krzywych rozpraszania domeny TPR białka SGT1a z *Arabidopsis thaliana*. **A.** Krzywe rozpraszania dla monomeru (kolor ciemnozielony) i dimeru (kolor jasnozielony) domeny TPR. **B.** Wykres Guiniera i wyznaczony promień żyrcacji R_g dla dimeru (górną krzywą) i monomeru (dolną krzywą) domeny TPR. **C.** Zmiana równowagi monomer-dimer domeny AtTPR w roztworze o różnym stężeniu KCl (kolor ciemnozielony dimer, kolor jasnozielony monomer). **D.** Krzywe $P(r)$ dla monomeru (kolor ciemnozielony) i dimeru (kolor jasnozielony) domeny TPR.

Zależność równowagi monomer-dimer białka AtTPR ma bardzo podobny przebieg jak w przypadku domeny TPR białka HvSGT1. Punkt kiedy frakcje objętościowe monomeru i dimeru są sobie równe znajdują się przy stężeniu KCl około 300 mM czyli prawie tyle samo jak dla domeny TPR białka HvTPR aczkolwiek przesunięty jest minimalnie w stronę niższych stężeń. Kształt krzywych $P(r)$ dla dimeru i monomeru jest bardzo podobny do krzywych dla białka HvTPR. Jednak występują pewne różnice. Na rysunku 4.29 przedstawione jest porównanie krzywych $P(r)$ dla domeny TPR z białka SGT1 z *Arabidopsis thaliana* i jęczmienia. Z porównania krzywych dla monomerów domeny TPR wynika, że domena TPR białka AtSGT1a jest nieco bardziej kompaktowa w porównaniu do białka HvSGT1 o czym świadczy większa wartość krzywej $P(r)$ w maksimum i nieco mniejsze wartości R powyżej 2,3 nm. Natomiast z porównania krzywych $P(r)$ dla dimeru wynika, że dimer białka HvSGT1 jest bardziej kompaktowy w porównaniu z białkiem AtSGT1a pomimo tego, że wartość parametru D_{max} dla białka HvSGT1 jest nieco większa. Także pomimo dużego podobieństwa w

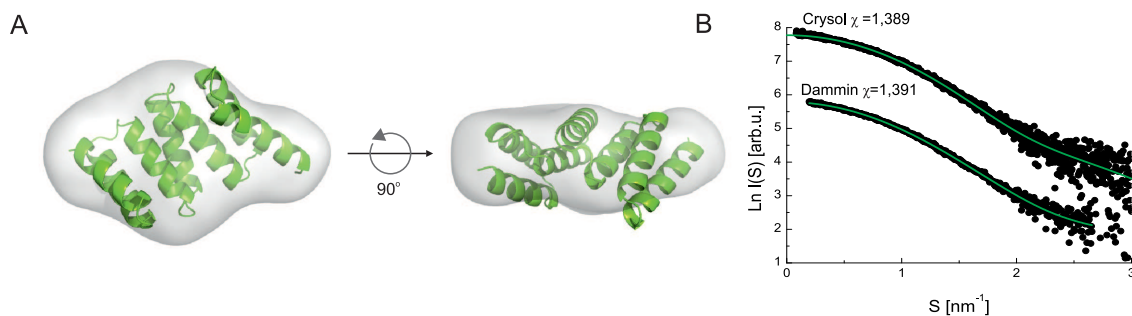
sekwencji pomiędzy domenami TPR białka SGT1 z różnych organizmów bardzo prawdopodobne są niewielkie różnice w strukturze przestrzennej.



Rysunek 4.29 Porównanie krzywe $P(r)$ dla monomeru (kolor jasnozielony) i dimeru (kolor ciemnozielony) białka AtTPR z krzywymi $P(r)$ dla monomeru (przerywana szara linia) i dimeru (ciągła szara linia) białka HvTPR.

4.6.3 Modelowanie niskorozdzielczej struktury monomeru i dimeru domeny TPR białka AtSGT1a

Na podstawie krzywych rozpraszania otrzymanych z analizy MCR-ALS wymodelowano niskorozdzielczą strukturę monomeru i dimeru białka AtSGT1a. Modele przedstawione są na rysunku 4.30 i rysunku 4.31.



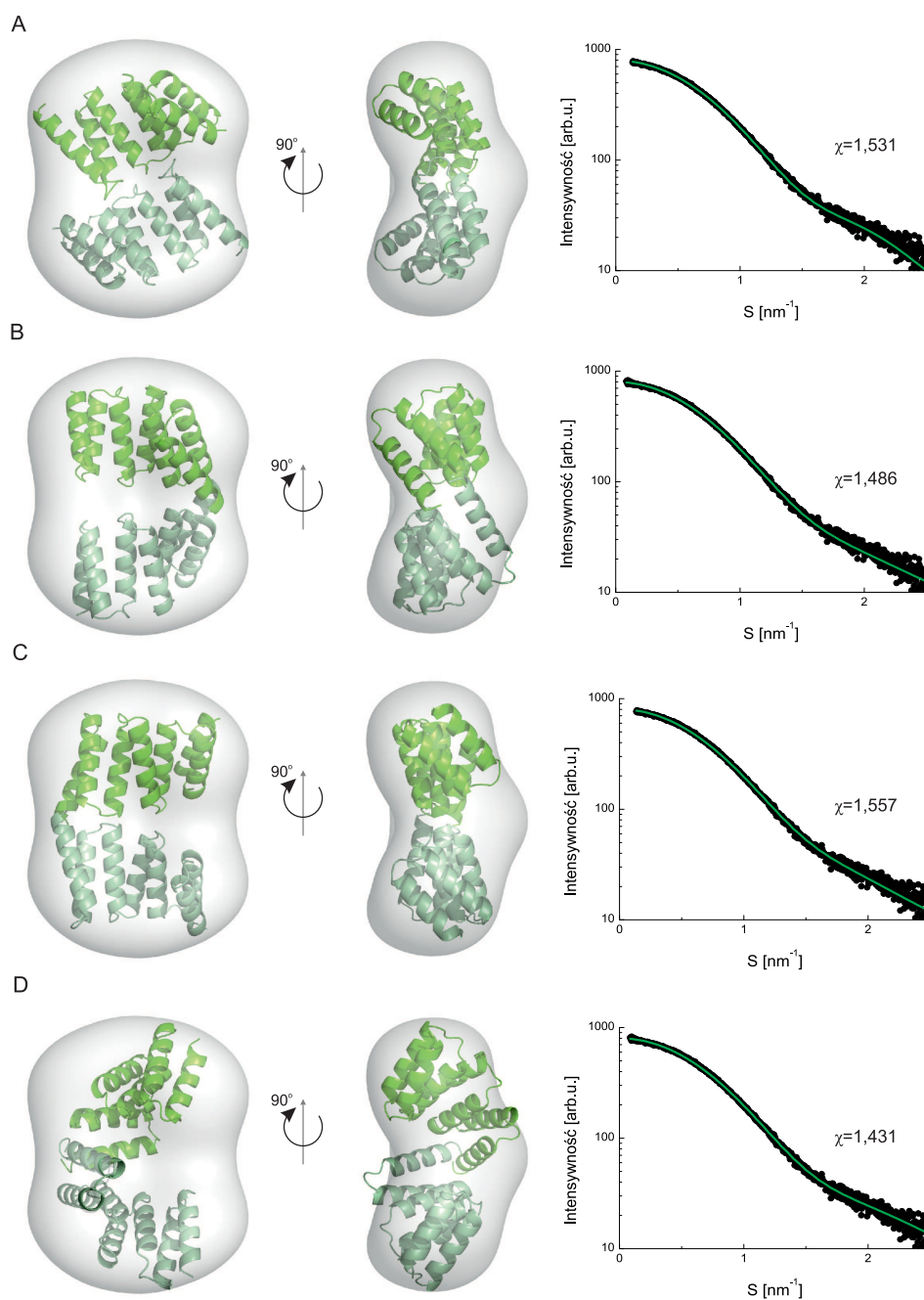
Rysunek 4.30 Modelowanie monomeru białka AtTPR w programie DAMMIN. **A.** Model monomeru białka AtTPR w postaci powierzchni (kolor szary) wraz z nałożonym modelem domeny TPR otrzymanym w programie I-Tasser. **B.** Dopasowanie do eksperymentalnej krzywej SAXS, krzywej SAXS dla modelu z programu DAMMIN i I-Tasser.

Na rysunku 4.30 A, w postaci szarej powierzchni przedstawiony jest uśredniony, niskorozdzielczy model monomeru domeny AtTPR otrzymany w programie DAMMIN. Dopasowanie krzywej SAXS

obliczonej dla najlepszego modelu z programu DAMMIN do eksperymentalnej krzywej SAXS przedstawione jest na rysunku 4.30 B. Jak widać z wykresu dopasowanie jest bardzo dobre o czym świadczy również niska wartość $\chi = 1,391$. Modelowanie struktury domeny TPR białka AtSGT1a przeprowadzono również w programie I-Tasser. Jak widać z rysunku 4.30 A model z programu I-Tasser (kolor zielony) bardzo dobrze pasuje do uśrednionego modelu z programu DAMMIN. Wartości dopasowania χ dla modelu I-Tasser wynosi 1,389 i jest bardzo podobna do wartości χ dla najlepszego modelu z programu DAMMIN. Kształt modelu jest typowy dla domeny składającej się z powtórzeń TPR.

Modelowanie niskorozdzielczej struktury dimeru domeny AtTPR również przeprowadzono w programie DAMMIN. Na rysunku 4.31 przedstawiony jest uśredniony, niskorozdzielczy model dimeru domeny AtTPR w postaci szarej powierzchni. Model dimeru domeny AtTPR ma kształt spłaszczonego dysku z przewężeniem w środkowej części modelu. Również sam model jest lekko zgięty w środkowej części. Ogólny kształt niskorozdzielczego modelu dimeru domeny AtTPR jest podobny do modelu dla dimeru domeny HvTPR. Jednak model dimeru białka HvTPR jest bardziej spłaszczony i nie posiada przewężenia w środkowej części, tylko posiada je w jednym z boków modelu.

Modelowanie dimeru domeny AtTPR przeprowadzono również w programie SASREF używając modelu otrzymanego w programie I-Tasser. W programie SASREF otrzymano kilka modeli, z których cztery najbardziej reprezentacyjne nałożone są na niskorozdzielczą strukturę dimeru domeny AtTPR otrzymanej w programie DAMMIN (rysunek 4.31). W przeciwieństwie do dimeru domeny TPR białka HvSGT1, dla domeny TPR z białka AtSGT1a nie udało się otrzymać jednego modelu z programu SASREF. Dla czterech otrzymanych modeli wartość dopasowania do eksperymentalnej krzywej SAXS zmienia się w zakresie od $\chi = 1,431$ dla modelu na rysunku 4.31 D do $\chi = 1,557$ dla modelu na rysunku 4.31 C. Model o najniższej wartości χ jest podobny do modelu otrzymanego dla dimeru domeny TPR dla białka HvSGT1. Niemniej jednak jego kształt jest najmniej podobny do niskorozdzielczej struktury otrzymanej w programie DAMMIN. W tym modelu główne oddziaływania zachodzą w obrębie C-końca białka. Modele na rysunkach 4.31 B i 4.31 C są bardzo podobne do siebie. Dla tych modeli oddziaływanie pomiędzy monomerami domen TPR również zachodzi głównie w obrębie C-końcowych helis jednak wzajemna orientacja podjednostek jest inna niż dla modelu na najniższej wartości χ . Dla modelu o najwyższej wartości dopasowania χ , oddziaływania pomiędzy domenami zachodzi poprzez oddziaływanie pierwszej i drugiej pętli które łączą ze sobą helisy z w obrębie pierwszego i drugiego powtórzenia TPR. Model ten najlepiej pasuje do niskorozdzielczego modelu dla dimeru domeny AtTPR z programu DAMMIN.

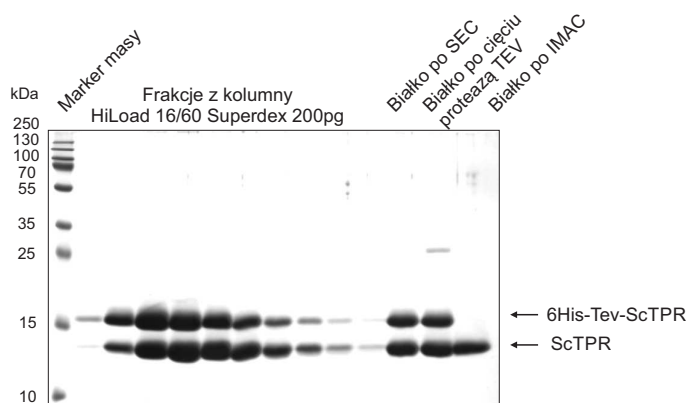


Rysunek 4.31 Modelowanie struktury dimeru domeny TPR białka AtTPR w programach DAMMIN i SASREF. Na kolejnych rysunkach A-D, na niskorozdzielczy model z programu DAMMIN (w postaci szarej powierzchni) nałożone są modele otrzymane w programie SASREF.

4.7 Badania struktury domeny TPR białka SGT1 z drożdży piekarskich w roztworze

4.7.1 Oczyszczanie białka ScTPR z bakterii *E.coli*

Białko SGT1 z drożdży piekarskich *Saccharomyces cerevisiae* (ScSGT1), podobnie jak roślinne białko SGT1 ulega dimeryzacji. W celu zbadania czy proces dimeryzacji domeny TPR z białka ScSGT1 (ScTPR) zachodzi w podobny sposób jak dla domen TPR białka SGT1 z jęczmienia i *Arabidopsis thaliana*, przeprowadzono badania małokątowego rozpraszania promieniowania rentgenowskiego w różnych warunkach siły jonowej. Białko ScTPR uzyskano według podobnego schematu jak białka HvTPR i AtTPR. Najpierw białko oczyszczano za pomocą chromatografii powinowactwa na kolumnie z jonami Ni^{2+} , odcięto etykietę 6xHis za pomocą proteazy TEV, ponownie za pomocą chromatografii powinowactwa a następnie chromatografii żelowej. Po chromatografii żelowej na kolumnie HiLoad 16/60 Superdex 200pg białko ScTPR niestety stanowiło mieszaninę białka z odciętą etykietą 6xHis i białka z nieodciętą etykietą (rysunek 4.32). Dodatkowe trawienie proteazą TEV nie przyniosło rezultatu. Dopiero dodatkowe oczyszczanie za pomocą chromatografii powinowactwa pozwoliło na otrzymanie czystego białka ScTPR odpowiedniego do badań SAXS. Na rysunku 4.32 przedstawiony jest rozdział elektroforetyczny frakcji białka ScTPR po oczyszczeniu na kolumnie HiLoad 16/60 Superdex 200pg oraz po ponownym cięciu proteazą TEV i oczyszczeniu za pomocą chromatografii powinowactwa.



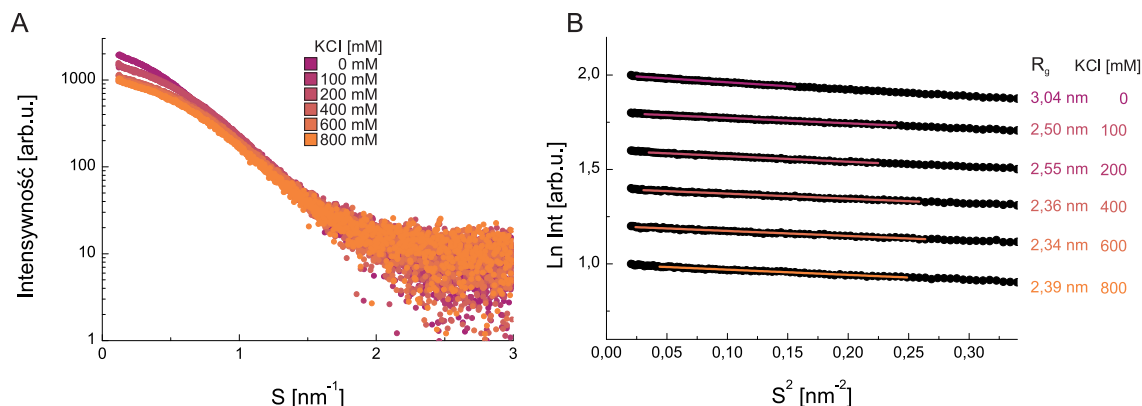
Rysunek 4.32 Rozdział elektroforetyczny białka ScTPR po sączeniu molekularnym, dodatkowym trawieniu proteazą TEV i chromatografii powinowactwa.

4.7.2 Badania małokątowego rozpraszania promieniowania rentgenowskiego białka ScTPR

Pomiary małokątowego rozpraszania przeprowadzono podobnie jak dla białek HvTPR i AtTPR przy stężeniu białka 2 mg/ml. Krzywe rozpraszania zarejestrowano dla różnych stężeń KCl: 0 mM, 100 mM, 200 mM, 400 mM, 600 mM i 800 mM. Na podstawie krzywych rozpraszania (z wykresu Guiniera) wyznaczono wartości promienia żyracji R_g . Krzywe rozpraszania i wykresy Guiniera przedstawione są na rysunku 4.33. Jak można odczytać z wykresu na rysunku 4.33 A wraz ze wzrostem siły jonowej maleje intensywność krzywych rozpraszania co świadczy o zmianach stopnia oligomeryzacji

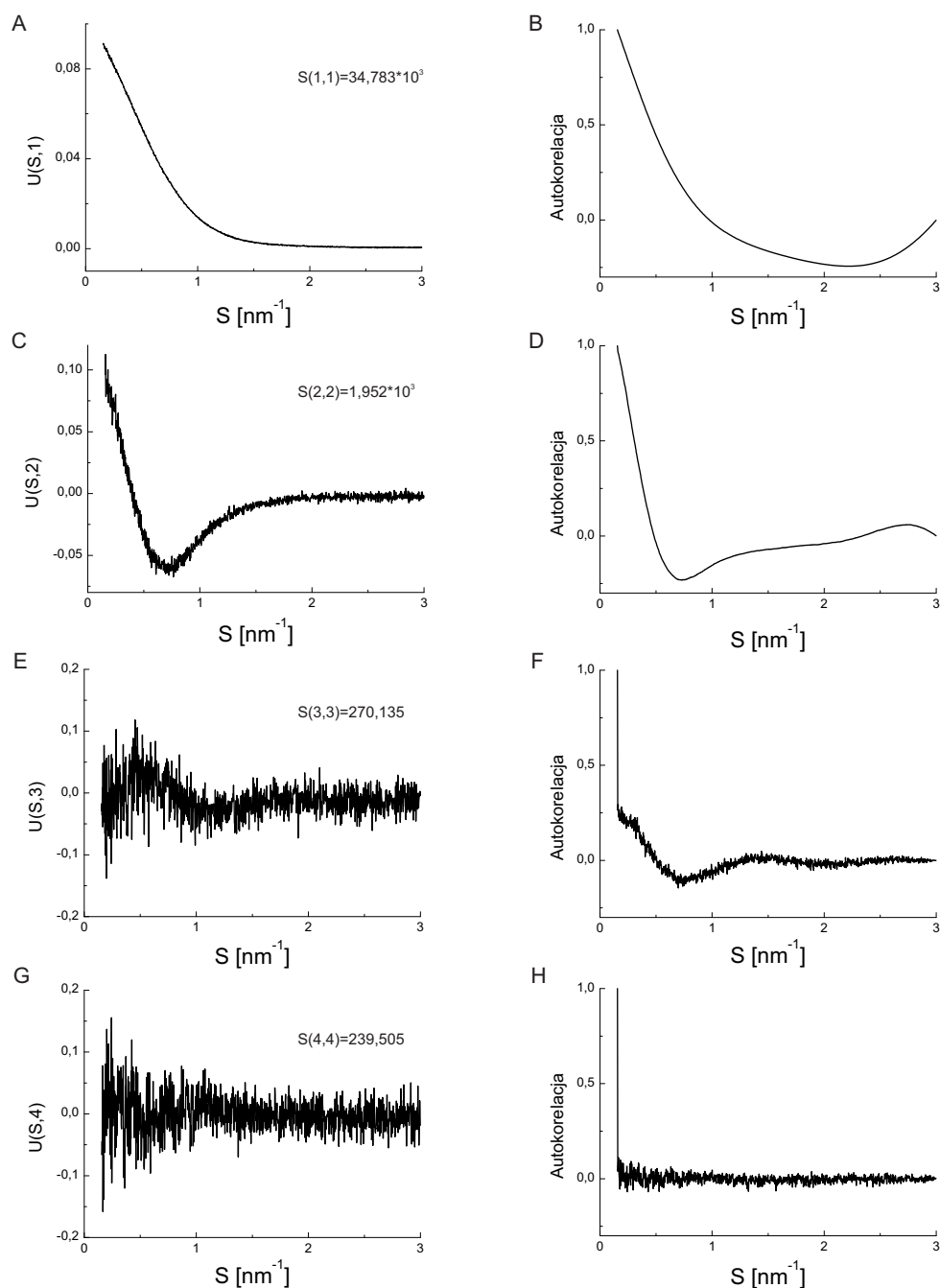
białka ScTPR. Również wartość promienia żyrcji R_g maleje wraz ze wzrostem stężenia KCl. Przy 0 mM wartość R_g wynosi 3,04 nm. Natomiast wraz ze wzrostem stężenia KCl wartość R_g zmienia się nieznacznie pomiędzy 2,55 nm przy 100 mM do 2,34 nm przy 600 mM.

Wstępna analiza danych SAXS wykazała, że zmiany w stężeniu KCl powodują zmiany w oligomeryzacji białka ScSGT1. W celu zbadania dokładniej tych zmian podobnie jak w poprzednich przypadkach przeprowadzono analizę SVD. Wyniki analizy SVD przedstawione są na rysunku 4.34. Z analizy SVD wynika, że dla białka ScTPR pierwsze trzy wektory własne są niezerowe i różne od szumu. Ich funkcje autokorelacji również są różne od szumu. Na rysunku 4.34 G i H, przedstawiony jest czwarty wektor własny oraz jego funkcja autokorelacji. Z wykresów wynika, że ma on charakter szumu. Trzy pierwsze wektory własne charakteryzują się wartościami własnymi: $34,783 \cdot 10^3$, $1,952 \cdot 10^3$, 270,135. Natomiast wartość własna dla czwartego wektora jest równa 239,505. Z analizy SVD wynika istnienie trzech, a nie jak w przypadku domen TPR białka SGT1 z jęczmienia i *Arabidopsis thaliana* dwóch, stanów oligomeryzacji białka ScTPR.



Rysunek 4.33 Eksperyment małąkątowego rozpraszania promieniowania rentgenowskiego białka ScTPR. **A.** Krzywe rozpraszania dla białka ScTPR zarejestrowane w różnych stężeniach KCl. **B.** Wykres Guiniera i wartość promienia żyrcji R_g wyznaczone na podstawie krzywych rozpraszania.

Następnie przeprowadzono analizę MCR-ALS w celu otrzymania krzywych rozpraszania dla poszczególnych stanów oligomeryzacji białka ScTPR. Wyniki tej analizy przedstawione są na rysunku 4.35. Na podstawie porównania wartości $I(0)$ dla poszczególnych stanów oligomeryzacji założono, że w przypadku białka ScTPR występuje równowaga pomiędzy monomerem, dimerem a tetramerem białka ScTPR. Wartości $I(0)$ wyznaczone dla krzywych rozpraszania z analizy MCR-ALS wynosiły odpowiednio: 1983, 1343, 585,4. Wartości R_g wynoszą natomiast 3,04 nm, 2,08 nm i 2,85 nm. Wartość $I(0)$ wskazują na występowanie tetrameru, dimeru i monomeru. Wartość $I(0)$ dla tetrameru jest mniejsza niż można by się spodziewać z porównania do wartości dla monomeru. Zarówno R_g , D_{max} i kształt krzywej $P(r)$ dla dimeru są bardzo podobne do wartości otrzymanych dla dimeru domeny TPR białka SGT1 z jęczmienia i *Arabidopsis thaliana*.



Rysunek 4.34 Analiza SVD krzywych rozpraszania dla białka ScTPR. **A, C, E, G** - Cztery pierwsze wektory własne oraz odpowiadające im wartości własne otrzymane z analizy SVD. **B, D, F, H** - Funkcje autokorelacji czterech pierwszych wektorów własnych.

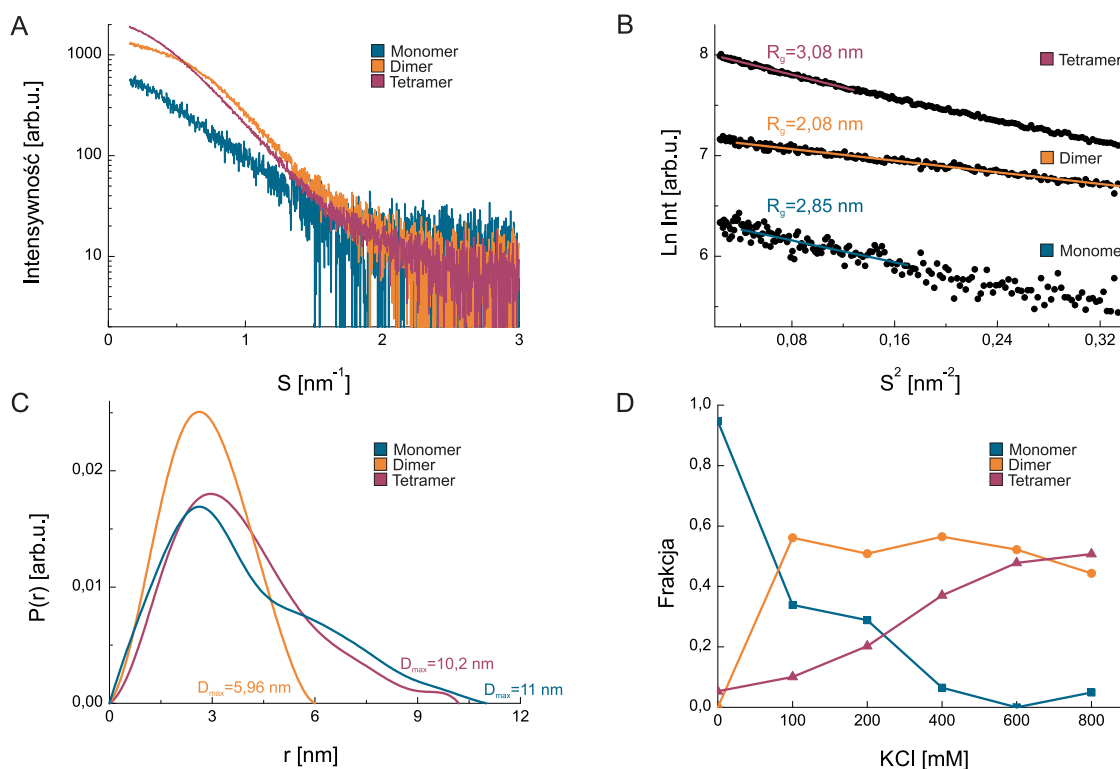
Natomiast sytuacja jest odwrotna w przypadku monomeru domeny TPR. Wartość R_g i D_{max} oraz krzywa $P(r)$ różnią się zasadniczo od krzywej $P(r)$ oraz wartości R_g i D_{max} otrzymanych dla

białek roślinnych. Kształt krzywej wskazuje, że monomer domeny ScTPR ma wydłużony kształt. Ponieważ domena ScTPR składa się z trzech powtórzeń TPR tak jak roślinne domeny TPR białka SGT1, powinna przyjmować strukturę podobną do białek roślinnych czyli bardziej kompaktową.

Równie interesująco przedstawia się zależność od siły jonowej występowania poszczególnych stanów oligomerycznych. W 0 mM KCl białko ScTPR występuje głównie jako tetramer. Wraz ze wzrostem stężenia KCl jego udział spada i w 600 mM spada do zera. Równocześnie wzrasta udział frakcji dimeru i monomeru tak jak można by było się spodziewać. Jednak frakcja dimeru pozostaje praktycznie stała w zakresie od 100 mM do 800 mM i stanowi około 50 %. Frakcja monomeru wzrasta stopniowo od 5 do około 50 % w zakresie stężenia KCl od 0 mM do 800 mM. W tabeli 4.2 zebrane są dokładne wartości zawartości poszczególnych frakcji w danym stężeniu siły jonowej.

Tabela 4.2 Zależności występowania stanów oligomerycznych białka ScTPR od stężenia KCl.

	C KCl					
	0 mM	100 mM	200 mM	400 mM	600 mM	800 mM
Monomer	0,053	0,100	0,202	0,370	0,478	0,507
Dimer	0	0,561	0,509	0,565	0,522	0,443
Tetramer	0,947	0,339	0,289	0,065	0	0,049

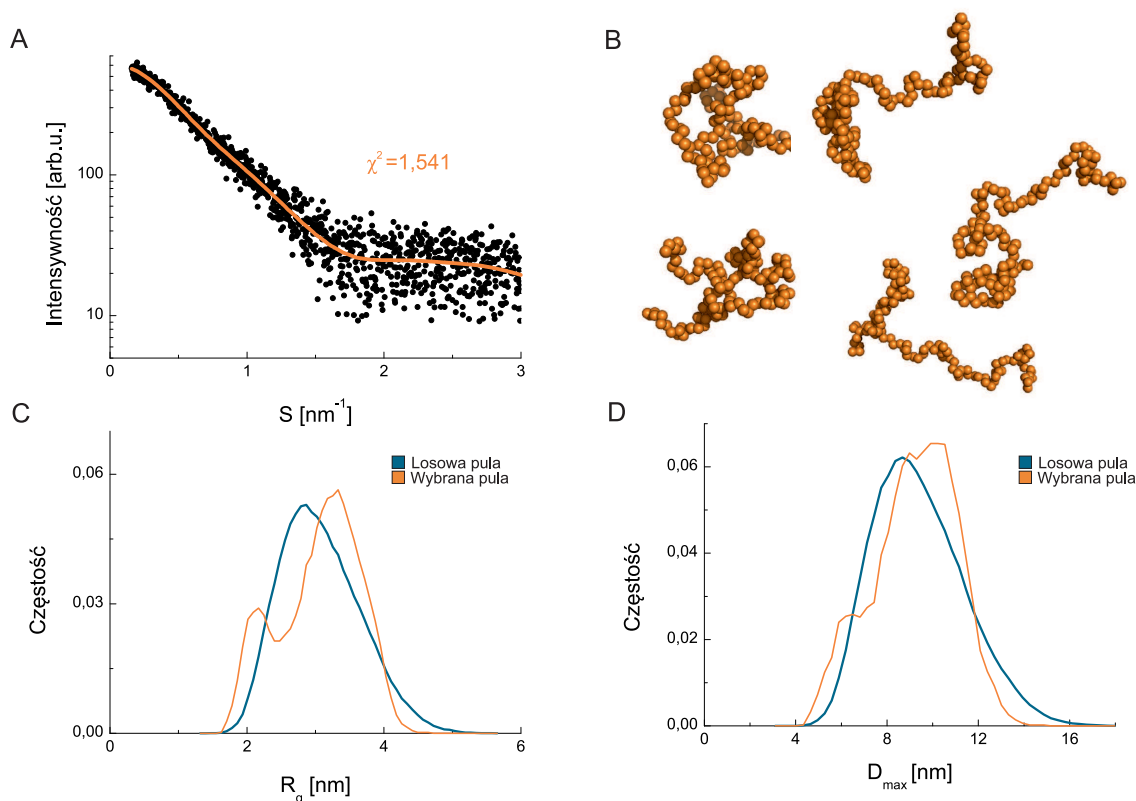


Rysunek 4.35 Analiza MCR-ALS krzywych rozpraszania białka ScTPR. **A.** Krzywe rozpraszania dla trzech stanów oligomerycznych białka ScTPR. **B.** Wykres Guiniera i wyznaczone wartości

promienia żyracji. **C.** Krzywe $P(r)$ dla trzech stanów oligomerycznych białka ScTPR. **D.** Zmiany we frakcji poszczególnych stanów oligomerycznych białka ScTPR od siły jonowej.

4.7.3 Modelowanie struktury monomeru białka ScTPR

Kształt krzywej $P(r)$ oraz wartości parametrów D_{max} - 11 nm oraz R_g - 2,85 nm dla białka ScTPR pozwalają przypuszczać, że monomer białka ScTPR wykazuje cechy białka nieuporządkowanego. Niestety mała wartość stosunku sygnału do szumu nie pozwoliła na dokładną analizę kształtu krzywej rozpraszania w oparciu o wykres Kratky'ego. Analizę struktury monomeru białka ScTPR przeprowadzono w programie EOM. Wyniki analizy przedstawione są na rysunku 4.36.



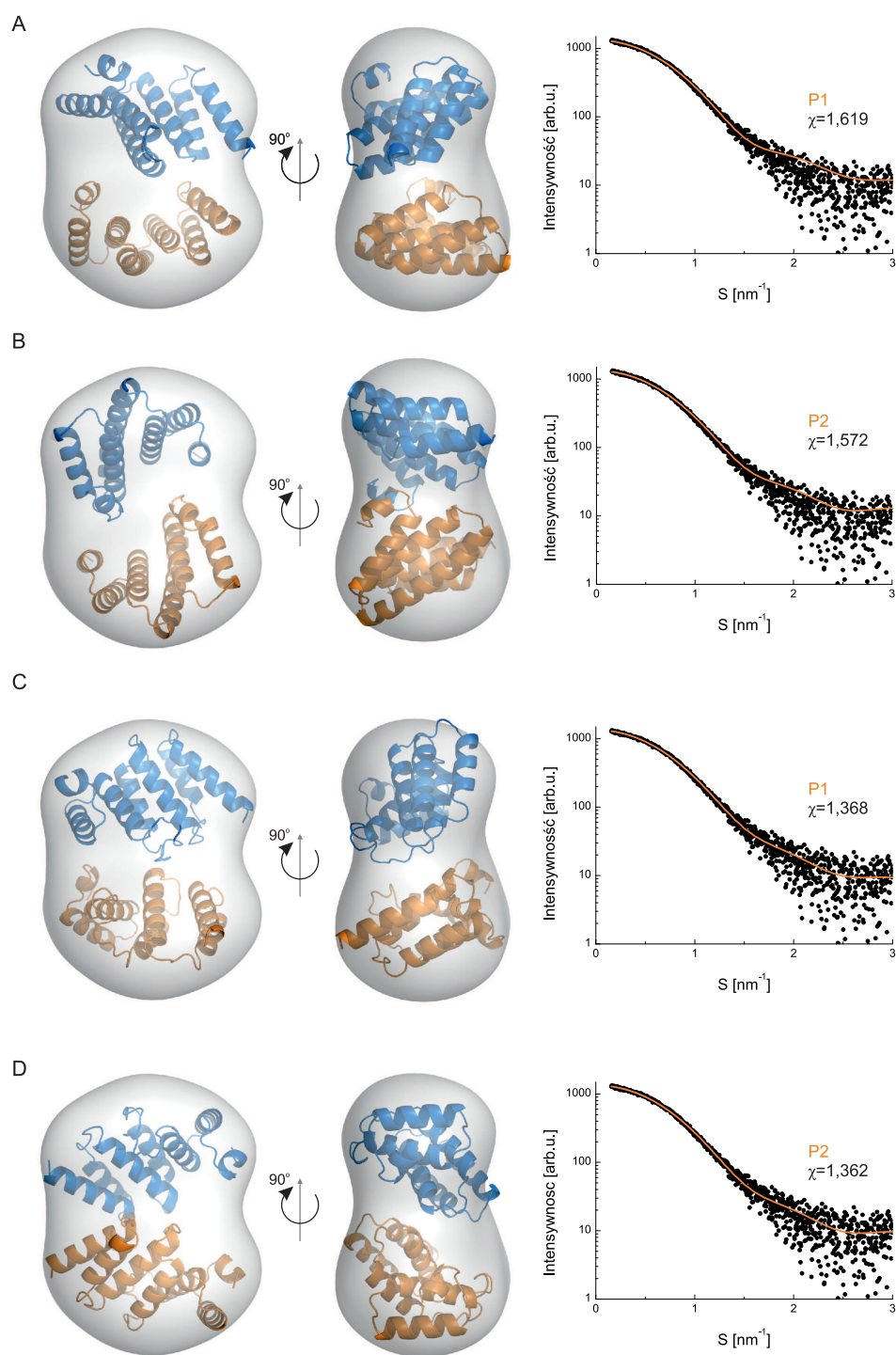
Rysunek 4.36 Analiza EOM monomeru białka ScTPR. **A.** Dopasowanie krzywej SAXS otrzymanej w programie EOM do eksperymentalnych danych SAXS. **B.** Przykładowe modele z wybranej puli z programu EOM. **C.** Rozkład wartości promienia żyracji R_g dla losowej puli modeli (kolor pomarańczowy) oraz puli modeli wybranych z losowej puli w programie EOM (kolor niebieski). **D.** Rozkład wartości maksymalnego rozmiaru cząsteczki D_{max} dla losowej puli modeli (kolor pomarańczowy) oraz puli modeli wybranych z losowej puli w programie EOM (kolor niebieski).

Wartość dopasowania krzywej SAXS otrzymanej w programie EOM do danych eksperymentalnych ($\chi^2=1,541$) jest bardzo dobra. Oznacza to, że istotnie monomer domeny ScTPR jest nieuporządk-

kowany w roztworze (przynajmniej w warunkach panujących podczas eksperymentu). Jak można zauważyć z rysunku 4.36 C i D, rozkład promienia żyracji R_g oraz maksymalnego rozmiaru cząsteczki D_{max} ma dwa maksima. Pierwsze maksimum występuje dla R_g i D_{max} odpowiednio równych około 2 nm i 5 nm. Drugie minimum występuje przy R_g i D_{max} równym odpowiednio 3,5 nm i 10 nm. Pierwsze maksimum jest też dużo węższe i jego intensywność jest mniejsza w porównaniu do drugiego, które jest szersze i ma większą intensywność. Monomer białka ScTPR występuje więc w postaci dwóch konformacji. Pierwsza konformacja, bardziej kompaktowa, charakteryzuje się R_g równym około 2 nm druga zaś charakteryzuje się szerokim zakresem wartości promienia żyracji w zakresie do około 2,5 nm do 5 nm i ma bardziej dynamiczny charakter.

4.7.4 Modelowanie struktury dimeru białka ScTPR

Modelowanie dimeru domeny ScTPR najpierw przeprowadzono w programie DAMMIN. Najlepsze wyniki otrzymano przy założeniu lokalnej symetrii P2. Model w postaci powierzchni zaprezentowany jest w postaci dwóch orientacji na rysunku 4.37. Niskorozdzielczy model z programu DAMMIN ma kształt spłaszczonego dysku, podobnie jak w przypadku modelu dimeru białka HvTPR, jednak posiada lekkie przewężenie w środkowej części. Wartość dopasowania dla najlepszego modelu wynosiła $\chi = 1,222$. Strukturę domeny TPR białka ScSGT1 wymodelowano również za pomocą programu I-Tasser. Model domeny TPR z programu I-Tasser ma strukturę typową dla domen zawierających powtórzenia TPR. Strukturę dimeru domeny ScTPR wymodelowano za pomocą metody modelowania bryły sztywnej w programie SASREF. Do modelowania użyto dwa modele z programu I-Tasser by sprawdzić czy w zależności od modelu będzie można otrzymać inne wyniki. Wyniki modelowania przedstawione są na rysunku 4.37 w postaci nałożenia na model otrzymany w programie DAMMIN (szara powierzchnia). Modelowanie przeprowadzono przy braku więzów na symetrię lokalną i z założeniem symetrii lokalnej P2. Na rysunku 4.37 obok modeli przedstawione jest dopasowanie krzywej SAXS dla modelu (krzywa pomarańczowa) do eksperymentalnych danych SAXS (kolor czarny) wraz z wartością dopasowania w postaci wartości χ . Najlepszy model otrzymano dla drugiego modelu z programu I-Tasser z założeniem symetrii lokalnej P2 (rysunek 4.37 D). Dla tego modelu wartość χ wynosiła 1,362. Niewiele mniejszą wartość $\chi = 1,368$ otrzymano dla tego samego modelu z programu I-Tasser przy braku założonej symetrii lokalnej. Oba modele jednak różnią się od siebie zasadniczo. W przypadku modelu z braku założenia o symetrii lokalnej (rysunek 4.37 C) jedna podjednostka domeny TPR oddziałuje poprzez boczną stronę (tam gdzie znajdują się pętle łączące α helisy w powtórzeniu TPR) z drugą podjednostką zwróconą wklęsłą stroną w kierunku pierwszej podjednostki. W przypadku modelu o założonej symetrii lokalnej P2 (rysunek 4.37 D) obie podjednostki domeny TPR zwrócone są do siebie stroną wypukłą, jednak oddziaływanie zachodzi głównie przez N-końcową stronę domeny TPR. Podobne wyniki otrzymano dla pierwszego modelu z programu I-Tasser. Model otrzymany z modelowania w programie I-Tasser przy braku założenia symetrii lokalnej (rysunek 4.37 A) ma bardzo podobne ułożenie podjednostek względem siebie. Natomiast model otrzymany w programie SASREF z założeniem lokalnej symetrii P2 (rysunek 4.37 B) ma podobne ułożenie jak model otrzymany dla drugiego modelu z programu I-Tasser z różnicą, że w tym modelu podjednostki w dimerze oddziałują ze sobą na całej wklęsłej stronie domeny TPR.

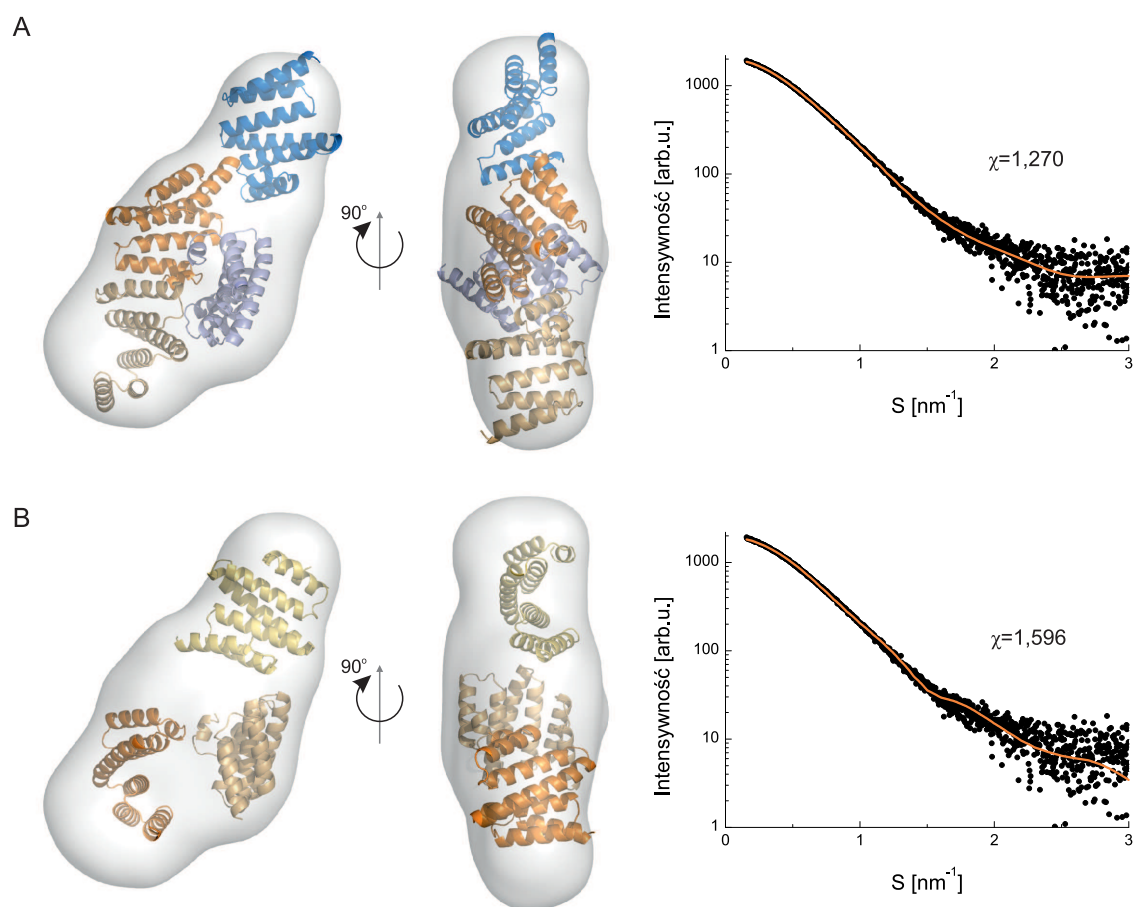


Rysunek 4.37 Modelowanie struktury dimeru domeny TPR białka ScTPR w programach DAMMIN i SASREF. Nałożenie modelu otrzymanego w programie SASREF (z założeniem braku symetrii lokalnej (A), oraz założeniem lokalnej symetrii P2 (B) wraz z dopasowaniem do danych eksperymentalnych. Na rysunkach C i D przedstawione są wyniki modelowania z takimi samymi założeniami jak

na rysunkach A i B lecz z użyciem innego modelu otrzymanego w programie I-Tasser. Kolorem szarym przedstawiony jest uśredniony model dimeru domeny ScTPR otrzymany w programie DAMMIN (w postaci powierzchni) dwa monomery domeny TPR zaprezentowane są w kolorze pomarańczowym i niebieskim.

4.7.5 Modelowanie struktury tetrameru domeny TPR białka ScSGT1

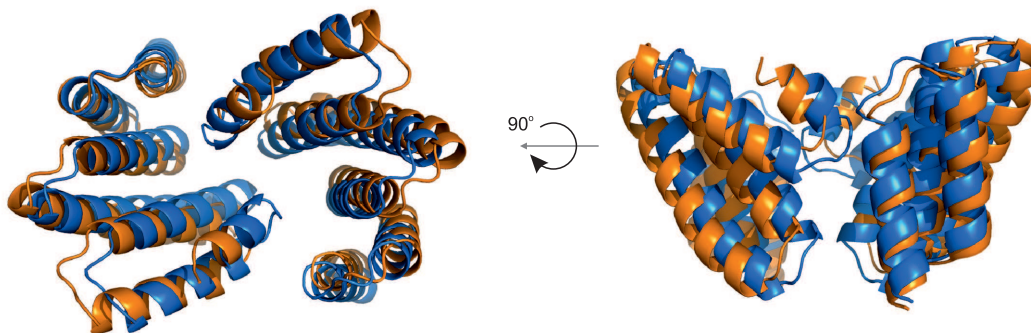
Niskorozdzielczą strukturę tetrameru białka ScTPR wymodelowano w programie DAMMIN. Dziecięc otrzymanych modeli uśredniono w programie DAMAVER. Na rysunku 4.38 uśredniony model niskorozdzielczej struktury tetrameru białka ScTPR przedstawiony jest w postaci szarej powierzchni.



Rysunek 4.38 Modelowanie struktury tetrameru domeny TPR białka ScTPR w programach DAMMIN i SASREF. **A.** Nałożenie modelu otrzymanego w programie SASREF (z użyciem modelu z programu I-Tasser) wraz z dopasowaniem do danych eksperymentalnych. **B.** Nałożenie wyników modelowania trimeru białka ScTPR w programie SASREF na model tetrameru otrzymany w programie DAMMIN. Przy wykresach podane są wartości χ dopasowania krzywej SAXS dla modelu z

programu SASREF do eksperymentalnej krzywej SAXS dla tetrameru.

Kształt tetrameru jest wydłużony z lekkim poszerzeniem w środkowej części modelu. W kolejnym etapie badań, w celu otrzymania dokładniejszego modelu dla tetrameru posłużono się modelowaniem metodą bryły sztywnej. Wyniki modelowania w programie SASREF przedstawione są na rysunku 4.38 w postaci nałożenia na niskorozdzielczy model otrzymany w programie DAMMIN. Dodatkowo przeprowadzono w programie SASREF modelowanie trimeru i pentameru białka ScTPR. Wyniki modelowania trimeru przedstawione są na rysunku 4.38 B. Model tetrameru składa się z centralnej części składającej się z dimeru domen TPR skierowanych do siebie wklęsłą powierzchnią. Model środkowej części tetrameru jest bardzo podobny do struktury dimeru białka ScTPR otrzymanego w programie SASREF (model dimeru z rysunku 4.37 B). Nałożenie obu modeli przedstawione jest na rysunku 4.39. Dwie dodatkowe domeny TPR oddziałują ze środkową częścią jedynie za pomocą drugiej i trzeciej helisy z N-terminalnej części domeny. Model dla tetrameru bardzo dobrze pasuje do modelu otrzymanego w programie DAMMIN. Środkowa część tetrameru pasuje bardzo dobrze do poszerzonej części modelu z programu DAMMIN. Wartość dopasowania dla tetrameru wynosiła $\chi = 1,270$ w porównaniu do $\chi = 1,596$ dla trimeru i $\chi = 1,448$ dla pentameru. Z tym, że w modelu pentameru piąta podjednostka nie oddziaływała z resztą modelu i była od niego znacząco oddalona. Model dla trimeru otrzymany w programie SASREF można nałożyć na model tetrameru otrzymany w programie DAMMIN. Niemniej jednak podjednostki są upakowane dość luźno a dopasowanie w środkowej części modelu z programu DAMMIN jest gorsze niż w przypadku modelu tetrameru otrzymanego w programie SASREF.

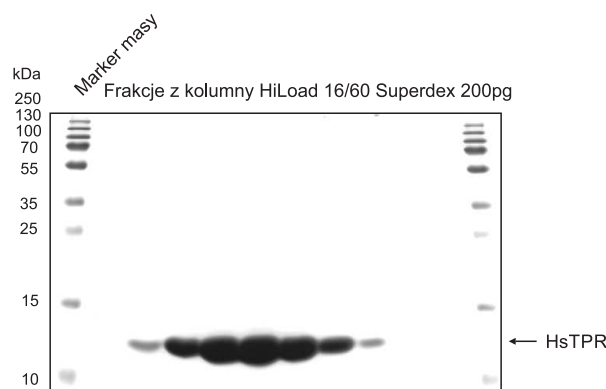


Rysunek 4.39 Porównanie struktury modelu dimeru białka ScTPR otrzymanego w programie SASREF (kolor niebieski) ze środkową częścią modelu tetrameru białka ScTPR otrzymanego w programie SASREF (kolor pomarańczowy).

4.8 Badanie struktury domeny TPR ludzkiego białka SGT1

4.8.1 Oczyszczanie białka HsTPR z bakterii *E.coli*

Ludzkie białko SGT1 w roztworze nie ulega dimeryzacji tak jak roślinne i drożdżowe białko SGT1 (Nyarko i wsp., 2007). Nie jest znany mechanizm dimeryzacji domeny TPR białek roślinnych i drożdżowych a nie też dlaczego domena TPR ludzkiego białka SGT1 nie ulega dimeryzacji. W celu zbadania jaka jest niskorozdzielcza struktura domeny TPR ludzkiego białka SGT1 oraz jak zachowuje się ona w roztworze przeprowadzono eksperymenty małokątowego rozpraszania promieniowania rentgenowskiego zmieniając stężenie KCl w zakresie od 0 mM do 800 mM. Podobnie jak domeny TPR z białka SGT1 z jęczmienia, *Arabidopsis thaliana* i drożdży piekarskich domenę HsTPR otrzymywano z bakterii *E. coli* i oczyszczano za pomocą chromatografii powinowactwa na kolumnie z jonami Ni^{2+} , odcięto etykietę 6xHis za pomocą proteazy TEV, ponownie oczyszczano za pomocą chromatografii powinowactwa a następnie chromatografii żelowej. Białko HsTPR po ostatnim etapie oczyszczania na kolumnie HiLoad 16/60 Superdex 200pg nie wymagało już dalszego oczyszczania. Czystość frakcji otrzymanych podczas sączenia molekularnego przedstawiona jest w postaci żelu po elektroforezie w żelu poliakrylamidowym na rysunku 4.40.



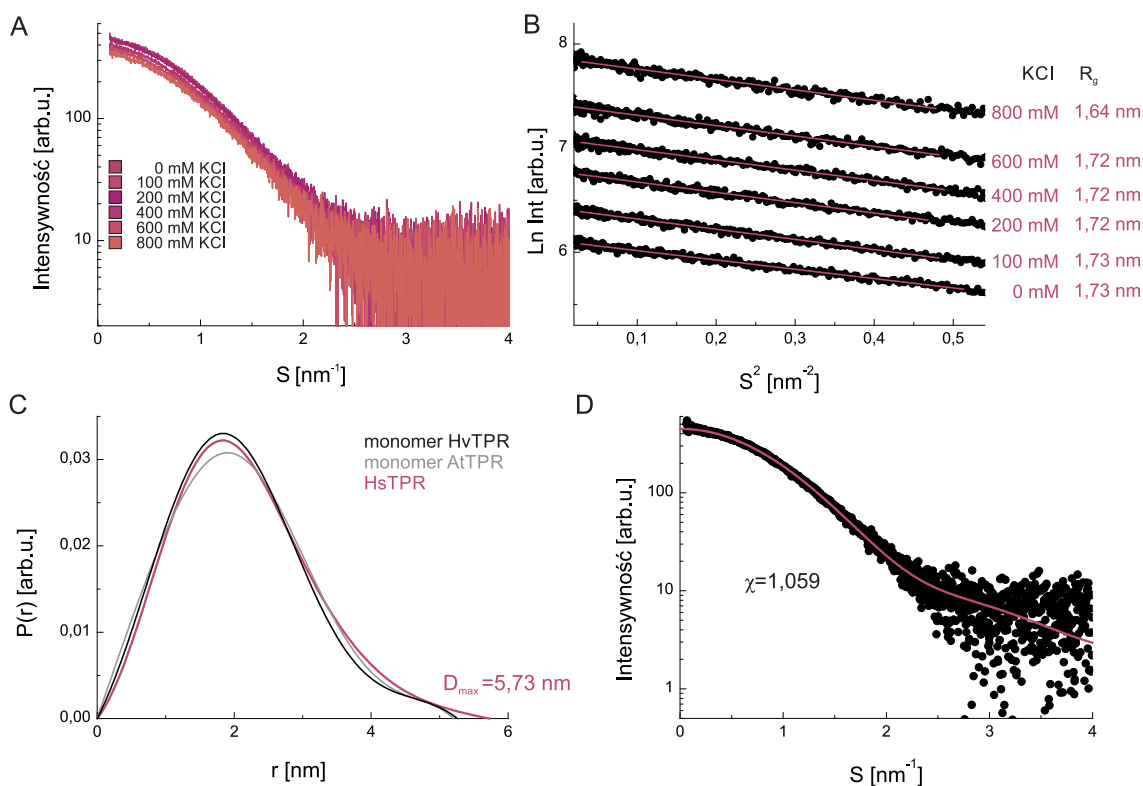
Rysunek 4.40 Rozdział elektroforetyczny w żelu poliakrylamidowym frakcji białka HsTPR po sączeniu molekularnym na złożu Superdex 200pg.

4.8.2 Badania małokątowego rozpraszania promieniowania rentgenowskiego białka HsTPR

Na rysunku 4.41 A, zaprezentowane są zmierzone krzywe SAXS dla białka HsTPR dla różnych wartości stężenia KCl. Bez względu na stężenie KCl, kształt krzywej SAXS jest praktycznie taki sam. Niewielkie różnice w ekstrapolowanej wartości intensywność dla zerowego kąta rozpraszania wynikają jedynie z błędów odjęcia krzywej SAXS dla buforu. Na rysunku 4.41 B, zaprezentowany jest wykres Guiniera wraz z wyznaczonymi wartościami promienia żyracji R_g dla krzywych SAXS zmierzonych dla różnego stężenia KCl. Wartość promienia żyracji w przedziale stężenia KCl od 0 mM do 800 mM zmienia się od 1,64 nm do 1,73 nm. W przedziale od 100 mM do 800 mM KCl promień żyracji praktycznie nie ulega zmianie. Niższa wartość R_g dla 0 mM może wynikać z różnicy w grubości i

składzie otoczki hydratacyjnej domeny HsTPR. W przeciwieństwie do domeny TPR z białek SGT1 z *Arabidopsis thaliana* i drożdży piekarskich. Zarówno kształt krzywych SAXS jak również wartość promienia żyracji R_g nie zmienia się wraz ze wzrostem siły jonowej roztworu. Domena TPR białka SGT1 także nie zmienia swojego stanu oligomerycznego. Wyznaczony maksymalny rozmiar cząsteczki dla domeny HsTPR wynosi 5,73 nm i jest porównywalny z wartościami otrzymanymi dla białek AtTPR i HvTPR. Również kształt krzywej $P(r)$ (rysunek 4.41 C) jest bardzo podobny do kształtu krzywych $P(r)$ dla białek AtTPR i HvTPR. Na rysunku widać drobne różnice pomiędzy krzywymi $P(r)$, jednak nie są one bardzo duże. Świadczy to o tym, że ogólny kształt i konformacja domeny HsTPR są podobny do domen AtTPR i HvTPR.

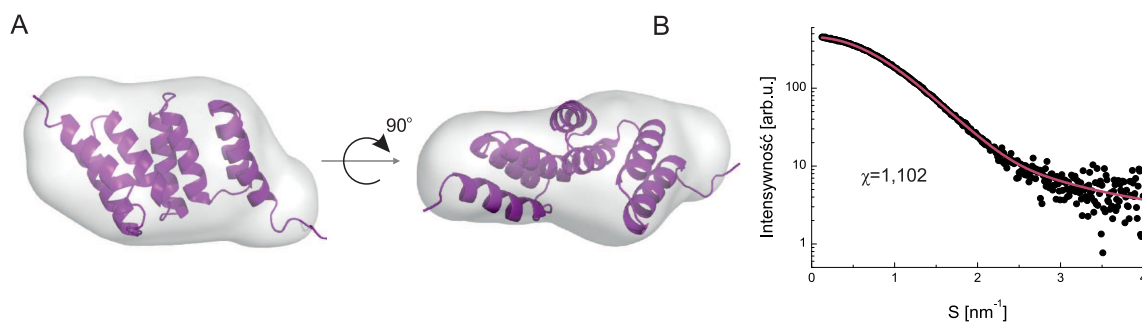
Strukturę domeny TPR ludzkiego białka SGT1 wymodelowano za pomocą serwisu do modelowania homologicznego I-Tasser. Dopasowanie krzywej SAXS obliczonej za pomocą programu CRY SOL do eksperymentalnej krzywej SAXS dla domeny HsTPR (0 mM KCl) przedstawione jest na rysunku 4.41 D. Jak widać z rysunku dopasowanie jest bardzo dobre, o czym świadczy również wartość dopasowania χ równa 1,059. Struktura modelu domeny TPR białka SGT1 przedstawiona jest na rysunku 4.42 A.



Rysunek 4.41 Eksperyment małokątowego rozpraszania promieniowania rentgenowskiego białka HsTPR. **A.** Krzywe SAXS zarejestrowane dla domeny HsTPR dla różnych wartości stężenia KCl. **B.** Wykres Guiniera dla krzywych SAXS wraz z wyznaczonym promieniem żyracji R_g . **C.** Krzywa $P(r)$ dla domeny HsTPR przy 0 mM stężeniu KCl (kolor fioletowy). Kolorem szarym i czarnym

zaznaczone są krzywe $P(r)$ odpowiednio dla monomeru domeny HvTPR i AtTPR. **D.** Dopasowanie krzywej SAXS dla modelu z programu I-Tasser do eksperymentalnych danych SAXS dla domeny HsTPR przy 0 mM stężeniu KCl.

Modelowanie niskorozdzielczej struktury białka HsTPR przeprowadzono w programie DAMMIN. Dziesięć modeli z programu DAMMIN uśredniono w programie DAMAVER. Wynik modelowania przedstawiony jest na rysunku 4.42 A w postaci szarej powierzchni. Jak widać na rysunku 4.42 A, model z programu I-Tasser bardzo dobrze pasuje do uśrednionego modelu otrzymanego w programach DAMMIN i DAMAVER. Kształt niskorozdzielczej struktury dla białka HsTPR jest bardzo podobny do niskorozdzielczej struktury białka AtTPR i HvTPR. Model z programu DAMMIN ma lekkie wybrzuszenie na jednym z końców, w które można wpasować nieuporządkowany fragment C-terminalnej helisy α modelu domeny HsTPR z programu I-Tasser. Wartość χ , dopasowania krzywej SAXS dla najlepszego modelu z programu DAMMIN do eksperymentalnej krzywej SAXS, wynosi 1,102 co świadczy bardzo dobrym dopasowaniu.



Rysunek 4.42 Modelowanie monomeru białka HsTPR w programie DAMMIN. **A.** Model monomeru białka HsTPR w postaci powierzchni (kolor szary) wraz z nałożonym modelem domeny HsTPR otrzymanym w programie I-Tasser. **B.** Dopasowanie do eksperymentalnej krzywej SAXS krzywej obliczonej dla najlepszego modelu otrzymanego w programie DAMMIN.

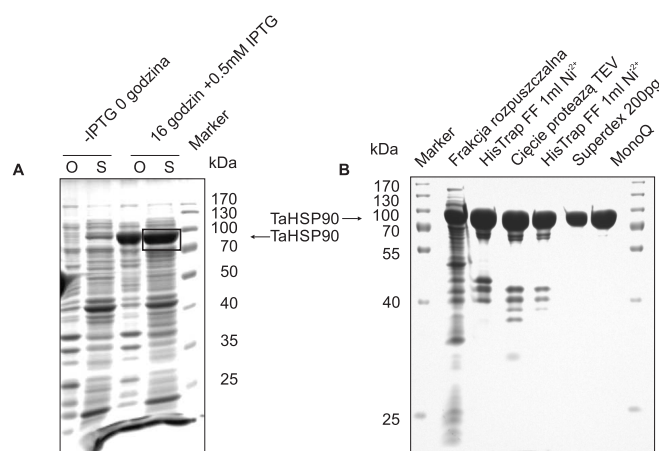
4.9 Wpływ siły jonowej i temperatury na konformację białka HSP90 w roztworze

Jak już było opisane we wstępie, białko HSP90 z różnych organizmów może występować w roztworze w równowadze kilku konformacji. U białka HtpG, które jest homologiem białka HSP90 z bakterii *E. coli*, zaobserwowano dwie konformacje w roztworze: zamkniętą (konformacja obserwowana w strukturze krystalograficznej w kompleksie z AMPPNP) i otwartą (Krukenberg i wsp., 2008). Podobnie w przypadku ludzkiego i drożdżowego białka HSP90 zaobserwowano równowagę pomiędzy konformacją zamkniętą i otwartą (Southworth i wsp., 2008). W kompleksie białka HtpG z ADP zaobserwowano jeszcze inną konformacja: tzw. kompaktowa charakteryzująca się mniej rozciągniętą konformacją białka HSP90 (Shiau i wsp., 2008). W przypadku eukariotycznych homologów dominuje konformacja otwarta. Różne czynniki mają wpływ na przesunięcie tej równowagi np. pH czy też stężenie osmotyków (Krukenberg i wsp., 2009a, Street i wsp., 2010). Jak do tej pory brak jest doniesień na temat wpływu temperatury i siły jonowej na równowagę konformacyjną białka. Z eksperymentów biochemicznych wiemy, że białko HSP90 oligomeryzuje w wyższej temperaturze jednak nie są znane podstawy strukturalne tego procesu. W celu zbadania efektu temperatury i siły jonowej na białko HSP90 przeprowadzono więc eksperymenty małokątowego rozpraszania promieniowania rentgenowskiego, których rezultaty zostaną przedstawione w poniższych podrozdziałach. Do testów wybrano białko HSP90 z pszenicy ponieważ, białko to dużo lepiej oczyszczało się z bakterii *E. coli* w porównaniu z białkiem HSP90 z jęczmienia. Białko HSP90 z pszenicy posiada 98 % podobieństwa sekwencyjnego do białka HSP90 z jęczmienia.

4.9.1 Oczyszczanie białka TaHSP90 z bakterii *E. coli*

Sekwencja cDNA kodująca białko TaHSP90 została wklonowana do wektora ekspresyjnego pETM11 w celu nadprodukcji tego białka w komórkach *E. coli*. Po kilku próbach w różnych warunkach hodowli udało się znaleźć optymalne warunki nadekspresji pozwalające na otrzymanie jak największej ilości białka we frakcji rozpuszczalnej. Dokładny protokół opisany jest w sekcji Materiały i Metody.

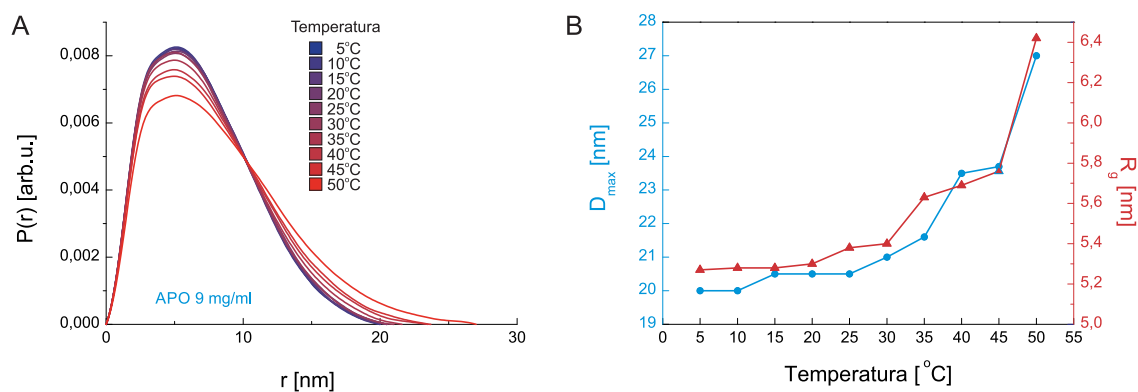
Przed dodaniem induktora nadekspresji (IPTG do stężenia 0,5 mM), hodowlę komórek poddawano szokowi termicznemu w temperaturze 47 °C w celu indukcji ekspresji natywnych bakteryjnych chaperonów, które miałyby pomóc w poprawnym fałdowaniu się białka TaHSP90. Zastosowano również dodatek osmotytu - betainy, w stężeniu 1 mM oraz podwyższenie stężenia NaCl w pożywce bakteryjnej do 0,5 M w celu indukcji szoku osmotycznego i syntezy osmotyków, które stabilizują nowo powstające białka. Samą nadekspresję prowadzono przez 16 godzin w temperaturze 18 °C. Na rysunku 4.43 przedstawione są wyniki nadekspresji białka TaHSP90 w bakteriach. Białko oczyszczano za pomocą chromatografii powinowactwa na złożu niklowym, cięcia za pomocą proteazy TEV w celu odcięcia etykiety 6xHis, ponownej chromatografii powinowactwa w celu związania proteazy TEV i białka z nieodciętą etykietą 6xHis, filtracji żelowej i ostatecznie chromatografii jonowymiennej. Stopień nadekspresji oraz czystość białka w kolejnych krokach oczyszczania przedstawiono jako obraz elektroforezy w żelu poliakrylamidowym przedstawionym na rysunku 4.43. Ostatecznie otrzymano czysty i homogeny preparat białka TaHSP90 o stosunku absorbancji przy 260 nm i 280 nm: $A_{260nm}/A_{280nm} \approx 0.5$. Świadczy to o tym, że białko TaHSP90 nie jest związane z nukleotydami, w przeciwnym wypadku wartość stosunku A_{260nm}/A_{280nm} byłaby wyższa.



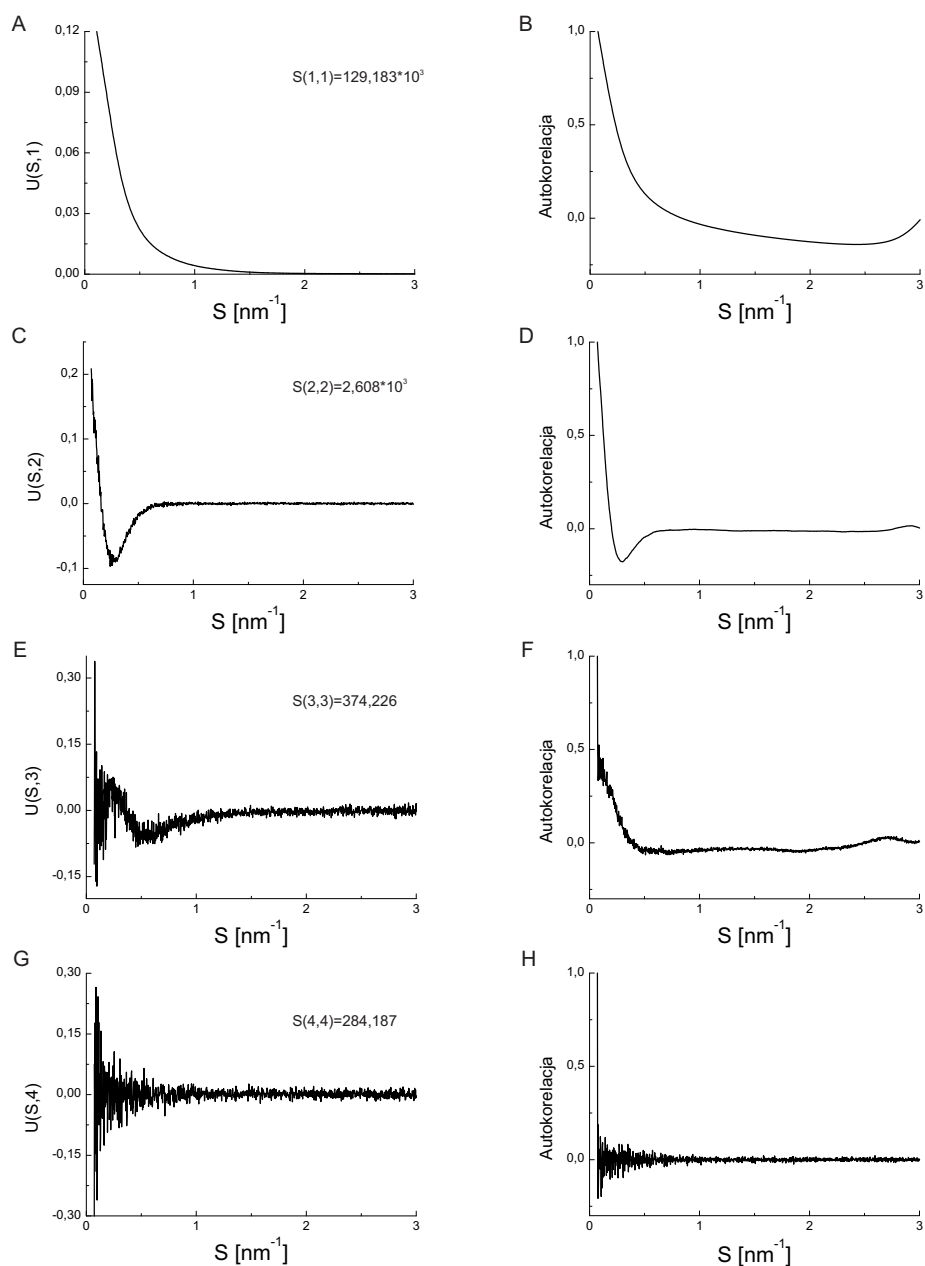
Rysunek 4.43 Nadekspresja i oczyszczanie białka TaHSP90. **A.** Żel poliakrylamidowy ukazujący nadekspresję białka TaHSP90 w bakteriach *E. coli*. (O-osad, frakcja nierozpuszczalna, S -supernatant, frakcja rozpuszczalna) **B.** Oczyszczanie białka TaHSP90 przedstawione w postaci zdjęcia elektrofo-rezy w żelu poliakrylamidowym.

4.9.2 Badanie wpływu temperatury na dynamikę konformacyjną białka HSP90 z pszenicy w roztworze

W pierwszej kolejności postanowiono sprawdzić jaki jest wpływ temperatury na konformację białka HSP90 z pszenicy (TaHSP90). W tym celu przeprowadzono eksperyment SAXS dla białka TaHSP90 o stężeniu 9 mg/ml w buforze 50 mM Tris pH 7,5 150 mM NaCl 1 mM DTT. Roztwór białka mierzono w zakresie temperatur od 5 °C do 50 °C zmieniając temperaturę co 5 °C. Za każdym razem białko inkubowano przez 1 minutę przed dokonaniem pomiaru. Dla każdej krzywej rozpraszania obliczono krzywą $P(r)$ a także wartości: maksymalnego rozmiaru cząsteczki D_{max} i promień żyracji R_g . Wyniki te przedstawione są na rysunku 4.44.



Rysunek 4.44 **A.** Krzywe $P(r)$ obliczone dla danych zebranych w przedziale temperatur od 5 °C do 50 °C. **B.** Zależność parametrów D_{max} oraz R_g dla białka TaHSP90 od temperatury.

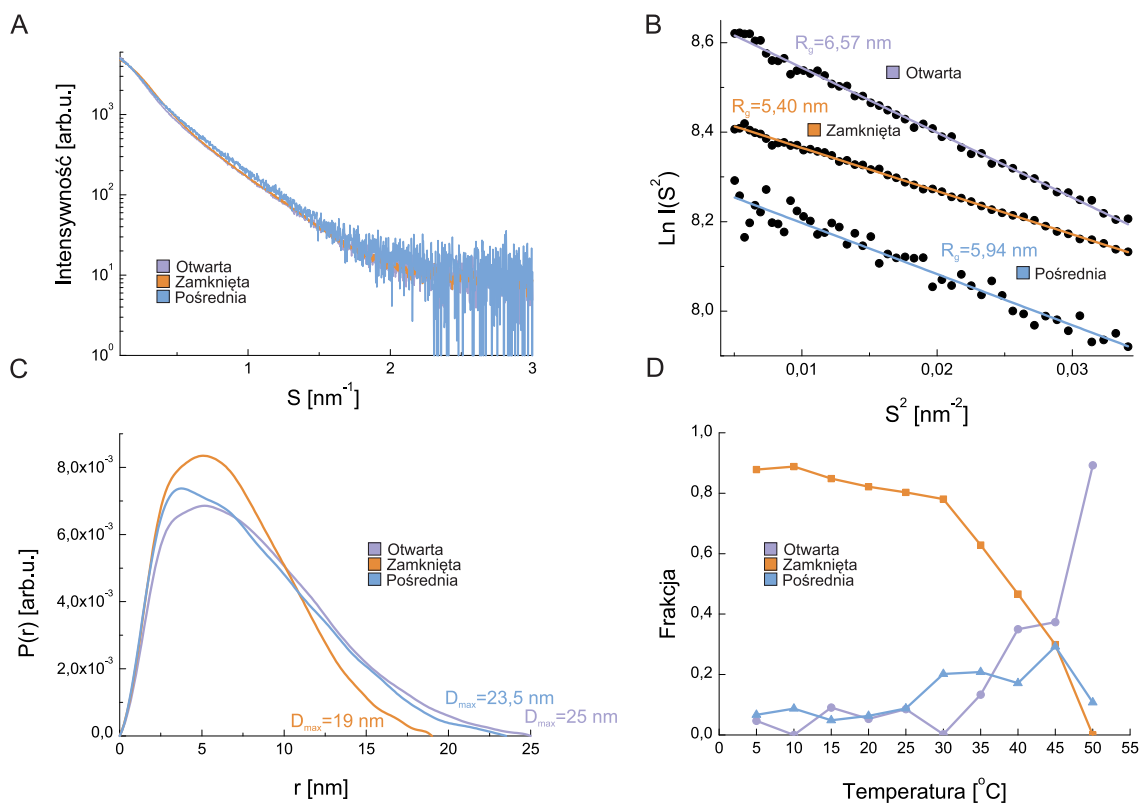


Rysunek 4.45 Analiza SVD krzywych rozpraszania dla eksperymentu SAXS kompleksowania białek TaHSP90 i HvSGT1 Δ SGS z ADP. **A, C, E, G.** Cztery pierwsze wektory własne (U_1 - U_4) uzyskane z zestawu danych SAXS. **B, D, F, H.** Odpowiadające im funkcje autokorelacji.

Z analizy wykresów przedstawionych na rysunku 4.44 można zauważyć, że do temperatury 30 °C konformacja białka TaHSP90 nie zmienia się w sposób znaczący. Dla tego zakresu temperatur krzywe $P(r)$ są praktycznie takie same. Również zmiana wartości D_{max} (20 nm w temperaturze 5 °C,

21 nm w temperaturze 30 °C) oraz R_g (5,27 nm w temperaturze 5 °C; 5,40 nm w temperaturze 30 °C) jest mała. Znaczne zmiany na krzywej $P(r)$ można zaobserwować w temperaturach powyżej 30 °C. Wartość maksimum przy około 6 nm maleje wraz ze wzrostem temperatury. Jednocześnie rośnie intensywność krzywej $P(r)$ w zakresie powyżej 10 nm. Jednocześnie można zaobserwować wzrost wartości parametrów D_{max} : od 21 nm w temperaturze 30 °C do 27 nm w temperaturze 50 °C, oraz R_g : od 5,40 nm w temperaturze 30 °C do 6,42 nm w temperaturze 50 °C. Oznacza to, że białko HSP90 z pszenicy wraz ze wzrostem temperatury zmienia swoją konformację na bardziej otwartą, charakteryzującą się większym maksymalnym rozmiarem cząsteczki D_{max} oraz R_g .

Dla zarejestrowanych krzywych rozpraszania przeprowadzono analizę SVD w celu sprawdzenia ile konformacji można zidentyfikować dla roztworu białka TaHSP90. Wyniki analizy przedstawione są na rysunku 4.45. Na rysunku 4.45 (A, C, E, G) przedstawione są cztery pierwsze wektory własne. Pierwsze trzy wektory własne są różne od szumu i mają wartości własne równe odpowiednio: $129,183 \cdot 10^3$, 2608 oraz 374.



Rysunek 4.46 Analiza MCR-ALS **A.** Krzywe rozpraszania dla trzech konformacji białka TaHSP90. **B.** Wykres Guiniera wraz z wyznaczonymi wartościami promienia żyracji R_g . **C.** Rozkład odległości wewnątrzcząsteczkowych wraz z wyznaczonymi wartościami maksymalnego rozmiaru cząsteczki D_{max} . **D.** Zależność populacji poszczególnych konformacji białka TaHSP90 od temperatury.

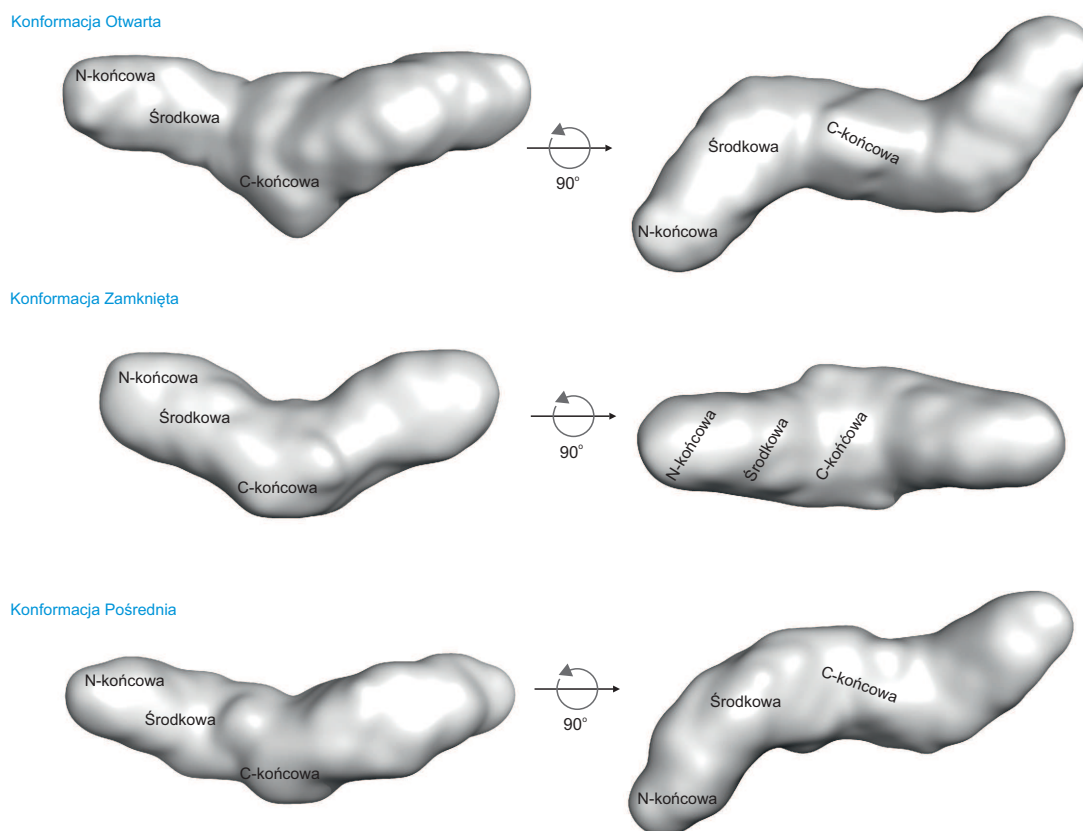
Czwarty wektor własny ma kształt szumu a jego wartość własna jest równa 284. Kolejne wektory własne mają taki sam kształt jak czwarty wektor własny i ich wartości własne mają wartości od: 194 do 263. Funkcje autokorelacji dla trzech pierwszych wektorów własnych również są różne od szumu. Natomiast dla pozostałych wektorów własnych ich funkcje autokorelacji mają charakter szumu. Dzięki analizie SVD zidentyfikowano trzy potencjalne stany konformacyjne białka TaHSP90, których stężenie zmienia się wraz z temperaturą.

Na podstawie wyników analizy SVD przeprowadzono analizę MCR-ALS w celu otrzymania krzywych rozpraszania dla poszczególnych stanów konformacyjnych oraz określenia jak zmieniają się populacje danej konformacji wraz z temperaturą. Wyniki analizy MCR-ALS przedstawione są na rysunku 4.46. Dla wyznaczonych krzywych rozpraszania wyznaczono najpierw promień żyracji R_g oraz maksymalny rozmiar cząsteczki D_{max} . Jak wynika z otrzymanych wartości promienia żyracji możemy wyróżnić konformację otwartą, analogicznie jak dla białka HtpG z *E.coli*, która charakteryzuje się największą wartością R_g wynoszącą 6,57 nm. Dwie pozostałe konformacje charakteryzują się mniejszą wartością promienia żyracji wynoszące odpowiednio 5,94 nm (konformacja pośrednia) i 5,40 nm (konformacja zamknięta). Konformacja zamknięta ma maksymalny rozmiar cząsteczki równy 19 nm w porównaniu do 23,5 nm dla konformacji pośredniej i 25 nm dla konformacji otwartej. Kształt krzywej P(r) dla konformacji zamkniętej charakteryzuje się szerokim maksimum przy $r = 5$ nm, wartość P(r) w maksimum jest większa niż dla dwóch pozostałych konformacji. Natomiast intensywność funkcji P(r) powyżej $r = 10$ nm jest mniejsza w stosunku do pozostałych konformacji. Świadczy to o mniejszych i bardziej kompaktowych wymiarach konformacji zamkniętej. W funkcji P(r) dla konformacji pośredniej możemy zauważyć, że maksimum na krzywej P(r) ma bardziej asymetryczny kształt a wartość maksymalna jest przesunięta poniżej $r = 5$ nm. Ogólnie rozkład odległości wewnątrzcząsteczkowych dla konformacji pośredniej i otwartej ma podobny kształt. Bliższe przyjrzenie się obu krzywym ukazuje jednak różnice, które wskazują na zmiany konformacyjne pomiędzy tymi dwoma konformacjami. Na szczególną uwagę zasługuje wykres (rysunek 4.46 D) przedstawiający zależność wkładu poszczególnych konformacji od temperatury. W roztworze o stężeniu 9 mg/ml białko TaHSP90 w zakresie temperatur od 5 °C do 30 °C posiada konformację zamkniętą, wkład której waha się w granicach 80 % - 88 %. Pozostałe dwie konformacje do temperatury 25 °C mają bardzo podobny wkład. Powyżej 30 °C bardzo szybko maleje wkład konformacji zamkniętej, by całkowicie zniknąć w temperaturze 50 °C. Jednocześnie wyraźnie wzrasta udział konformacji otwartej, której wkład w temperaturze 50°C wynosi 89 %. Wkład konformacji pośredniej rośnie do 29 % w temperaturze 45°C lecz spada do 10 % w temperaturze 50°C. Ewidentnie więc widoczny jest efekt dominacji konformacji otwartej w funkcji temperatury.

4.9.3 Modelowanie *ab-initio* struktury w roztworze białka HSP90 z pszenicy

Niskorozdzielczą strukturę białka TaHSP90 wymodelowano za pomocą programu DAMMIN. Dla każdej konformacji wymodelowano od dziesięciu do piętnastu niskorozdzielczych struktur z założeniem symetrii lokalnej P2 dimeru białka HSP90. Następnie uzyskane modele uśredniono w programie DAMAVER z uwzględnieniem symetrii dimeru białka HSP90, otrzymując w rezultacie model reprezentujący każdą z trzech konformacji białka HSP90. Otrzymane modele, w postaci powierzchni, przedstawiane są na rysunku 4.47 w postaci dwóch różnych orientacji. Jak widać na rysunku 4.47

trzy konformacje białka HSP90 są raczej rozciągnięte. Konformacja oznaczona jako zamknięta, jest bardziej kompaktowa w przeciwieństwie do dwóch pozostałych charakteryzuje się mniejszym kątem utworzonym przez dwie domeny środkowe. Również konformacja domeny N-terminalnej jest inna. Dla konformacji zamkniętej domena N-terminalna tworzy jedno, proste ramię z domeną środkową. Natomiast w przypadku dwóch pozostałych konformacji można zaobserwować przegięcie pomiędzy domeną N-terminalną a środkową. Przegięcie to jest bardziej uwidocznione w przypadku konformacji otwartej. Dla konformacji pośredniej przegięcie to jest mniej widoczne niż dla konformacji otwartej, ale wciąż widoczne.



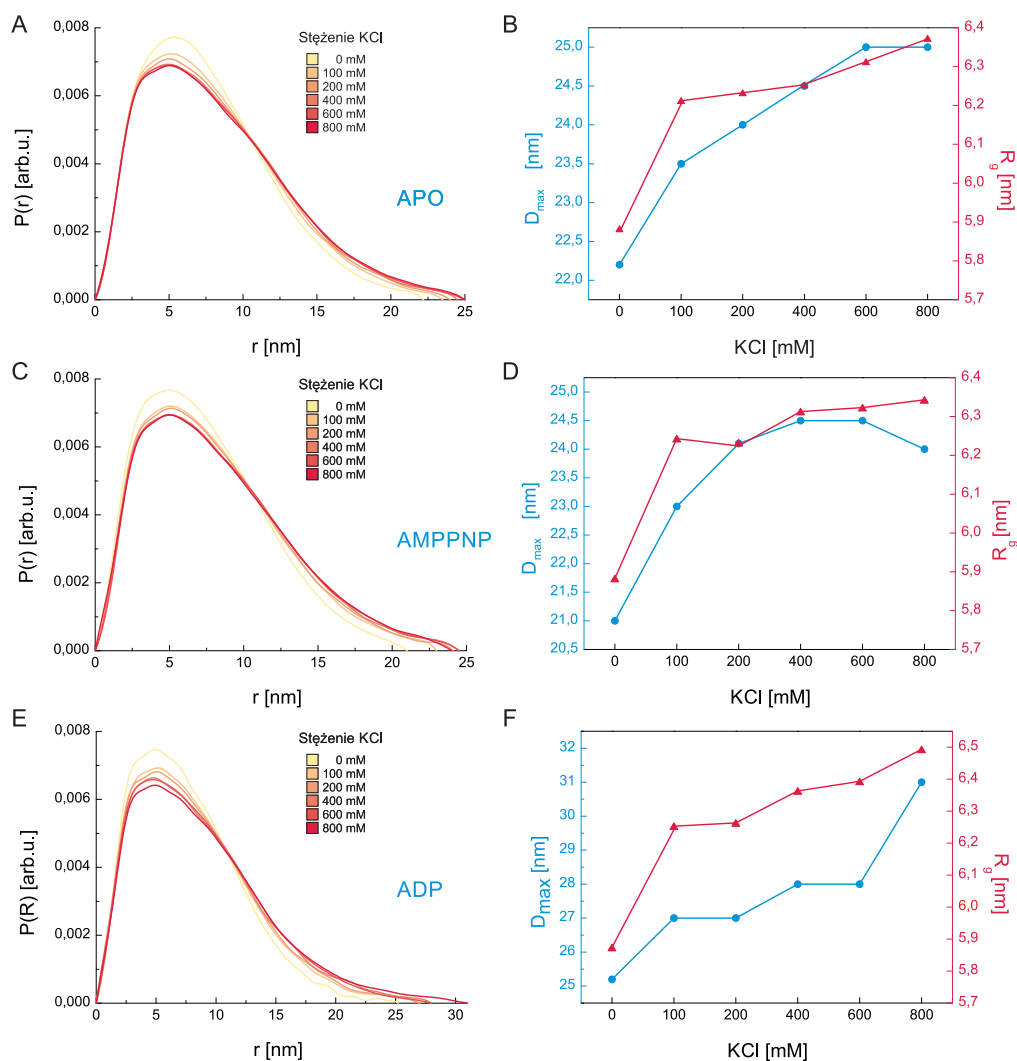
Rysunek 4.47 Modelowanie poszczególnych konformacji białka TaHSP90 z pomocą programu DAMMIN. Na modelach zaznaczone są potencjalne umiejscowienie trzech domen białka HSP90: N-końcowej, środkowej i C-końcowej.

Domena N-terminalna również odchylona jest w płaszczyźnie prostopadłej do płaszczyzny wyznaczonej przez dwie domeny środkowe. Jest do doskonale widoczne na lewej części rysunku dla konformacji otwartej i pośredniej. W konformacji zamkniętej N-terminalna domena jest bardziej

związana z domeną środkową. Natomiast w przypadku pozostałych dwóch konformacji, domena N-terminalna ma inną orientację względem domeny środkowej. W tej orientacji domena N-terminalna najprawdopodobniej dużo jest związana z domeną środkową dużo słabiej i/lub wykazują większą dynamikę.

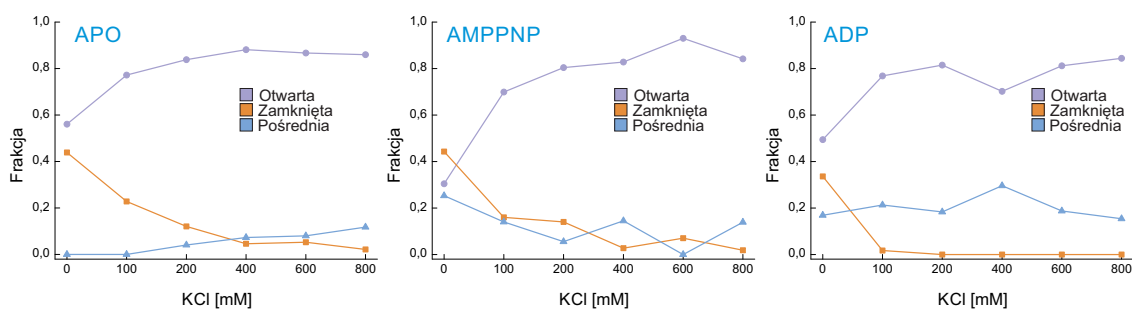
4.9.4 Wpływ siły jonowej na dynamikę konformacyjną białko HSP90 z pszenicy

Znając już jak temperatura wpływa na konformację białka HSP90 z pszenicy w roztworze, sprawdzono także, jaki wpływ ma siła jonowa na jego dynamikę konformacyjną.

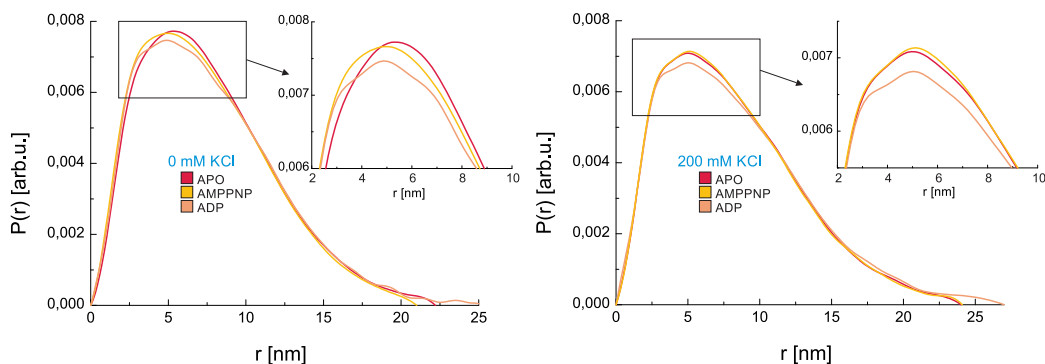


Rysunek 4.48 Krzywe $P(r)$ obliczone dla białka TaHSP90 **A.** APO, **C.** związanego z AMPPNP, **E.** związanego z ADP, w różnych stężeniach KCl. Zależność parametrów D_{max} oraz R_g od stężenia KCl dla białka TaHSP90 **B.** APO, **D.** związanego z AMPPNP, **F.** związanego z ADP.

W tym celu przeprowadzono eksperymenty małokątowego rozpraszania promieniowania rentgenowskiego dla białka HSP90 o stężeniu 4 mg/ml bez związanych nukleotydów (apo), ze związanymi nukleotydami ADP oraz analogiem ATP - AMPPNP z dodatkiem 10 mM $MgCl_2$ i KCl w stężeniu od 0 mM do 800 mM. Dla każdej krzywej rozpraszania obliczono krzywą $P(r)$ oraz parametry D_{max} oraz R_g . Wyniki te przedstawione są na rysunku 4.48. Analizując krzywe $P(r)$ dla białka TaHSP90 można zauważyć, że bez względu na to czy białko było związane z nukleotydami czy też nie, wraz ze wzrostem siły jonowej zmienia się kształt krzywej $P(r)$. Zmiana ta jest analogiczna do zmiany zaobserwowanej dla eksperymentów temperaturowych. Wraz ze wzrostem siły jonowej maleje wartość maksimum krzywej $P(r)$ przy około 6 nm natomiast rosną wartości powyżej 10 nm. Nie mniej jednak wartość D_{max} oraz R_g są większe niż dla eksperymentów temperaturowych. Również zmiany krzywej $P(r)$ są mniej widoczne. Najprawdopodobniej jest to związane z mniejszym stężeniem białka wykorzystywanym w tych eksperymentach ponieważ w analogicznych warunkach tj. stężenia soli pomiędzy 100 mM a 200 mM i temperatury 15 °C, przy stężeniu 9 mg/ml białko TaHSP90 cechuje się mniejszą wartością D_{max} oraz R_g . Pomimo tego, że zmiany są mniej widoczne, łatwo można dostrzec że, wraz ze wzrostem siły jonowej, bez względu na związane nukleotydy białko TaHSP90 zmienia stopniową swoją konformację na bardziej otwartą.



Rysunek 4.49 Wpływ siły jonowej na populacje poszczególnych konformacji białka TaHSP90.



Rysunek 4.50 Porównanie konformacji białka TaHSP90: bez oraz z związanymi nukleotydami AMPPNP i ADP. Krzywe $P(r)$ dla białka TaHSP90 przy stężeniu soli: **A.** 0 mM oraz **B.** 200 mM.

Z eksperymentów temperaturowych wynika, że białko HSP90, może występować w trzech różnych konformacjach. Zestaw krzywych SAXS dla tych trzech konformacji wykorzystano w celu zrekonstruowania danych otrzymanych przy różnych stężeniach siły jonowej za pomocą ich kombinacji liniowej. Dopasowywanie przeprowadzono w programie Oligomer z pakietu ATSAS. Wyniki dopasowania (w postaci wkładu poszczególnych konformacji do danej krzywej SAXS) przedstawione są na rysunku 4.49 oraz w tabeli 4.3. Jak wynika z dopasowania w programie Oligomer białko TaHSP90 w różnych stężeniach soli, jak również związane z różnymi nukleotydami może być opisane jako superpozycję trzech konformacji wyznaczonych z eksperymentów temperaturowych. Wartości χ^2 dopasowania prawie we wszystkich przypadkach są mniejsze niż 2 co świadczy o bardzo dobrym dopasowaniu. Jedynie w przypadku białka TaHSP90 bez związanego nukleotydu przy stężeniu 0mM dopasowanie jest gorsze o czym świadczy wartość $\chi^2 = 5,66$. TaHSP90 w stężeniach KCl od 100 mM występuje głównie w konformacji otwartej, niezależnie od związanych nukleotydów. Jedynie przy stężeniu równym 0 mM frakcja konformacji otwartej jest mniejsza niż 0,6. Dla białka ze związanym nukleotydem AMPPNP przeważa konformacja zamknięta przy 0 mM KCl, której frakcja wynosi 0,443. Natomiast dla związanego ADP jak również dla białka bez nukleotydów przeważa konformacja otwarta, której frakcja wynosi odpowiednio 0,494 i 0,561. Przy 0 mM KCl różnice pomiędzy krzywymi $P(r)$ dla różnych związanych nukleotydów i białka bez nukleotydu są najbardziej widoczne jak można zauważyć z rysunku 4.50, ale kształt całej krzywej $P(r)$ jest bardzo podobny. Związanie nukleotydu najprawdopodobniej powoduje lokalne zmiany konformacyjne w obrębie domeny N-terminalnej gdzie znajduje się miejsce wiązania nukleotydów. Przy 200 mM kształt krzywej dla białka bez nukleotydu i ze związanym AMPPNP jest praktycznie identyczny natomiast dla związanego ADP, kształt jest bardzo podobny jedynie wartość maksimum jest mniejsza. Białko ze związanym nukleotydem ADP ma dużo mniejszy udział konformacji zamkniętej w porównaniu do białka bez nukleotydu i związanego z AMPPNP.

Tabela 4.3 Dopasowanie trzech konformacji białka TaHSP90 do danych SAXS w różnych stężeniach soli oraz związanych nukleotydów. W tabeli podane są wartości χ^2 opisujące jakość dopasowania. Dla każdej z trzech konformacji podany jest również jej wkład przy danym stężeniu siły jonowej.

APO						
C KCl	0 mM	100 mM	200 mM	400 mM	600 mM	800 mM
χ^2	5,66	1,54	1,26	1,32	1,51	0,95
Otwarta	0,561	0,772	0,838	0,881	0,867	0,860
Zamknięta	0,439	0,288	0,121	0,046	0,053	0,022
Pośrednia	0	0	0,041	0,073	0,08	0,118

AMPPNP						
C KCl	0 mM	100 mM	200 mM	400 mM	600 mM	800 mM
χ^2	1,98	1,31	1,22	1,14	1,48	1,15
Otwarta	0,304	0,699	0,804	0,828	0,93	0,842
Zamknięta	0,443	0,16	0,14	0,027	0,07	0,018
Pośrednia	0,253	0,14	0,055	0,145	0	0,139

ADP						
C KCl	0 mM	100 mM	200 mM	400 mM	600 mM	800 mM
χ^2	1,17	1,72	1,48	1,29	1,71	1,98
Otwarta	0,494	0,768	0,815	0,702	0,812	0,844
Zamknięta	0,336	0,017	0	0	0	0
Pośrednia	0,169	0,213	0,183	0,296	0,187	0,154

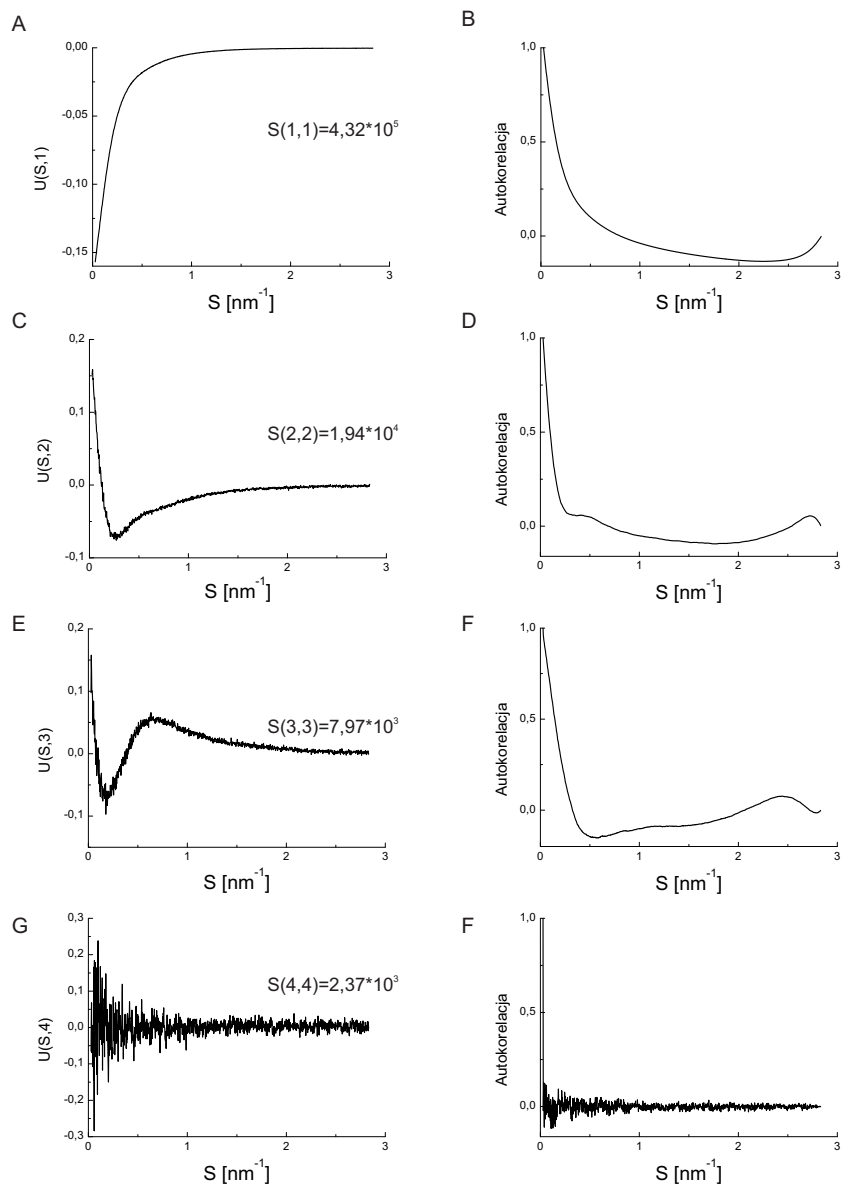
4.10 Badanie kompleksu białek HSP90 i SGT1 z ADP

Białko SGT1 oddziałuje najsilniej z białkiem HSP90 w kompleksie z adenozyno-5- di-fosforanem (ADP). Jak dotąd udało się zbadać strukturę kompleksu domeny N-terminalnej HSP90 z domeną CS białka SGT1 oraz w kompleksie z dodatkową domeną CHORD II białka RAR1 (Zhang i wsp., 2010a). Brak jest jednak informacji na temat stechiometrii łączenia się białek kompleksu oraz jego struktury w kontekście pełnej długości białka HSP90. W celu zbadania kompleksu białek HSP90 i SGT1 z ADP przeprowadzono eksperymenty małokątowego rozpraszania promieniowania synchrotronowego dla kompleksu utworzonego z pełnej długości białka HSP90 z pszenicy (TaHSP90) z białkiem SGT1 z jęczmienia bez domeny SGS (HvSGT1 Δ SGS), kompleksu białka HSP90 z jęczmienia bez domeny C-terminalnej (HvHSP90 Δ C) z białkiem SGT1 z jęczmienia bez domeny SGS oraz kompleksu białka HSP90 z jęczmienia bez domeny C-terminalnej z domeną CS białka SGT1 (HvSGT1-CS) z jęczmienia. Wszystkie kompleksy zbadano w obecności ADP.

4.10.1 Struktura kompleksu TaHSP90-HvSGT1 Δ SGS z ADP

W celu zbadania struktury kompleksu TaHSP90-HvSGT1 Δ SGS przeprowadzono pomiary SAXS dla różnych wartości stosunków molowych pełnej długości białka TaHSP90 oraz białka HvSGT1 Δ SGS, tak aby można było przeprowadzić analizę SVD oraz analizę MCR-ALS. Podejście te miało na celu otrzymanie krzywej rozpraszania dla kompleksu bez artefaktów pochodzących od rozpraszania nieskompleksowanych białek TaHSP90 oraz HvSGT1 Δ SGS. Otrzymanie krzywej dla samego kompleksu jest również potrzebne do poprawnego określenia wartości $I(0)$ potrzebnych do oszacowania masy molekularnej kompleksu oraz zbadania jego stechiometrii. W eksperymencie zarejestrowano krzywe rozpraszania dla białka TaHSP90 z ADP, HvSGT1 Δ SGS oraz potencjalnego kompleksu tych białek zmierzonych w stosunku molowym HvSGT1 Δ SGS do TaHSP90: 2:1, 1,7:1, 1,5:1, 1,2:1, 1:1, 1:1,2. Dla zarejestrowanych krzywych SAXS przeprowadzono analizę SVD. W wyniku analizy SVD spodziewano się otrzymać trzy niezerowe wektory własne odpowiadające krzywom rozpraszania dla białek: TaHSP90, HvSGT1 Δ SGS oraz ich kompleksu. Cztery pierwsze wektory własne wraz z charakteryzującymi je wartościami własnymi oraz funkcjami autokorelacji przedstawione są na rysunku 4.51. Jak wynika z analizy wektorów własnych i ich funkcji autokorelacji, w przypadku zestawu danych użytego w analizie SVD, trzy pierwsze wektory własne są różne od szumu i reprezentują trzy różne krzywe rozpraszania dla białek: TaHSP90, HvSGT1 Δ SGS oraz kompleksu białka TaHSP90 z białkiem HvSGT1 Δ SGS. Istnienie tylko dwóch niezerowych wektorów własnych sugerowałoby, że mamy do czynienia jedynie z mieszaniną białek TaHSP90 z HvSGT1 Δ SGS w różnych proporcjach. Następnie przeprowadzono analizę MCR-ALS w celu otrzymania krzywej rozpraszania dla kompleksu białek TaHSP90 i HvSGT1 Δ SGS z ADP. Wyniki analizy MCR-ALS przedstawione są na rysunku 4.52. Na podstawie krzywej rozpraszania dla kompleksu obliczoną krzywą $P(r)$ i porównano ją z krzywą obliczoną dla białka TaHSP90. Znormalizowane do wartości $I(0)$ krzywe $P(r)$ dla TaHSP90 i kompleksu różnią się zasadniczo. Maksimum na krzywej $P(r)$ dla kompleksu ma mniejszą wartość niż odpowiadający jemu obszar na krzywej dla białka TaHSP90, natomiast powyżej wartości 9 nm krzywa $P(r)$ przyjmuje wartości większe w stosunku do białka TaHSP90 co wskazuje na utworzenie się kompleksu białek TaHSP90 i HvSGT1 Δ SGS w obecności ADP. Również utworzenie kompleksu ewidentnie jest widoczne w analizie maksymalnego rozmiaru cząsteczki D_{max} oraz promienia żyracji

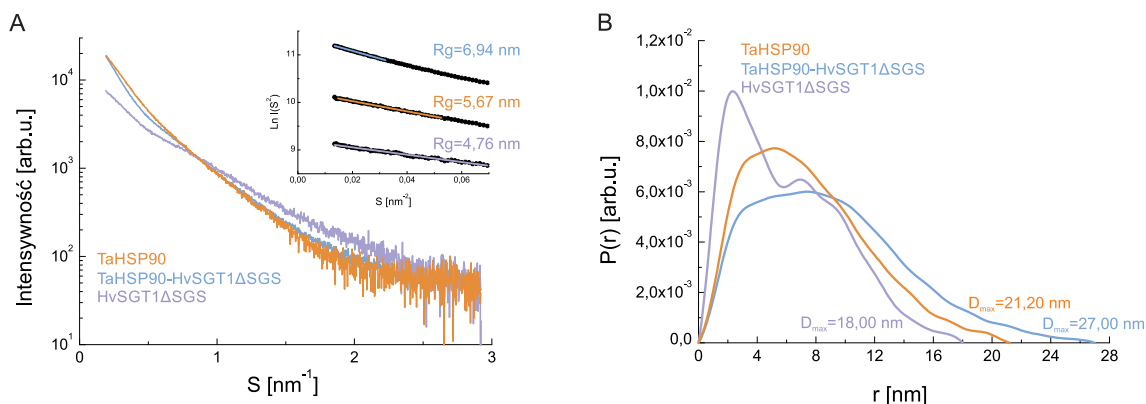
R_g . Dla TaHSP90 wynoszą one odpowiednio 21 nm oraz 5,67 nm w porównaniu do 28 nm oraz 6,94 nm dla kompleksu TaHSP90-HvSGT1 Δ SGS.



Rysunek 4.51 Analiza SVD krzywych rozpraszania dla eksperymentu SAXS monitorującego proces kompleksowania białek TaHSP90 i HvSGT1 Δ SGS z ADP. **A, C, E, G** Cztery pierwsze wektory własne (U_1 - U_4) uzyskane z zestawu danych SAXS; **B, D, F, H** odpowiadające im funkcje autokorelacji.

Analiza ekstrapolowanego sygnału $I(0)$ niespodziewanie ujawniła, że w kompleksie TaHSP90-HvSGT1 Δ SGS znajduje się tylko jedna kopia białka HvSGT1 Δ SGS pomimo tego, że HvSGT1 Δ SGS jest dimerem w roztworze jak również pomimo tego, że w dimerze białka HSP90 są dwa miejsca

wiązania domeny CS w obrębie domen N-terminalnych. Także, można byłoby spodziewać się symetrycznego kompleksu o stechiometrii 4:2 (podjednostki SGT1 do podjednostek HSP90). W tabeli 4.4 znajdują się porównanie wartości $I(0)$ dla krzywych rozpraszania dla białek: TaHSP90, HvSGT1 Δ SGS oraz kompleksu. Wartość D_{max} dla białka HvSGT1 Δ SGS wynosi 18 nm, i jest około dwa razy większy niż różnica D_{max} , pomiędzy białkiem TaHSP90 a kompleksem co może wskazywać na wiązanie się jednej podjednostki białka HvSGT1 Δ SGS. Oczywiście nie można odrzucić całkowicie hipotezy, że różnica D_{max} wynika ze zmian konformacyjnych w kompleksie a nie z wiązania tylko jednej podjednostki białka HvSGT1 Δ SGS.



Rysunek 4.52 A. Eksperymentalne krzywe rozpraszania dla białek TaHSP90 i HvSGT1 Δ SGS oraz kompleksu TaHSP90-HvSGT1 Δ SGS (otrzymane z procedury MCR-ALS). Na mniejszym wykresie przedstawiony jest wykres Guiniera wraz z wyznaczonymi wartościami promienia żyracji R_g . **B.** Krzywe $P(r)$ dla białek TaHSP90 i HvSGT1 Δ SGS oraz kompleksu TaHSP90-HvSGT1 Δ SGS oraz wyznaczone w programie GNOM maksymalne rozmiary cząsteczki D_{max} .

Tabela 4.4 Analiza wartości $I(0)$ dla zestawu danych otrzymanych z analizy MCR-ALS.

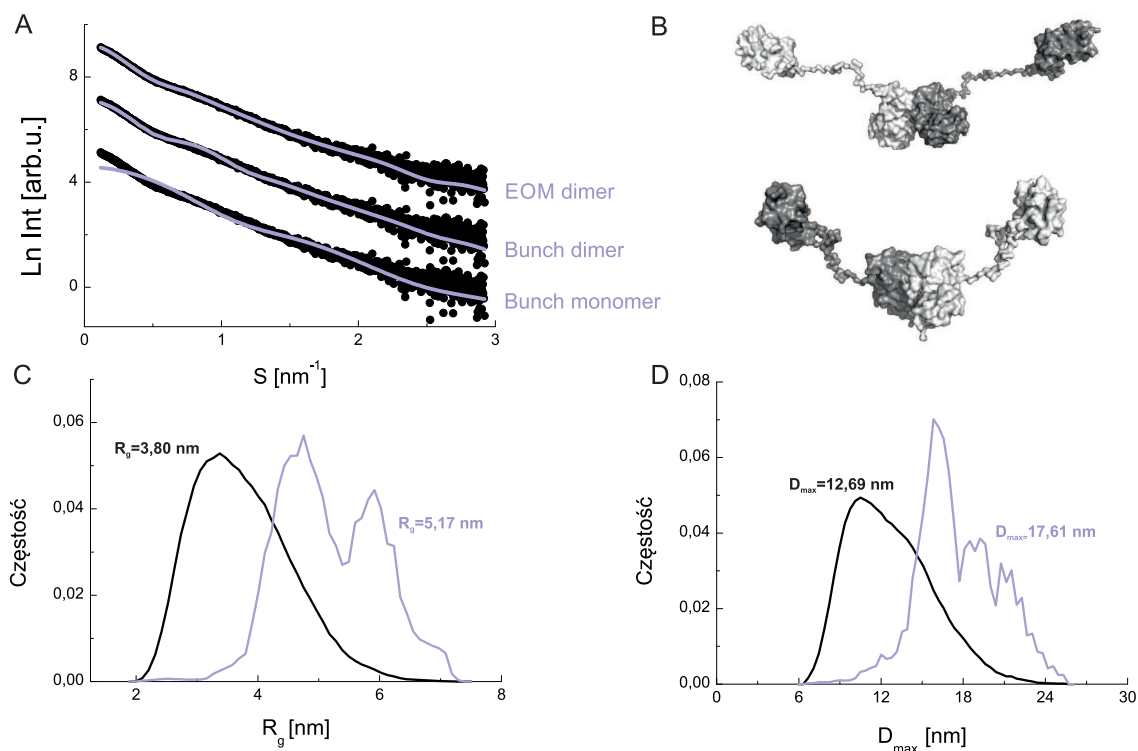
Białko	Masa molekularna ¹	$I(0)_{SAXS}$ ²	$I(0)_{białko}/$ $I(0)_{HvSGT1\Delta SGS}$	$MM_{białko}/$ $MM_{HvSGT1\Delta SGS}$
HvSGT1 Δ SGS	56,833 kDa	10040	1	1
TaHSP90	161,167 kDa	28580	2,847	2,836
TaHSP90-HvSGT1 Δ SGS	218,000 kDa (2:2)	33560	3,343	3,836
TaHSP90-HvSGT1 Δ SGS	274,833 kDa (2:4)	33560	3,343	4,836
TaHSP90-HvSGT1ΔSGS	189,583 kDa (2:1)	33560	3,343	3,336

¹Wyliczona na podstawie sekwencji dla dimerów białek HvSGT1 Δ SGS i TaHSP90 oraz stosunków molowych tych białek podanych w nawiasie (TaHSP90:HvSGT1 Δ SGS), ²Wartość $I(0)$ obliczona w programie GNOM.

4.11 Modelowanie *ab-initio* struktury białek TaHSP90, HvSGT1 Δ SGS oraz kompleksu TaHSP90-HvSGT1 Δ SGS

4.11.1 Modelowanie białka HvSGT1 Δ SGS na podstawie danych SAXS

W celu sprawdzenia jaka jest stechiometria i konformacja białka HvSGT1 Δ SGS w roztworze przeanalizowano zebrane dane SAXS za pomocą programu DAMMIN oraz modelowania bryły sztywnej w programach BUNCH i EOM. Analizę bryły sztywnej najpierw przeprowadzono zakładając, że białko HvSGT1 Δ SGS jest monomerem. Otrzymane dopasowanie do danych eksperymentalnych przedstawione jest na rysunku 4.53. Następnie tą samą analizę przeprowadzono dla dimeru HvSGT1 Δ SGS z założeniem symetrii lokalnej P2. Ostatnim etapem analizy tego układu było modelowanie bryły sztywnej, które przeprowadzono w programie EOM z założeniem symetrii lokalnej P2. Najlepsze dopasowanie do danych eksperymentalnych otrzymano w programie EOM.

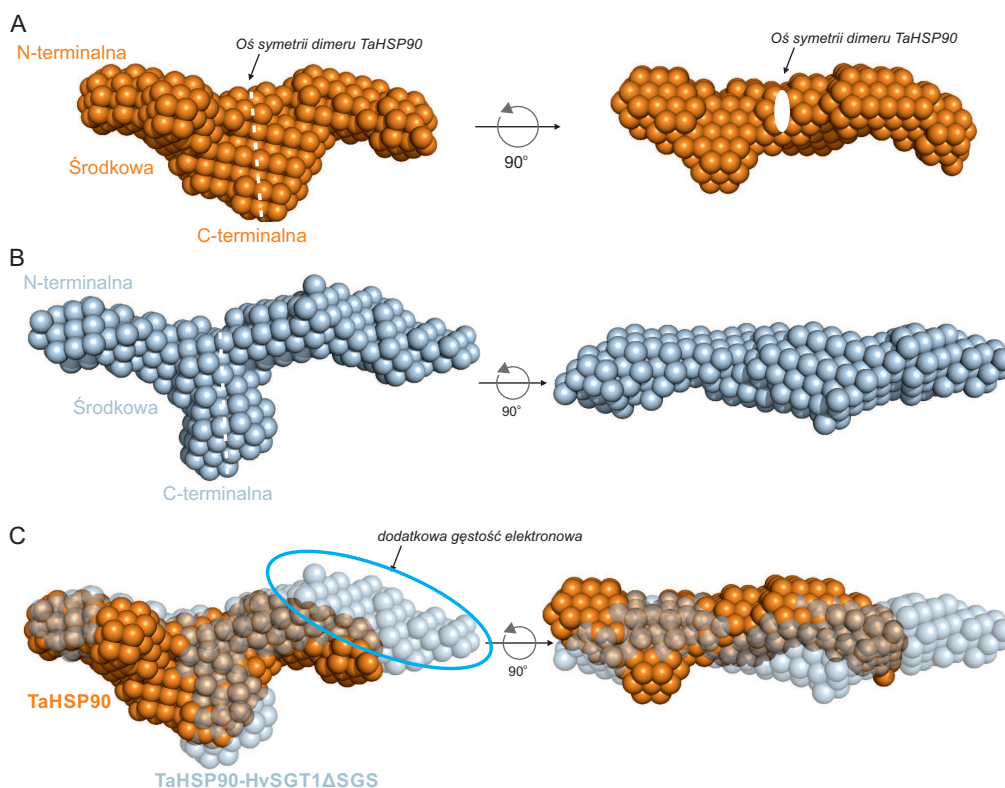


Rysunek 4.53 Analiza wyników modelowania bryły sztywnej białka HvSGT1 Δ SGS. **A.** Dopasowanie krzywych otrzymanych w programie BUNCH oraz EOM do danych eksperymentalnych. **B.** Przykładowe modele z puli modeli wybranej w programie EOM. **C, D** Rozkład promienia żyracji R_g (C) oraz maksymalnego rozmiaru cząsteczki D_{max} (D) dla losowej puli (kolor czarny) oraz wybranej puli modeli (kolor fioletowy). Na rysunkach przedstawione są również średnie wartości R_g i D_{max} dla losowych i wybranych puli modeli.

Z analizy tej wynika, że białko HvSGT1 Δ SGS w roztworze jest dimerem, który występuje w równowadze kilku konformacji. Różnią się one między sobą strukturą łącznika VR1 pomiędzy domeną TPR i CS. Rozkłady R_g i D_{max} dla białka HvSGT1 Δ SGS otrzymane w programie EOM różnią się od rozkładu dla losowej puli modeli i charakteryzują się dość dużą wartością zarówno R_g jak i D_{max} . Świadczy to o tym, że podobnie jak pełnej długości białko SGT1 z jęczmienia, białko HvSGT1 Δ SGS przyjmuje raczej rozciągniętą konformację.

4.11.2 Modelowanie struktury białka TaHSP90 oraz kompleksu białek TaHSP90 - HvSGT1 Δ SGS

Modelowanie niskorozdzielczej struktur badanych białek przeprowadzono z wykorzystaniem algorytmu Monte Carlo zaimplementowanego w programie DAMMIN. Modele otrzymane dla białka TaHSP90 oraz kompleksu TaHSP90-HvSGT1 Δ SGS przedstawione są na rysunku 4.54. Każdy model jest wynikiem uśrednienia w programie DAMAVER dziesięciu modeli otrzymanych w programie DAMMIN. Dla białek TaHSP90 oraz kompleksu TaHSP90-HvSGT1 Δ SGS nie zakładano żadnej symetrii lokalnej w celu dokładnego zbadania konformacji kompleksu bez narzucania jakichkolwiek więzów.



Rysunek 4.54 Modelowanie białka TaHSP90 oraz kompleksu białek TaHSP90-HvSGT1 Δ SGS w programie DAMMIN. **A.** Model dimeru białka TaHSP90. **B.** Model kompleksu TaHSP90-HvSGT1 Δ SGS.

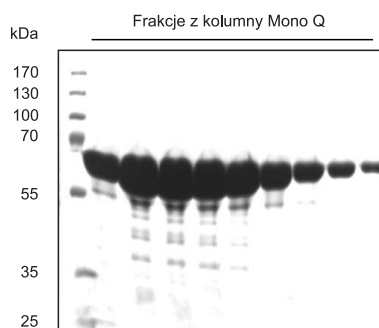
C. Nałożenie modelu białka TaHSP90 na model kompleksu.

Model białka TaHSP90 jest symetryczny nawet bez zakładania symetrii P2, która charakteryzuje dimer białka HSP90. Sam model ma kształt zbliżony do struktur otrzymanych na podstawie danych SAXS dla białka HSP90 z *Homo sapiens* i *Sus scrofa* f. *domestica* (świnia domowa) (Krukenberg i wsp., 2009b, Bron i wsp., 2008). W przeciwieństwie do modelu białka TaHSP90 model kompleksu białek TaHSP90-HvSGT1 Δ SGS jest asymetryczny. Jedno ramię w modelu, które składa się z środkowej i N-terminalnej domeny białka TaHSP90 jest znacznie dłuższe. Ponieważ miejsce wiązania białka SGT1 znajduje się w obrębie domeny N-terminalnej, przedłużenie ramienia można zidentyfikować jako białko HvSGT1 Δ SGS. Wymiary przedłużenia, które przedstawiono na rysunku 4.54, są o połowę mniejsze niż maksymalny wymiar cząsteczki HvSGT1 Δ SGS (otrzymanej w programie DAMMIN). Jest to zgodne z wynikami otrzymanymi z analizy wartości $I(0)$, które sugerowało wiązanie jednej podjednostki białka HvSGT1 Δ SGS do dimeru białka TaHSP90.

4.12 Badanie struktury kompleksu białek HvHSP90 Δ C i HvSGT1 Δ SGS

4.12.1 Oczyszczanie białka HvHSP90 Δ C z bakterii *E. coli*

Do pomiarów SAXS białko HvHSP90 Δ C oczyszczano z bakterii *E. coli*. Białko HvHSP90 Δ C zostało oczyszczone według protokołów opisanych w sekcji 3.6.1-3.6.3, przy użyciu chromatografii powinowactwa, cięcia proteazą TEV, ponownym oczyszczaniu za pomocą chromatografii powinowactwa, sączenia molekularnego oraz chromatografii jonowymiennej jako ostatniego etapu oczyszczania. Na rysunku 4.55 pokazana jest czystość frakcji białka HvHSP90 Δ C otrzymanych po oczyszczaniu za pomocą chromatografii jonowymiennej na kolumnie MonoQ.



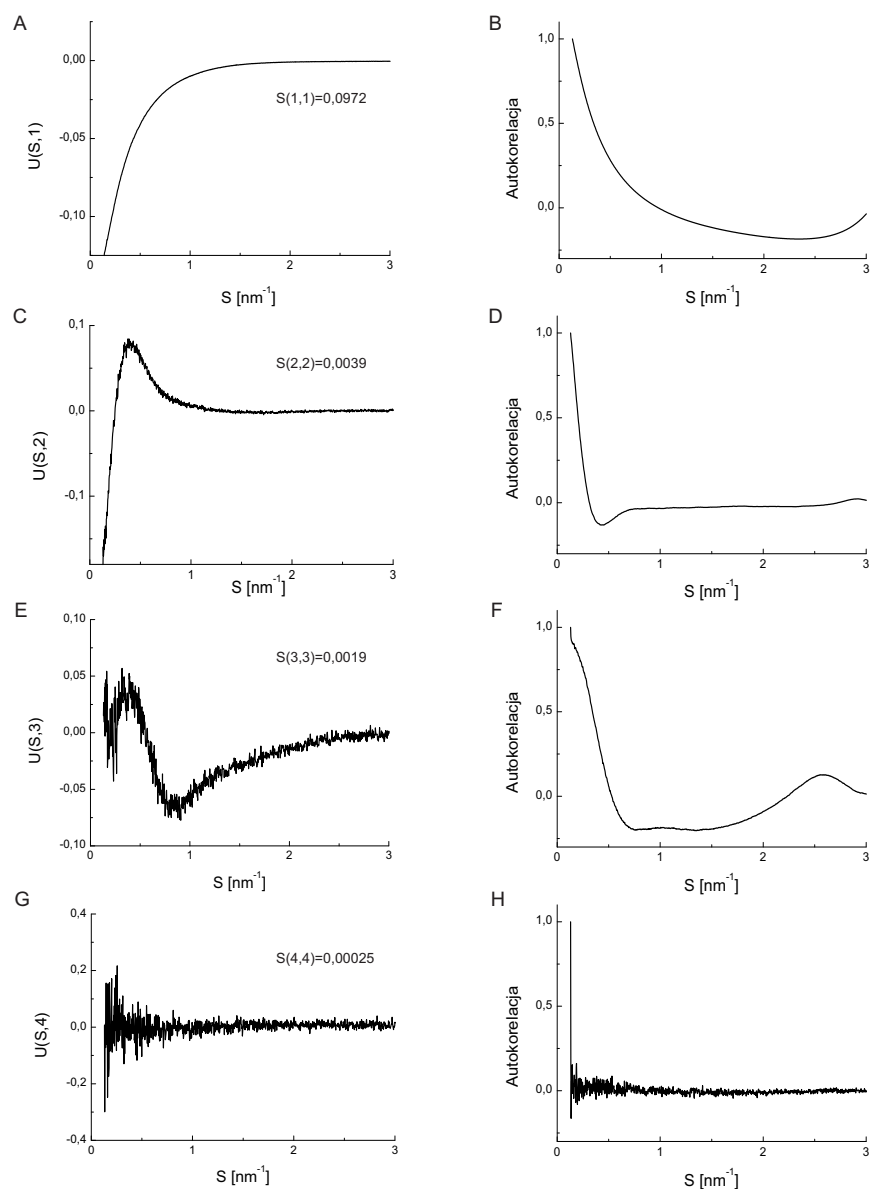
Rysunek 4.55 Czystość frakcji białka HvHSP90 Δ C podczas ostatniego etapu oczyszczania za pomocą chromatografii jonowymiennej na kolumnie MonoQ przedstawione w postaci zdjęcia elektroforezy w żelu polikarylamidowym.

Za pomocą opisanego wyżej protokołu udało się otrzymać białko o odpowiedniej czystości oraz stosunku absorbancji przy 260 nm i 280 nm: $A_{260nm}/A_{280nm} \approx 0,5$, który świadczy to o tym, że białko HvHSP90 Δ C nie jest związane z nukleotydami.

4.12.2 Badania SAXS kompleksu białek HvHSP90 Δ C i HvSGT1 Δ SGS

Aby stwierdzić czy dimeryzacja białka HSP90 jest potrzebna do wiązania białka HvSGT1 Δ SGS oraz czy w kompleksie tym białko to występuje w formie monomerycznej czy dimerycznej przeprowadzono badania małokątowego rozpraszania promieniowania rentgenowskiego dla kompleksu białek HvHSP90 Δ C z białkiem HvSGT1 Δ SGS. Białko HvHSP90 Δ C pozbawione jest domeny C-terminalnej odpowiedzialnej za dimeryzację białka HSP90 i w roztworze występuje w formie monomeru. Poprzez zbadanie kompleksu białka HvHSP90 Δ C z białkiem HvSGT1 Δ SGS zamierzano sprawdzić czy dimeryzacja HSP90 jest potrzebna do wiązania monomeru białka HvSGT1 Δ SGS, czy też w kompleksie z monomerem białka HvHSP90 Δ C białko HvSGT1 Δ SGS występuje w postaci dimeru. W eksperymencie tym zmierzono krzywe rozpraszania dla pojedynczych białek: HvHSP90 Δ C i HvSGT1 Δ SGS oraz kompleksu białek HvHSP90 Δ C z HvSGT1 Δ SGS w roztworze o stosunku molowym 1,5:1; 1,2:1; 1:1; 1:1,2; 1:1,5. Podobnie jak w przypadku kompleksu białek TaHSP90 i HvSGT1 Δ SGS na początku przeprowadzono analizę SVD i MCR-ALS w celu sprawdzenia czy utworzył się kompleks pomiędzy białkami HvHSP90 Δ C i HvSGT1 Δ SGS oraz w celu otrzymania krzywej rozpraszania dla kompleksu.

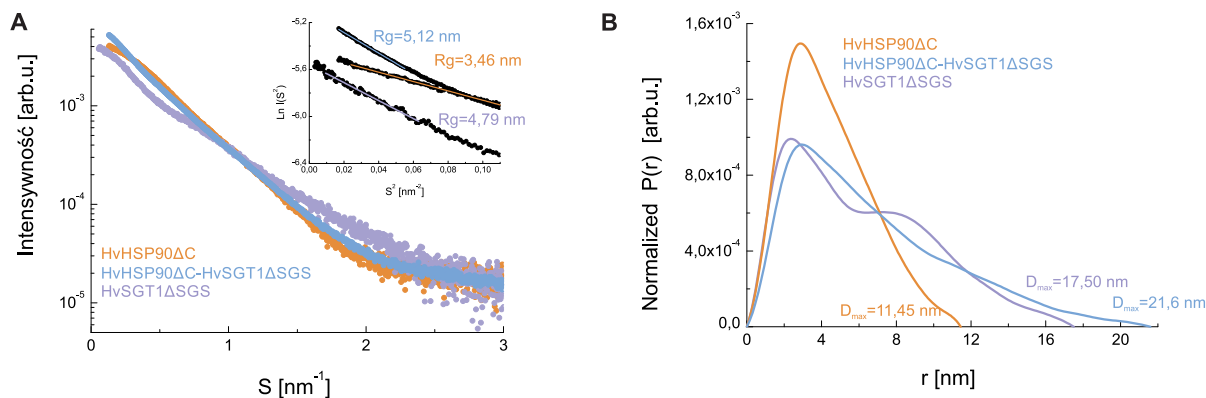
Wyniki analizy SVD w postaci pierwszych czterech wektorów własnych oraz ich funkcji autokorelacji przedstawione są na rysunku 4.56.



Rysunek 4.56 Analiza SVD krzywych rozpraszania dla eksperymentu SAXS kompleksowania białek HvHSP90 Δ C i HvSGT1 Δ SGS z ADP. **A, C, E, G** Cztery pierwsze wektory własne (U1-U4) uzyskane z zestawu danych SAXS. **B, D, F, H** Odpowiadające im funkcje autokorelacji.

Jak wynika z rysunku 4.56 pierwsze trzy wektory własne są różne od szumu jak również odpowiadające im funkcje autokorelacji. Wynik ten potwierdza istnienie trzech komponentów w analizowanym zestawie danych SAXS, które można interpretować jako nieskompleksowane białka HvHSP90 Δ C i HvSGT1 Δ SGS oraz ich kompleks. Następnie przeprowadzono analizę MCR-ALS, w której otrzy-

mano krzywą rozpraszania dla kompleksu białek HvHSP90 Δ C i HvSGT1 Δ SGS. Krzywe rozpraszania dla białek HvHSP90 Δ C i HvSGT1 Δ SGS oraz ich kompleksu przedstawione są na rysunku 4.57 A. Na podstawie danych SAXS wyznaczono promień żyrcji R_g . Białko HvHSP90 Δ C charakteryzuje się promieniem żyrcji równym 3,46 nm a białko HvSGT1 Δ SGS posiada promień żyrcji 4,79 nm. Dla kompleksu białek HvHSP90 Δ C i HvSGT1 Δ SGS wartość R_g wynosi 5,12 nm. W programie GNOM wyznaczono również krzywe rozkładu odległości wewnątrzcząsteczkowych, które przedstawione są na rysunku 4.57 B. Wyznaczone krzywe $P(r)$ różnią się od siebie. Dla białka HvHSP90 Δ C rozkład jest bardziej zwarty i jest bardziej charakterystyczny dla białka o kompaktowym kształcie. Natomiast zarówno kształty krzywych $P(r)$ dla białka HvSGT1 Δ SGS jak również kompleksu wskazują, że białka te mają wydłużony kształt. Białka HvHSP90 Δ C oraz HvSGT1 Δ SGS charakteryzują się odpowiednio wartością D_{max} równą 11,45 nm i 17,50 nm. Wartość D_{max} wyznaczona dla kompleksu wynosi natomiast 21,6 nm. Wartość ta jest mniejsza niż suma wartości D_{max} dla pojedynczych białek. Może to sugerować, że białka te występują w kompleksie w stosunku molowym 1:1, ponieważ wartość D_{max} dla kompleksu jest bardziej zbliżona do sumy połowy wartości D_{max} dla białka HvSGT1 Δ SGS (które jest dimerem) i wartości D_{max} dla białka HvHSP90 Δ C.



Rysunek 4.57 A. Eksperymentalne krzywe SAXS dla białek HvHSP90 Δ C i HvSGT1 Δ SGS oraz kompleksu (otrzymana z procedury MCR-ALS). Na mniejszym wykresie przedstawiony jest wykres Guiniera wraz z wyznaczonymi wartościami promienia żyrcji R_g . **B.** Krzywe $P(r)$ dla białek HvHSP90 Δ C i HvSGT1 Δ SGS oraz kompleksu HvHSP90 Δ C-HvSGT1 Δ SGS oraz wyznaczone w programie GNOM maksymalne rozmiary cząsteczki D_{max} .

W celu dokładnego zbadania stechiometrii kompleksu przeprowadzono szczegółową analizę wartości parametru $I(0)$ wyznaczonego w programie GNOM. Wynik analizy przedstawiony jest w tabeli 4.5. W tabeli 4.5 przeanalizowano stosunek masy molekularnej (w kDa), obliczonej na podstawie sekwencji aminokwasowej badanych białek, do masy molekularnej dimeru białka HvSGT1 Δ SGS. Podobnie przeanalizowano stosunek wartości $I(0)$ dla badanych białek do wartości $I(0)$ dla dimeru białka HvSGT1 Δ SGS. Ponieważ wartość $I(0)$ jest proporcjonalna do masy molekularnej cząsteczki rozpraszającej, stosunki masowe oraz stosunek $I(0)$ powinny być takie same. W przypadku kompleksu o

stosunku molowym 1:2 (HvHSP90 Δ C do HvSGT1 Δ SGS) wartości stosunków $I(0)$ i masy molekularnej różnią się od siebie. Natomiast w przypadku kompleksu o stechiometrii 1:1 wartości stosunku $I(0)$ oraz masy molekularnej są bardzo podobne do siebie. Z analiz wartości parametrów D_{max} oraz stosunków wartości $I(0)$ wynika, że w kompleksie z białkiem HvHSP90 Δ C, białko HvSGT1 Δ SGS występuje w postaci monomeru.

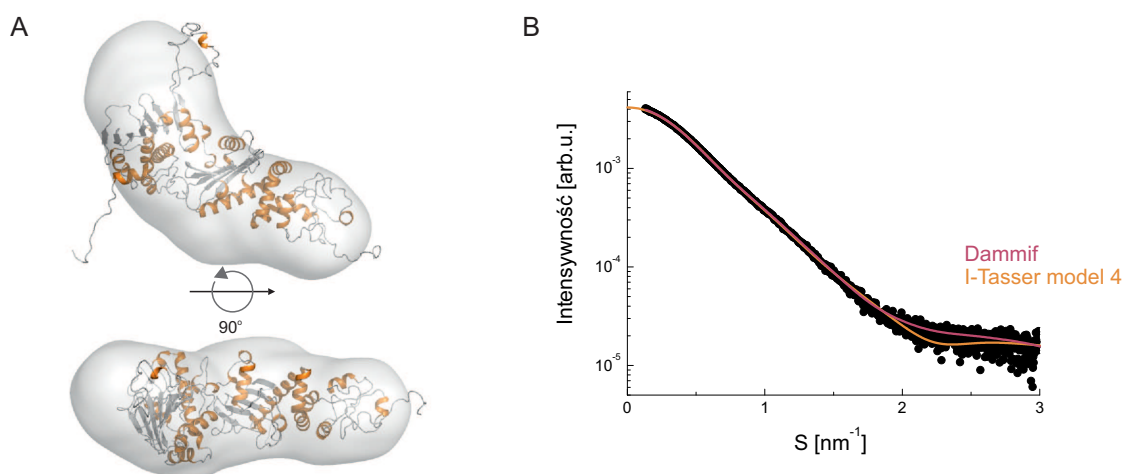
Tabela 4.5 Analiza wartości $I(0)$ dla zestawu danych otrzymanych z analizy MCR-ALS.

Białko	Masa molekularna ¹	$I(0)_{SAXS}$ ²	$I(0)_{białko}/$ $I(0)_{HvSGT1\Delta SGS}$	$MM_{białko}/$ $MM_{HvSGT1\Delta SGS}$
HvSGT1 Δ SGS	56,833 kDa	0,00393	1	1
HvHSP90 Δ C	61,537 kDa	0,00422	1,075	1,112
HvHSP90ΔC-HvSGT1ΔSGS	89,195 kDa (1:1)	0,00618	1,574	1,449
HvHSP90 Δ C-HvSGT1 Δ SGS	118,370 kDa (1:2)	0,00618	1,574	2,083

¹Wyliczona na podstawie sekwencji dla dimeru białka HvSGT1 Δ SGS i monomeru białka HvHSP90 Δ C oraz stosunków molowych tych białek podanych w nawiasie (HvHSP90 Δ C:HvSGT1 Δ SGS), ²Wartość $I(0)$ obliczona w programie GNOM.

4.12.3 Modelowanie struktury białka HvHSP90 Δ C oraz kompleksu białek HvHSP90 Δ C-HvSGT1 Δ SGS z ADP

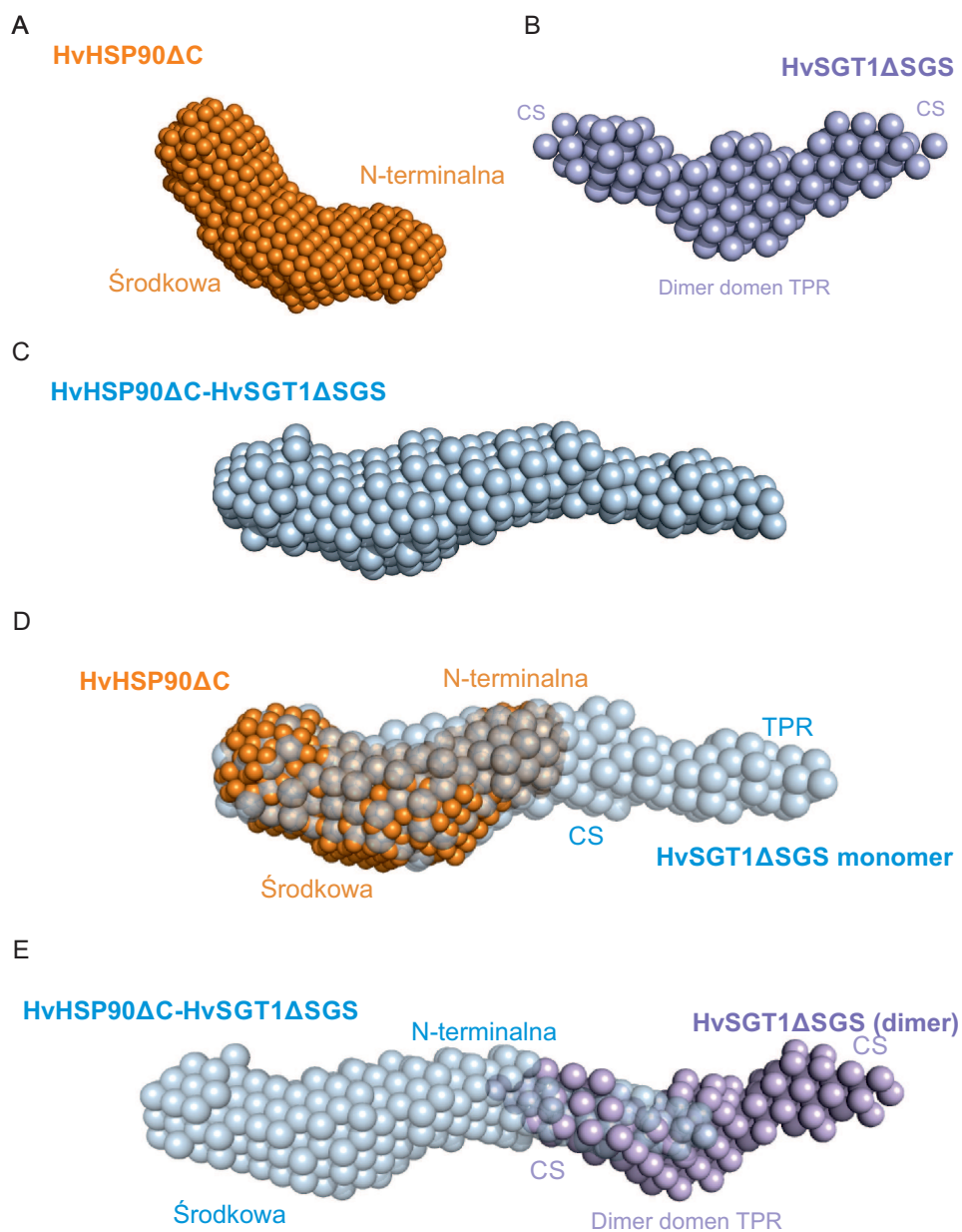
Modelowanie niskorozdzielczej struktury białka HvHSP90 Δ C oraz kompleksu białek HvHSP90 Δ C-HvSGT1 Δ SGS z ADP przeprowadzono w programie do modelowania *ab-initio* DAMMIF. Dodatkowo dla białka HvHSP90 Δ C wymodelowano jego strukturę za pomocą serwera do modelowania homologicznego I-Tasser.



Rysunek 4.58 Modelowanie białka HvHSP90 Δ C w programie DAMMIF i serwerze I-Tasser. **A.** Nałożenie na model z programu DAMMIF (szara powierzchnia) modelu otrzymanego z serwisu

do modelowania homologicznego I-Tasser. **B.** Dopasowanie do danych eksperymentalnych krzywej obliczonej dla modelu otrzymanego z programów DAMMIF i I-Tasser.

Na rysunku 4.58 przedstawione są wyniki modelowania struktury białka HvHSP90 Δ C. Model białka HvHSP90 Δ C ma lekko wydłużony kształt, z lekkim przegięciem w środkowej jego części. Najlepszy model z programu I-Tasser bardzo dobrze pasuje do modelu z programu DAMMIF jak również do danych eksperymentalnych. Jak widać z rysunku 4.58 białko HvHSP90 Δ C jest monomerem. Modelowanie kompleksu białek HvHSP90 Δ C-HvSGT1 Δ SGS również przeprowadzono w programie DAMMIF. Wyniki modelowania przedstawione są na rysunku 4.59 (rysunek C). Model kompleksu jest wydłużony tak jak wynikało to z analizy kształtu krzywej P(r). Na model kompleksu można nałożyć model uzyskany dla białka HvHSP90 Δ C, który bardzo dobrze pasuje do jednego z końców modelu kompleksu. Pozostała część modelu kompleksu ma wymiary zbliżone do połowy modelu dla kompleksu białka HvSGT1 Δ SGS rysunek 4.59 B. Na rysunku 4.59 E pokazane jest dopasowanie modelu białka HvSGT1 Δ SGS do modelu kompleksu. Jak widać z rysunku jedynie połowa modelu białka HvSGT1 Δ SGS mieści się w części nie zajmowanej przez model białka HvHSP90 Δ C. Porównanie modeli otrzymanych w programie DAMMIF potwierdza, że białko HvHSP90 Δ C tworzy kompleks o stechiometrii 1:1 z białkiem HvSGT1 Δ SGS. W kompleksie tym białko HvHSP90 Δ C ma strukturę bardzo podobną do struktury izolowanego białka. Białko HvSGT1 Δ SGS, które w formie nieskompleksowanej z innymi białkami tworzy homodimery, wiąże się z białkiem HvHSP90 Δ C w postaci monomeru. Niestety z danych otrzymanych w eksperymentach SAXS nie można wywnioskować mechanizmu, który by tłumaczył dysocjacje dimeru białka HvSGT1 Δ SGS, szczególnie gdy z modelu wynika, że domena TPR (która jest odpowiedzialna za dimeryzację) nie oddziałuje z białkiem HvHSP90 Δ C.

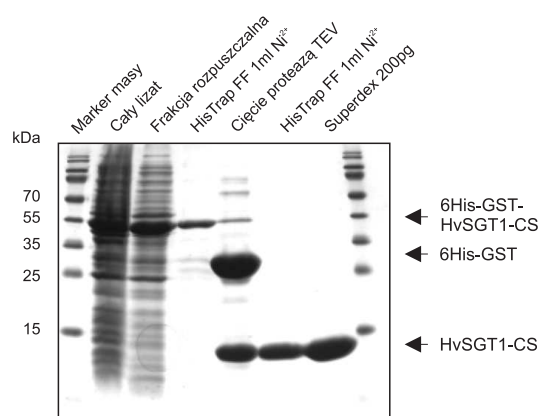


Rysunek 4.59 Modelowanie struktury kompleksu białek HvHSP90ΔC-HvSGT1ΔSGS z ADP. **A.** Model białka HvHSP90ΔC otrzymany w programie DAMMIF. **B.** Model białka HvSGT1ΔSGS otrzymany w programie DAMMIF. **C.** Model kompleksu białek HvHSP90ΔC-HvSGT1ΔSGS z ADP otrzymany w programie DAMMIF. **D.** Porównanie modelu białka HvHSP90ΔC z modelem kompleksu białek HvHSP90ΔC-HvSGT1ΔSGS z ADP. **E.** Porównanie modelu białka HvSGT1ΔSGS z modelem kompleksu białek HvHSP90ΔC-HvSGT1ΔSGS z ADP.

4.13 Badania struktury kompleksu białek HvHSP90 Δ C i HvSGT1-CS

4.13.1 Oczyszczanie białka HvSGT1-CS do pomiarów SAXS

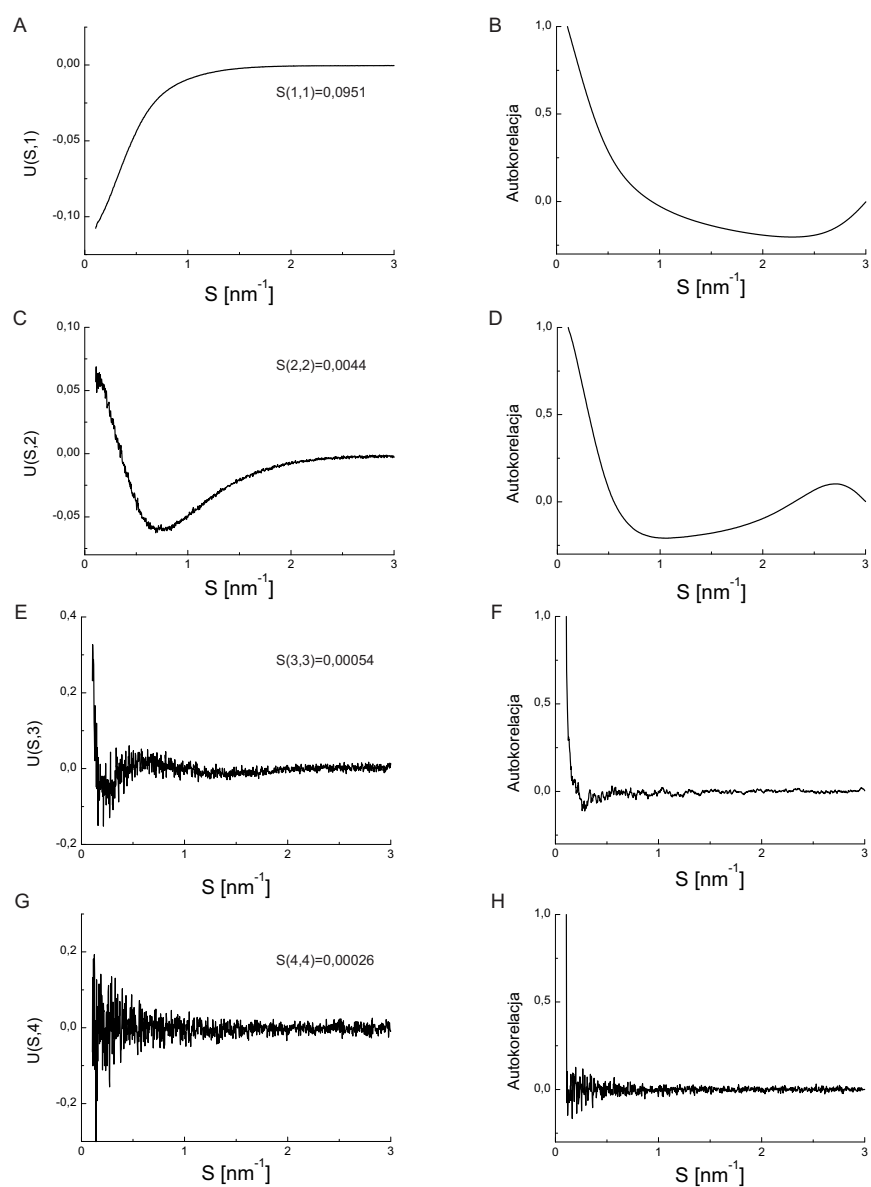
Do badań SAXS przygotowano konstrukt domeny CS białka HvSGT1 zawierający aminokwasy 167-276 z pełnej długości białka HvSGT1. Konstrukt ten jest dłuższy na N- i C-końcu w porównaniu z konstruktem domeny CS użytym do badań dichroizmu kołowego, ze względu na lepszą stabilność w roztworze. Białko HvSGT1-CS oczyszczano z bakterii *E. coli* jako białko fuzyjne z białkiem GST. Białko HvSGT1-CS zostało oczyszczone według protokołów opisanych w sekcji 3.7.1-3.7.2, przy użyciu chromatografii powinowactwa, cięcia proteazą TEV, ponownym oczyszczeniu za pomocą chromatografii powinowactwa oraz sączenia molekularnego. Na rysunku 4.60 pokazana jest czystość białka HvSGT1-CS podczas kolejnych kroków oczyszczania.



Rysunek 4.60 Oczyszczanie białka HvSGT1-CS przedstawione w postaci zdjęcia elektroforezy w żelu polikarylamidowym.

4.13.2 Badania SAXS kompleksu białek HvHSP90 Δ C i HvSGT1-CS

W celu zbadania miejsca wiązania domeny CS białka HvSGT1 (HvSGT1-CS) przeprowadzono badania kompleksowania białka HvHSP90 Δ C z domeną CS białka HvSGT1. Podobnie jak dla pozostałych kompleksów zarejestrowano krzywe rozpraszania dla pojedynczej domeny CS, samego białka HvHSP90 Δ C oraz kompleksu białka HvHSP90 Δ C z białkiem HvSGT1-CS i ADP w stosunkach molarowych 1,5:1; 1,2:1; 1:1; 1:1,2; 1:1,5. Dla zebranych danych wykonano analizę SVD i dla pierwszych czterech wektorów własnych obliczono krzywe autokorelacji. Wyniki tej analizy przedstawione są na rysunku 4.61. Trzy pierwsze wektory własne i ich funkcje autokorelacji są wyraźnie różne od szumu w przeciwieństwie do czwartego wektora własnego. Z analizy SVD wynika istnienie trzech komponentów w zestawie danych SAXS: nieskompleksowanych białek HvHSP90 Δ C i HvSGT1-CS oraz kompleksu HvHSP90 Δ C-HvSGT1-CS.



Rysunek 4.61 Analiza SVD krzywych rozpraszania dla eksperymentu SAXS kompleksowania białek HvHSP90 Δ C i HvSGT1-CS z ADP. **A, C, E, G.** Cztery pierwsze wektory własne (U_1-U_4) uzyskane z zestawu danych SAXS wraz z odpowiadającymi im wartościami własnymi. **B, D, F, H.** Odpowiadające im funkcje autokorelacji.

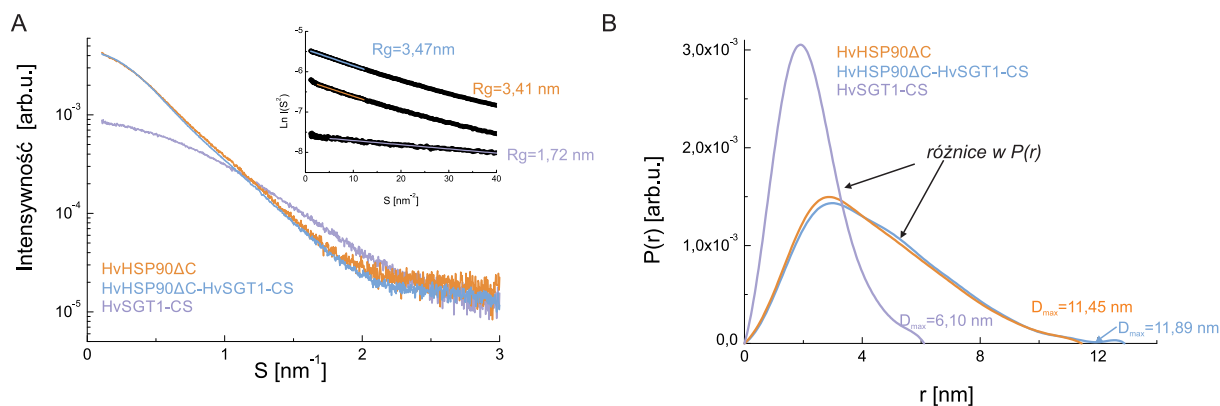
Na podstawie analizy SVD wykonano następnie analizę MCR-ALS i wyznaczono krzywą rozpraszania dla kompleksu białek HvHSP90 Δ C-HvSGT1-CS z ADP. Dla eksperymentalnych krzywych rozpraszania dla białek HvHSP90 Δ C i HvSGT1-CS oraz dla krzywej rozpraszania dla kompleksu z analizy MCR-ALS obliczono krzywe rozkładu odległości wewnątrzcząsteczkowych $P(r)$. Wyznaczono także wartości promienia żyracji R_g oraz maksymalnego rozmiaru cząsteczki D_{max} . Wyniki tych analiz przedstawione są na rysunku 4.62.

W przeciwieństwie do kompleksów białek TaHSP90-HvSGT1 Δ C oraz HvHSP90 Δ C-HvSGT1 Δ SGS na krzywej $P(r)$ utworzenie kompleksu jest słabo widoczne. Kształt krzywej jest bardzo zbliżony do krzywej dla białka HvHSP90 Δ C, jednak widoczne są drobne różnice, które jednak wskazują na utworzenie się kompleksu. Maksimum krzywej $P(r)$ ma mniejszą wartość w przypadku kompleksu HvHSP90 Δ C-HvSGT1-CS. Krzywa $P(r)$ dla kompleksu natomiast przyjmuje większe wartości powyżej 4 nm. Analogiczne zachowanie się krzywej $P(r)$ można było zaobserwować dla kompleksu białek TaHSP90-HvSGT1 Δ SGS. Wartości promienia żyracji R_g oraz maksymalnego rozmiaru cząsteczki D_{max} w bardzo małym stopniu różnią się od wartości dla białka HvHSP90 Δ C aczkolwiek wartości te dla kompleksu są minimalnie większe. Różnice w masie molekularnej białek wyznaczone w programie Scatter, przedstawione w tabeli 4.6 są mniejsze niż teoretyczne masy białek. Pomimo tego że, zmiany parametrów opisujących badane białka są małe, jasno pokazują utworzenie się kompleksu.

Tabela 4.6 Porównanie masy molekularnej oraz wartości parametrów $I(0)$, R_g oraz D_{max} dla białek kompleksu HvHSP90 Δ C- HvSGT1-CS.

Białko	MM kDa ¹	$I(0)$	D_{max} [nm]	R_g [nm]	MM Scatter ²
HvSGT1-CS	12,623	0,000824	6,1	1,72	13,1 kDa
HvHSP90 Δ C	61,537	0,00422	11,46	3,41	58,6 kDa
HvHSP90 Δ C-HvSGT1-CS	74,160	0,00434	11,89	3,47	65,6 kDa

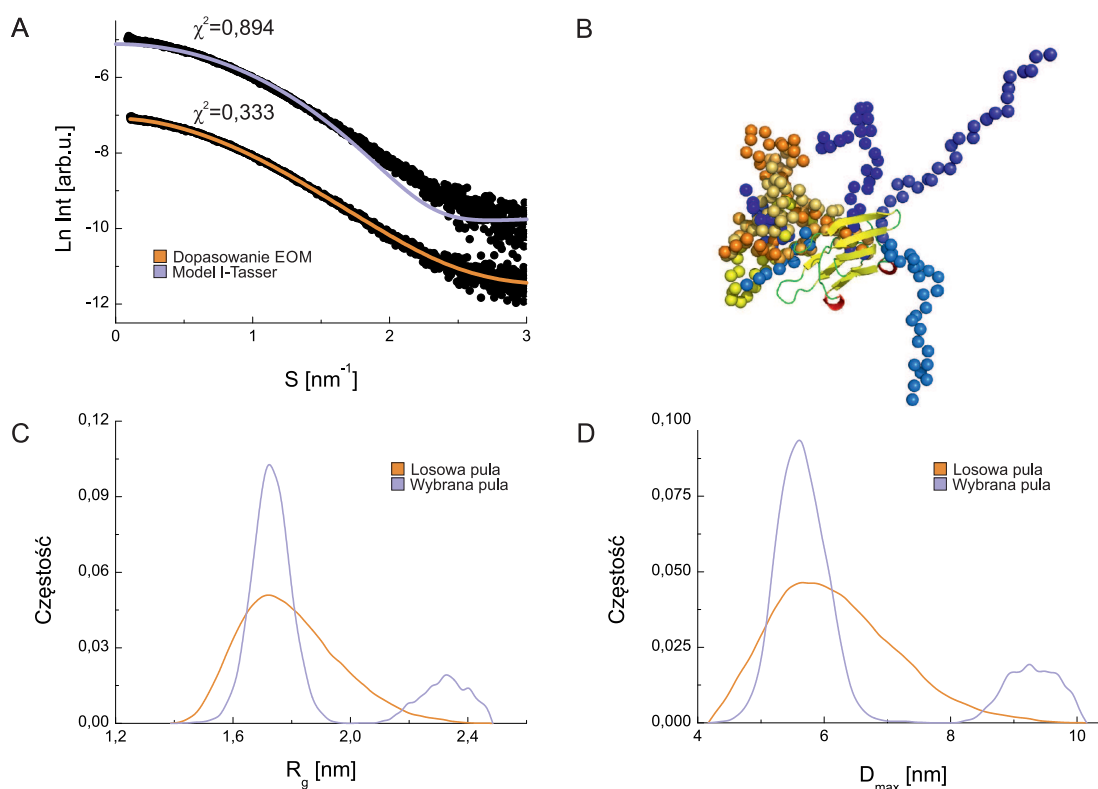
¹Wyliczona na podstawie sekwencji aminokwasowej, ²Wyznaczona w programie Scatter (patrz sekcja 3.9.7)



Rysunek 4.62 A. Krzywe doświadczalne SAXS dla białek HvHSP90 Δ C i HvSGT1-CS oraz kompleksu (otrzymana z procedury MCR-ALS). Na mniejszym wykresie przedstawiony jest wykres Guiniera wraz z wyznaczonymi wartościami promienia żyracji R_g . **B.** Krzywe $P(r)$ dla białek HvHSP90 Δ C i HvSGT1 Δ SGS oraz kompleksu HvHSP90 Δ C-HvSGT1-CS oraz wyznaczone w programie GNOM maksymalne rozmiary cząsteczki D_{max} .

4.13.3 Modelowanie struktury białka HvSGT1-CS

Modelowanie niskorozdzielczej struktury domeny białka SGT1 z jęczmienia przeprowadzono w programie DAMMIF. Uśredniony model otrzymany w programie DAMMIF ma lekko wydłużony kształt na jednym z końców (rysunek 4.64 D). Generalnie model ma kształt podobny do struktury krystalograficznej domeny CS z białka SGT1a z *Arabidopsis thaliana*. Fragment domeny CS wykorzystany w eksperymentach SAXS (aminokwasy 167-276 z pełnej długości białka SGT1 z jęczmienia) posiadał dodatkowe aminokwasy na N-końcu i C-końcu, znajdujące się poza regionem najbardziej zachowanym ewolucyjnie pod względem sekwencji. Struktura krystalograficzna domeny CS białka SGT1a z *Arabidopsis thaliana* zawiera jedynie region najbardziej zakonserwowany. Wydłużenie zaobserwowane w modelu z programu DAMMIF, może być właśnie miejscem gdzie znajdują się dodatkowe aminokwasy na N-końcu lub C-końcu. W celu sprawdzenia tej hipotezy wymodelowano strukturę domeny CS w programie EOM. Najbardziej zakonserwowany region domeny CS, w postaci modelu z programu I-Tasser, potraktowano jako bryłę sztywną natomiast dodatkowe aminokwasy na N-końcu i C-końcu modelowano za pomocą pseudo-łańcucha.



Rysunek 4.63 Modelowanie struktury domeny CS białka SGT1 z jęczmienia w programie EOM i I-Tasser. **A.** Dopasowanie krzywych SAXS obliczonej dla modelu z programu I-Tasser oraz dla puli modeli z programu EOM do eksperymentalnej krzywej SAXS. **B.** Nałożenie na siebie reprezentacyjnych modeli z wybranej puli modeli otrzymanej w programie EOM. Odcieniami koloru niebieskiego zaznaczone są fragmenty modeli o największej wartości D_{\max} i R_g natomiast odcieniami żółtego i

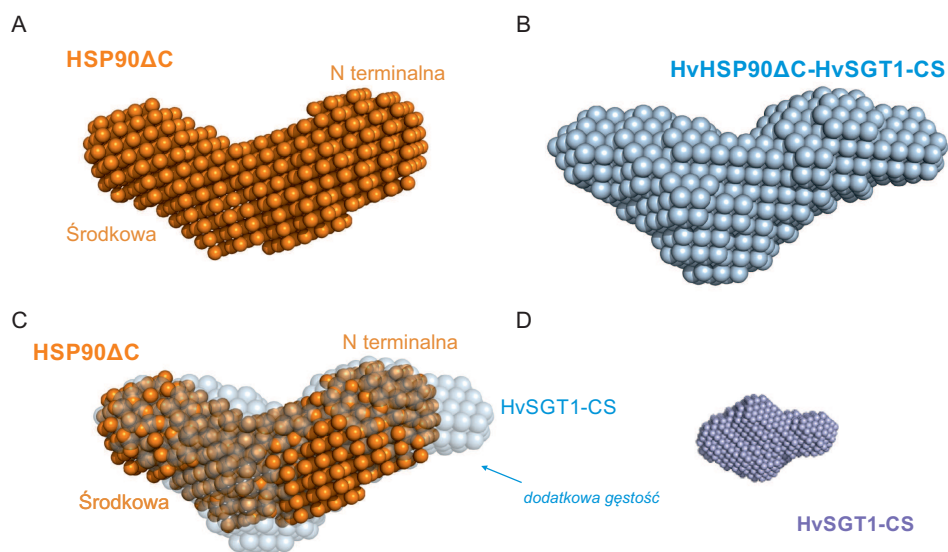
pomarańczowego zaznaczone są fragmenty dla bardziej zwartych modeli. Wspólna dla wszystkich modeli struktura domeny CS wymodelowana w programie I-Tasser zaprezentowana jest w postaci wstążkowej. **C**, **D** - rozkłady promienia żyracji R_g i maksymalnego rozmiaru cząsteczki D_{max} dla losowej puli i puli wybranej w programie EOM.

Wynik modelowania w programie EOM przedstawiony jest na rysunku 4.63. Rozkład promienia żyracji R_g oraz maksymalnego rozmiaru cząsteczki D_{max} dla wybranej puli modeli ma dwa maksima. Pierwsze maksimum jest związane z występowaniem kompaktowej struktury białka HvCS natomiast drugie jest związane z występowaniem konformacji o rozciągniętej konformacji C-końca białka HvCS. Na rysunku 4.63 B pokazane są modele z wybranej puli z programu EOM. Fragmenty modeli należące do konformacji kompaktowej zaznaczone są odcieniami pomarańczowego natomiast należące do konformacji rozciągniętej odcieniami niebieskiego. Dopasowanie krzywej SAXS dla modelu z programu I-Tasser do danych eksperymentalnych jest gorsze niż dopasowanie krzywej SAXS dla puli modeli z programu EOM. Wartości χ^2 wynoszą odpowiednio 0,333 oraz 0,894 dla wybranej puli modeli z programu EOM i modelu z programu I-Tasser.

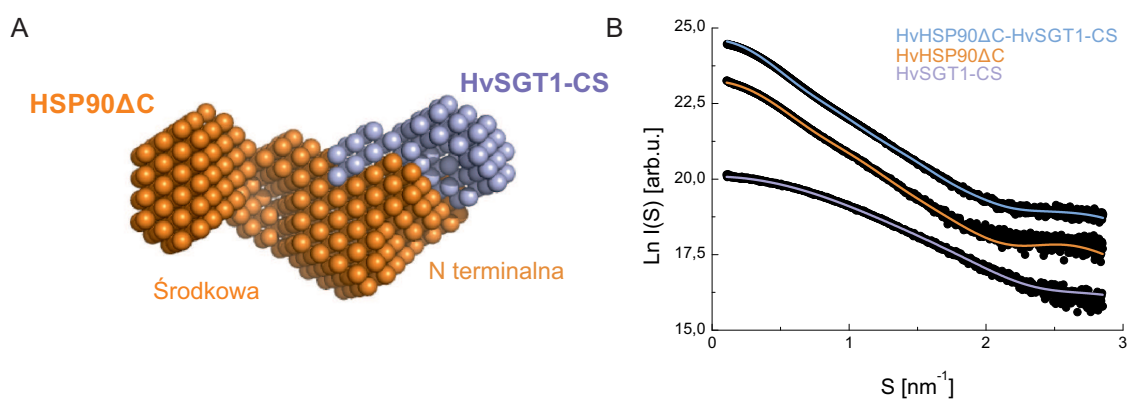
4.13.4 Modelowanie struktury kompleksu białek HvHSP90 Δ C-HvSGT1-CS z ADP

Z kolei miejsca oddziaływania domeny CS białka HvSGT1 z białkiem HvHSP90 Δ C określono za pomocą modelowania niskorozdzielczej struktury kompleksu w programach DAMMIN oraz MONSA. W programie MONSA można modelować w sposób analogiczny jak w programie DAMMIN, niskorozdzielcze struktury kompleksów składające się z N-komponentów (np. kompleks N-różnych białek). Jako wynik otrzymujemy model dla całego kompleksu składający się z N-faz (komponentów kompleksu). Program wykorzystuje krzywe rozpraszania dla podjednostek oraz całego kompleksu, do których dopasowywane są modele poszczególnych faz oraz całego kompleksu. Takie podejście pozwala na otrzymanie kształtu kompleksu, który podzielony jest na podjednostki. Pozwala to odpowiedzieć na pytanie gdzie oddziałują ze sobą podjednostki kompleksu i jaka jest ich struktura.

Modelowanie kompleksu najpierw przeprowadzono w programie DAMMIN. Wyniki modelowania zaprezentowane są na rysunku 4.64. Kształt kompleksu jest bardzo podobny do modelu białka HvHSP90 Δ C z charakterystycznym przegięciem w strukturze. Nałożenie modelu białka HvHSP90 Δ C na model białka HvHSP90 Δ C-HvSGT1-CS uwidacznia dodatkowy fragment modelu, który można zinterpretować jako białko HvSGT1-CS. Lokalizacja tego fragmentu znajdują się przy domenie N-terminalnej białka HvHSP90 Δ C, czyli w miejscu oddziaływania z domeną HvSGT1-CS. Modelowanie w programie MONSA potwierdziło, że domena CS białka HvSGT1 oddziałuje z N-terminalną domeną białka HvHSP90 Δ C. Sam kompleks ma te same charakterystyczne przegięcie (rysunek 4.65 A). Krzywe rozpraszania obliczone dla modeli białek HvHSP90 Δ C HvSGT1-CS oraz kompleksu bardzo dobrze pasują do krzywych eksperymentalnych.



Rysunek 4.64 Modelowanie struktury kompleksu HvHSP90 Δ C-HvSGT1-CS z ADP w programie DAMMIN. **A.** Model białka HvHSP90 Δ C otrzymany w programie DAMMIN. **B.** Model kompleksu białek HvHSP90 Δ C-HvSGT1-CS z ADP otrzymany w programie DAMMIN. **C.** Nałożenie modelu dla białka HvHSP90 Δ C na model kompleksu HvHSP90 Δ C-HvSGT1-CS z ADP. **D.** Model białka HvSGT1-CS otrzymany w programie DAMMIF.



Rysunek 4.65 Modelowanie kompleksu HvHSP90 Δ C-HvSGT1-CS z ADP w programie MONSA. **A.** Model otrzymany w programie MONSA. Kolorem pomarańczowym zaznaczony jest model białka HvHSP90 Δ C, fioletowym model białka HvSGT1-CS. **B.** Dopasowanie krzywych rozpraszania dla modeli z programu MONSA do danych eksperymentalnych.

5 Dyskusja

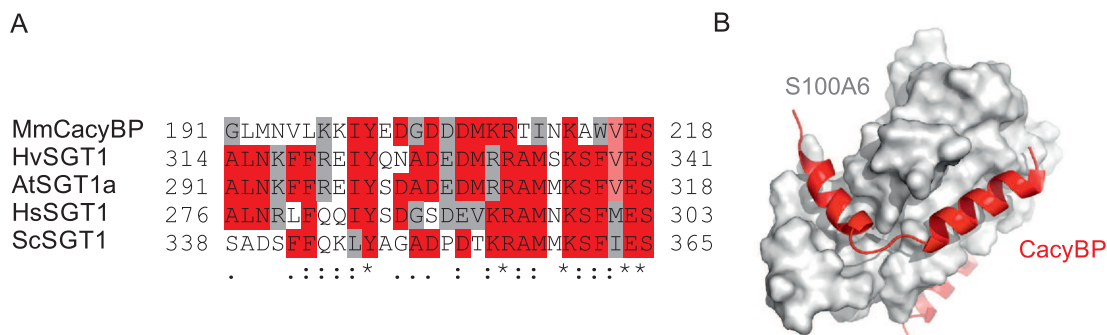
Głównym celem pracy było zbadanie struktury trzech białek zaangażowanych w procesy odpowiedzi obronnej na patogeny u roślin: HSP90, SGT1 i RAR1. Białko HSP90 jest jednym z kluczowych białek, które odpowiedzialne są za procesy fałdowania się innych białek w warunkach stresu, jak również za kontrolę wielu ważnych procesów komórkowych poprzez stabilizację kluczowych dla nich białek. Charakterystyczną cechą białka HSP90 jest jego występowanie w kompleksie z wieloma partnerami białkowymi zwanymi z ko-chaperonami (ang. co-chaperones). Białka te można podzielić na dwie grupy. Pierwsza moduluje jego funkcję poprzez inhibicję lub stymulację aktywności ATPazowej. Druga natomiast jest odpowiedzialna za rekrutację specyficznej klasy “klientów” do kompleksu HSP90. Do tej klasy należą białka SGT1 i RAR1. Białka SGT1 i RAR1 (jak i ich homologii z innych organizmów) biorą udział w przekazywaniu sygnału podczas reakcji na atak patogenów u roślin i zwierząt. Uważa się że, białka te u roślin odpowiedzialne są za rekrutację cytoplazmatycznych receptorów odporności R, które rozpoznają patogenne białka wewnątrz komórki. Jak dotąd niewiele jest dostępnych informacji na temat struktury pełnej długości białek SGT1, RAR1 oraz ich kompleksów z białkiem HSP90 u roślin.

W ramach niniejszej pracy udało się po raz pierwszy wymodelować niskorozdzielczą strukturę pełnej długości białek SGT1 oraz RAR1 z jęczmienia. Zbadano również niskorozdzielczą strukturę kompleksu białek TaHSP90 z białkiem HvSGT1 Δ SGS oraz białka HvHSP90 Δ C z białkami HvSGT1 Δ SGS i HvSGT1-CS. Dodatkowo przeanalizowano dimeryzację domeny TPR białka SGT1 z jęczmienia, *Arabidopsis thaliana*, drożdży piekarskich i człowieka w różnych warunkach siły jonowej, a także zaproponowano modele strukturalne dla dimeru domen TPR. Ponadto zbadano również wpływ siły jonowej i temperatury na równowagę konformacyjną białka HSP90 z pszenicy oraz ustalono niskorozdzielczą strukturę dla wyznaczonych stanów konformacyjnych.

5.1 Badanie struktury pełnej długości białka SGT1 z jęczmienia

Pierwszym krokiem eksperymentalnym, przed przystąpieniem do badań strukturalnych, było otrzymanie w ilości kilku miligramów białka SGT1 o odpowiedniej czystości. Z wielu doniesień literaturowych wynikało, że białko to, niezależnie od organizmu, z którego pochodzi, w komórkach bakteryjnych i owadziach ulega częściowej proteolizie (Nyarko i wsp., 2007). Ponadto, podczas oczyszczania oprócz pełnej długości, izolują się także krótsze fragmenty tego białka. Stąd też opracowano odpowiedni protokół pozwalający na oczyszczanie pełnej długości białka SGT1 spełniającego kryteria późniejszego zastosowania do badań strukturalnych. O czystości i poprawnym uformowaniu świadczą pomiary widma dichroizmu kołowego. Warto nadmienić iż tak przygotowane białko HvSGT1 odznaczało się większą czystością niż białka przygotowane według innych, opublikowanych wcześniej protokołów. Białko SGT1 w swojej strukturze zawiera fragmenty helisy α , arkuszy β oraz regiony nieuporządkowane. Szczególnie duży jest udział struktury nieuporządkowanej. Znaczący wkład w strukturę nieuporządkowaną mają fragmenty VR1 i VR2 oraz domena SGS. Wynika to z analizy, zarejestrowanych przeze mnie, widm dichroizmu kołowego dla krótszych fragmentów białka SGT1. Z eksperymentów NMR wynikało, że domena SGS jest nieuporządkowana (Botër i wsp., 2007). Co ciekawe ludzkie białko CacyBP, które wykazuje wysokie podobieństwo sekwencyjne do fragmentu do-

meny SGS (przyrównanie sekwencji fragmentów na rysunku 5.1 A), po związaniu się białkiem S100A uległo porządkowaniu i przyjmowało strukturę α helikalną (rysunek 5.1 B) (Lee i wsp., 2008).

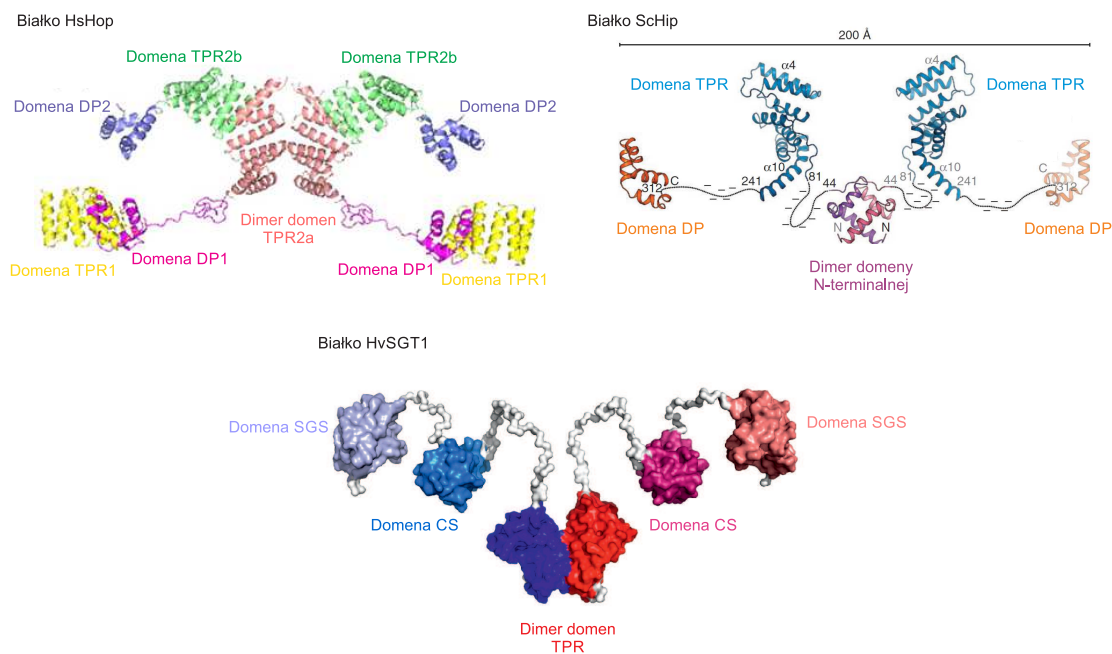


Rysunek 5.1 A. Przyrównanie sekwencji fragmentu domeny SGS białka CacyBP z fragmentami sekwencji domeny SGS białek SGT1 z różnych organizmów (Mm-*Mus musculus* (mysz domowa), Hv - *Hordeum vulgare*, At- *Arabidopsis thaliana*, Hs - *Homo sapiens*, Sc - *Saccharomyces cerevisiae*). Kolorem czerwonym oznaczone są najbardziej zakonserwowane aminokwasy. **B.** Struktura NMR fragmentu białka CacyBP (kolor czerwony, kod pdb: 2JTT, Lee i wsp., 2008) w kompleksie z białkiem S100A6 (kolor szary).

Komputerowe modele domeny SGS z jęczmiennego białka SGT1, otrzymane przez mnie w programach I-Tasser i Quark, charakteryzują się strukturą α helikalną. Z pomiarów dichroizmu kołowego wynikało, że w obrębie domeny SGS znajdują się fragmenty α helikalne aczkolwiek ich wkład procentowy jest bardzo mały i nie przekracza 5 %. Za pomocą dichroizmu kołowego wykazano, że domena SGS może się porządkować w zatłoczonym środowisku molekularnym. Widmo dichroizmu kołowego uzyskane dla fragmentu białka SGT1 zawierającego domeny CS, VR2 i SGS w roztworze zawierającym 30 % Ficoll 70K wykazało, że fragment ten ulega częściowemu porządkowaniu się.

Trzy domeny białka SGT1 wymodelowano za pomocą programu Quark (Xu i wsp., 2012). Uzyskany dla domeny TPR model ma strukturę typową dla domen TPR (np. domeny TPR białka HOP, ko-chaperona białka HSP90). Natomiast model domeny CS ma strukturę bardzo podobną do struktury domeny CS z ludzkiego białka SGT1 otrzymanej za pomocą NMR i struktury krystalograficznej domeny CS z białka SGT1a z *Arabidopsis thaliana* (Lee i wsp., 2004, Botër i wsp., 2007). W celu otrzymania niskorozdzielczej struktury białka SGT1 w roztworze posłużono się techniką mało-kątowego rozpraszania promieniowania rentgenowskiego SAXS. Ponieważ białko SGT1 występuje w roztworze w równowadze monomer-dimer, wyniki standardowej metody analizy danych SAXS byłyby zafałszowane poprzez istnienie dwóch stanów oligomerycznych jęczmiennego białka SGT1. Blobel i współpracownicy (2009) zaproponowali nową metodę, MCR-ALS, pozwalającą na wyodrębnienie z zestawu danych, które są kombinacją liniową krzywych SAXS, dane dla indywidualnych komponentów bez znajomości informacji na temat ich struktury czy stałej dysocjacji. Za pomocą analizy SVD i MCR-ALS otrzymano krzywe rozpraszania dla monomeru i dimeru białka SGT1. Z wyznaczonych funkcji $P(r)$ oraz parametrów: D_{max} i R_g wynika, że białko SGT1 w roztworze ma rozciągniętą

konformację zarówno w formie monomerycznej jak i dimerycznej. Za pomocą metod *ab initio* otrzymano niskorozdzielcze modele dla dimeru i monomeru białka SGT1. Modele te potwierdzają wnioski płynące z analizy krzywych $P(r)$, że białko SGT1 w roztworze ma rozciągnięty i wydłużony kształt, a regiony zmienne mają nieuporządkowaną strukturę. Modelowanie bryły sztywnej pokazało, że jęczmienne białko SGT1 ma dynamiczny charakter i w roztworze występuje w kilku konformacjach. Dimeryzacja nie wpływa na ogólną dynamikę białka, a sam dimer także posiada nieuporządkowane fragmenty VR1 i VR2. Dimeryzacja zachodzi jedynie poprzez oddziaływanie N-terminalnych domen TPR, bez udziału innych fragmentów białka. Niestety modelowanie za pomocą danych SAXS nie pozwoliło na pełne określenie stopnia nieuporządkowania domeny SGS.



Rysunek 5.2 Porównanie struktury dimeru białka HvSGT1 z dimerem ludzkiego białka HOP i Hip z drożdży piekarskich (Onuoha i wsp., 2008, Li i wsp., 2013b).

Na rysunku 5.2 zaprezentowano porównanie struktury dimeru białka SGT1 z jęczmienia z strukturą dimerów białek HOP i HIP (opis w tekście poniżej). Podobną, dynamiczną strukturę posiada ludzkie białko HOP (Onuoha i wsp., 2008), które także jest ko-chaperonem białka HSP90. Onuoha i współpracownicy wykorzystali technikę SAXS, by uzyskać niskorozdzielczy model pełnej długości białka HOP jak również model białka HOP w oparciu o modelowanie bryły sztywnej. W modelu tym, centralną część stanowi dimer domen TPR (domena TPR2A białka HOP) tak samo jak w przypadku białka SGT1, a same domeny TPR2A oddziałują boczną stroną, na którą składają się pętle pomiędzy helisami α w powtórzeniu TPR. Podobną dynamiczną strukturę posiada białko Hip z drożdży (Li i wsp., 2013b). Jest ono ko-chaperonem białka HSP70, które także asystuje w procesie fałdowania się białek. Białko Hip występuje w postaci dimeru, lecz dimeryzacja zachodzi poprzez N-terminalną domenę, a nie przez środkową domenę TPR. Wszystkie te białka są dimerami i mają

dynamiczną strukturę, która ma odzwierciedlenie w funkcji tych białek tj. pośredniczeniu w wiązaniu wielu partnerów białkowych w kompleksie biorącym udział w fałdowaniu i/lub aktywacji innych białek.

W modelu dimeru pełnej długości jęczmiennego białka SGT1, centralną część stanowi dimer domen TPR. Jak wykazano we wcześniejszych pracach (Nyarko i wsp., 2007, Bansal i wsp., 2009a) domena TPR jest niezbędna do dimeryzacji, a sama dimeryzacja domeny TPR jest zależna od siły jonowej i potencjału oksydoredukcyjnego roztworu. W celu podjęcia próby określenia struktury dimeru domeny TPR białka SGT1 przeprowadzono eksperymenty SAXS izolowanej domeny TPR z jęczmiennego białka SGT1. Domena HvTPR, tak jak pełnej długości białko, w roztworze występuje w równowadze monomer - dimer. W roztworach o niskiej sile jonowej domena TPR występuje głównie jako dimer, a wraz ze wzrostem siły jonowej równowaga przesuwa się w stronę monomeru. W roztworze dimer domeny TPR posiada kształt spłaszczonego dysku jak wynika z modelowania niskorozdzielczej struktury na podstawie danych SAXS. Model dimeru domeny TPR posiada lekkie przewężenie, które na podstawie modelowania metodą bryły sztywnej jest potencjalnym miejscem oddziaływania obu podjednostek TPR. Miejsce oddziaływania znajduje się w C-terminalnej części, którą w eksperymentach zszywania chemicznego zidentyfikowano jako potencjalne miejsce oddziaływania (Nyarko i wsp., 2007). Co ciekawe w tej części białka znajdują się też para cystein, które stanowią swoisty przełącznik regulujący dimeryzację domeny TPR. Tak więc otrzymany model zgadza się z dotychczasowymi danymi otrzymanymi przez innych badaczy.

5.2 Badanie niskorozdzielczej struktury białka RAR1 z jęczmienia

Białko RAR1 jak wynika z przedstawionych wcześniej wyników analizy strukturalnej jest częściowo nieuporządkowane w roztworze, a wiązanie cynku pełni rolę strukturalną (Heise i wsp., 2007). Z otrzymanych przeze mnie wyników wynika, że pełnej długości białko RAR1 z jęczmienia ma dynamiczną strukturę w roztworze, gdzie domeny CHORD I i CHORD II mają uporządkowaną strukturę a fragment pomiędzy domenami jest nieuporządkowany. Heise i współpracownicy (Heise i wsp., 2007) pokazali również, że motyw CCCH wiąże jeden atom cynku i najprawdopodobniej motyw ten musi być w pewnym stopniu uporządkowany. Z otrzymanych wyników dla modelowania białka RAR1 za pomocą programu EOM wynika, że fragment ten nie jest uporządkowany, gdyż użycie do modelowania struktury tego motywu otrzymanej z programu I-Tasser pogarsza jakość dopasowania (rysunek 4.24) do eksperymentalnych danych SAXS. Białko RAR1 podobnie jak białko SGT1 ma modułową budowę, a fragment pomiędzy domenami odpowiedzialnymi za oddziaływania z innymi białkami ma nieuporządkowany charakter. Wyniki te są zgodne z wynikami otrzymanymi przez Heise i współpracowników (Heise i wsp., 2007), którzy pokazali za pomocą spektroskopii dichroizmu kołowego, że białko RAR1 w swojej strukturze posiada głównie fragmenty nieuporządkowane. Także, oba białka oddziałujące z białkiem HSP90: RAR1 i SGT1 są białkami częściowo nieuporządkowanymi i mają podobny charakter. Być może, dynamika konformacyjna jest odpowiedzialna za poprawne formowanie się i funkcję kompleksu białek HSP90-SGT1-RAR1.

5.3 Zastosowanie metody SVD i MCR-ALS w analizie danych SAXS

Metody analizy zestawu danych za pomocą analizy SVD (ang. singular value decomposition) i analizy MCR-ALS (ang. multivariate curve resolution-alternating last squares) są bardzo popularnymi metodami w analizie danych chemometrycznych np. analizie chromatogramów mieszanin czy też różnego rodzaju danych spektroskopowych (de Juan i wsp., 2014). Za pomocą tych metody można analizować złożone zestawy danych, które są kombinacją liniową np. widm spektroskopowych dla poszczególnych składowych mieszaniny. Przykładem zastosowania tych metod może być analiza wpływu stężenia jonów magnezu oraz temperatury i siły jonowej na strukturę i stabilności oligonukleotydu DNA mierzona za pomocą spektroskopii UV i spektroskopii dichroizmu kołowego w zakresie UV (Jaumot i wsp., 2002). Dzięki zastosowaniu analizy MCR-ALS autorzy tej pracy wyodrębnili widma dla trzech różnych struktur i określili jak stężenie jonów Mg^{2+} , siła jonowa i temperatura zmieniają równowagę pomiędzy trzema strukturami DNA. Metoda ta znalazła także zastosowanie do analizy danych SAXS. Blobel i współpracownicy (2009) po raz pierwszy wykorzystali metodę MCR-ALS do wyizolowania krzywych rozpraszania, a następnie modelowania niskorozdzielczej struktury białka lmwPTP (ang. low molecular weight protein tyrosine phosphatase) oraz jego różnych form oligomerycznych. Najważniejszą zaletą metody MCR-ALS jest możliwość znalezienia krzywych SAXS dla składników mieszaniny bez żadnej wiedzy na temat ich struktury, czy też przy braku modelu termodynamicznego opisującego równowagę pomiędzy składnikami mieszaniny. Jedyne założenia jakie należy poczynić w analizie to to, że krzywa SAXS ma dodatnie wartości (zliczenia fotonów nie mogą być ujemne) oraz unimodalność (dla $s \rightarrow 0$ krzywa SAXS przyjmuje maksymalne wartości). Pomimo zalet jakie posiada metoda MCR-ALS w literaturze nie ma dużej liczby doniesień o wykorzystaniu tej metody do analizy danych SAXS.

W mojej pracy metoda SVD i MCR-ALS stanowiła główne narzędzie do analizy danych otrzymanych z eksperymentów SAXS. Dzięki tej metodzie udało mi się otrzymać krzywe SAXS dla monomeru i dimeru białka SGT1 z jęczmienia a następnie określić jego strukturę. Bez zastosowania metody MCR-ALS nie byłoby możliwe oddzielenie wkładu monomeru i dimeru do krzywych rozpraszania. Nawet dla krzywych SAXS zarejestrowanych w skrajnych stężeniach NaCl, gdzie dominuje wkład monomeru lub dimeru krzywe rozpraszania nie są czystymi krzywymi dla pojedynczych komponentów i modelowanie struktury na ich podstawie obarczone byłoby błędem. Analiza krzywych $P(r)$ również byłaby obarczona pewnym błędem a analiza subtelných różnic pomiędzy krzywymi $P(r)$ mogłaby prowadzić do błędnych wniosków. Szczególnie widoczne jest to na przykładzie analizy krzywych $P(r)$ dla trzech konformacji białka HSP90 z pszenicy. Ogólny kształt trzech konformacji jest bardzo podobny, wszystkie mają wydłużony kształt ale różnice w kształcie maksimum na krzywej $P(r)$ uwiadcniają różnice pomiędzy tymi konformacjami.

Krukenberg i współpracownicy (2008) zbadali równowagę konformacyjną białka HtpG homologa białka HSP90 z bakterii *E.coli*. Białko HtpG występowało w dwóch konformacjach otwartej lub zamkniętej, czyli bardzo podobnie jak białko HSP90 z pszenicy. Jednak w tej pracy zastosowano inne podejście do analizy danych SAXS. Za pomocą modelowania komputerowego autorzy wygenerowali różne struktury białka HtpG w oparciu o modelowanie bryły sztywnej i obliczyli dla każdego modelu wartość dopasowania do danych eksperymentalnych. Taka analiza jest bardziej czasochłonna i wymaga większej mocy obliczeniowej niż zastosowanie metody MCR-ALS. Analiza SVD pozwala na

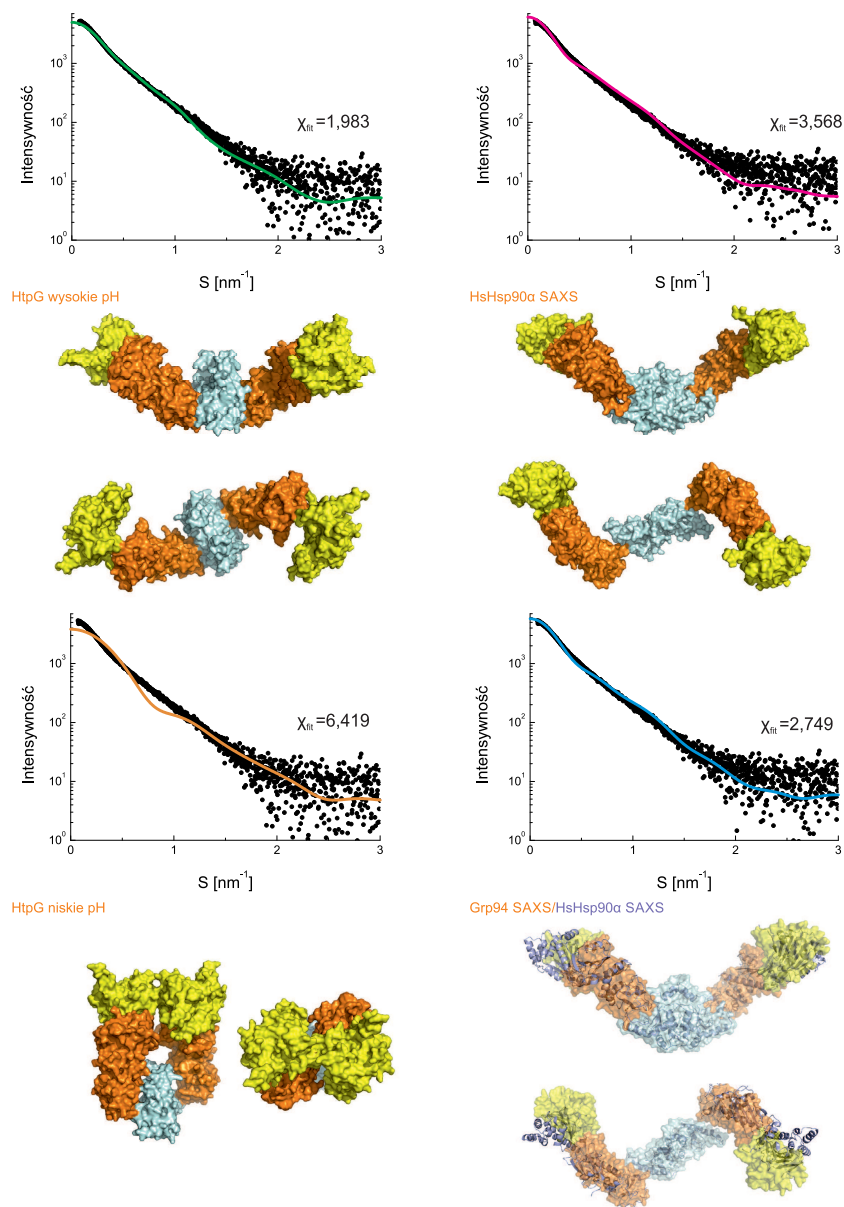
oszacowanie liczby istotnych stanów konformacyjnych a otrzymane krzywe SAXS z analizy MCR-ALS służą do modelowania w oparciu o metody Monte Carlo, co w rezultacie jest dużo bardziej wydajne. Białko HSP90 z pszenicy występuje w równowadze trzech różnych konformacji, dla których krzywe SAXS są bardzo podobne. Modelowanie za pomocą podejścia opisanego w pracy Krukenberg i współpracownicy (2008), byłoby bardzo niewydajne i wymagało wykorzystania dużej mocy obliczeniowej ponieważ jednocześnie trzeba by było znaleźć strukturę oraz udział w mieszaninie dla trzech konformerów.

Szczegółowa analiza ekstrapolowanej wartości intensywności dla zerowego kąta rozpraszania ($s = 0$) była kluczowa dla analizy stechiometrii kompleksów białka HSP90 z białkiem SGT1. Krzywe SAXS dla kompleksu białek zawsze zawierały pewien wkład od nieskompleksowanych białek. Wynika to z błędów pomiaru stężenia białek, jak również z niezbyt dużej wartości stałej dysocjacji kompleksu. Dlatego też wartość $I(0)$ wyznaczona bezpośrednio z danych eksperymentalnych nie dawała stuprocentowej pewności do otrzymanych wyników. Zastosowanie analizy MCR-ALS pozwoliło na otrzymanie wartości $I(0)$ dla krzywych SAXS dla kompleksów bez wkładu od niezwiązanych białek, a błąd wyniósł nie więcej niż 10 %, co jest wynikiem bardzo dobrym i wystarczająco potwierdza przyjęte założenia co do stechiometrii kompleksów białek HSP90 i SGT1 z ADP. Zastosowany w pracy protokół pomiarów i analizy danych SAXS dla kompleksu białek może być śmiało zastosowany do badania struktury innych kompleksów białkowych i kompleksów z kwasami nukleinowymi. Pomiarzy dla różnych stosunków molowych komponentów kompleksu pozwoliły by także na uzyskanie informacji na temat procesu formowania się kompleksu i istnienia stanów przejściowych.

5.4 Wpływ siły jonowej i temperatury na dynamikę konformacyjną białka HSP90 z pszenicy

Białko HSP90 jest jednym z najbardziej intensywnie badanych białek ze względu na bardzo ważną rolę w komórce - udział w fałdowaniu białek oraz w procesach nowotworzenia (Taipale i wsp., 2010). Badania struktury białka HSP90 wykazały dramatyczne zmiany konformacyjne towarzyszące wiązaniu różnych nukleotydów: ATP i ADP przez białka z rodziny HSP90 (Krukenberg i wsp., 2011). Ze względu na wielkość oraz występowanie fragmentów niepołałdowanych w strukturze, badania krystalograficzne białka HSP90 były bardzo trudne, a strukturę eukariotycznych białek HSP90 otrzymano dopiero po usunięciu łącznika pomiędzy N-terminalną i środkową domeną (Ali i wsp., 2006, Dollins i wsp., 2007). Z drugiej strony wyżej wymienione cechy białka HSP90 nie są przeszkodą w badaniach SAXS. Za pomocą tej metody odkryto, że białko HtpG z *E.coli* w roztworze ma bardziej otwartą konformację niż ta występująca w strukturze krystalograficznej (Krukenberg i wsp., 2008). W tej samej pracy autorzy wykazali, że białko HSP90 po związaniu analogu ATP występuje w równowadze dwóch konformacji: zamkniętej (analogicznej do struktury zaobserwowanej dla drożdżowego białka HSP90 w kompleksie z AMPPNP) oraz otwartej (takiej samej jak dla białka bez związanego nukleotydu). Także, metoda SAXS stała się standardem do badania struktury i konformacji białka HSP90. Za jej pomocą pokazano również, że HSP90 pod wpływem osmolitów takich jak tlenek N-trimetyloaminy (TMAO) czy betaina, występuje częściej w konformacji zamkniętej, podobnej do tej zaobserwowanej w strukturze białka Grp94 (Krukenberg i wsp., 2009b, Street i wsp., 2010). Krukenberg i współpracownicy (2009a) pokazali za jej pomocą, że białko HtpG z *E.coli* zmienia swoją konformację pod

wpływem pH. W niskich pH występuje w równowadze dwóch konformacji, głównie w konformacji zamkniętej. Natomiast w roztworach o wyższym pH występuje głównie w postaci konformacji otwartej. Tak samo jak białka HSP90 z innych organizmów, roślinne białko HSP90 (pszenica) występuje w roztworze w postaci równowagi kilku konformacji.



Rysunek 5.3 Porównanie modeli otrzymanych przez grupę Agarda na podstawie danych SAXS do krzywej SAXS dla konformacji pośredniej białka HSP90 z pszenicy.

W odróżnieniu od ludzkiego i drożdżowego białka HSP90, w roztworze białko z pszenicy występuje głównie w postaci konformacji bardziej zamkniętej przy wyższych stężeniach białka (około 9 mg/ml)

w porównywalnych warunkach siły jonowej. Dla niższego stężenia (około 4 mg/ml) białko HSP90 pszenicy występuje w równowadze konformacji otwartej i bardziej zamkniętej. Z badań SAXS wynika, że stężenie, a co za tym idzie również zatłoczenie molekularne, może mieć wpływ na równowagę konformacyjną białka HSP90 z pszenicy. W literaturze brak jest doniesień na temat wpływu stężenia lub zatłoczenia molekularnego na równowagę konformacyjną białka HSP90 z innych organizmów. Ponieważ konformacja białka HSP90 zależy od czynników takich jak pH czy osmolity, zbadano jak zachowuje się ono pod wpływem temperatury i siły jonowej. Zarówno wyższa temperatura i siła jonowa powodują przesunięcie równowagi w stronę konformacji bardziej otwartej. Podobnie wyższe pH powodowało przesunięcie równowagi białka HtpG w stronę konformacji otwartej. Zatem trzy czynniki: temperatura, siła jonowa i wysokie pH mają podobny wpływ na białka z rodziny HSP90. Przyjęcie bardziej otwartej konformacji sprzyjałoby zwiększeniu dostępności do środkowej części białka HSP90, która jak się uważa odpowiedzialna jest za wiązanie substratu. Wiązanie nukleotydów przez białko HSP90 z pszenicy nie zmienia diametralnie równowagi konformacyjnej. Dla niskiej siły jonowej (0 mM KCl) przeważa konformacja zamknięta, niezależnie od związanego nukleotydu. Jednak widoczny jest wzrost udziału konformacji pośredniej białka HSP90 po związaniu ADP i AMPPNP. Także, konformacja ta może być jedną z pośrednich konformacji podczas cyklu związanego z wiązaniem ATP i substratu, a następnie hydrolizy ATP do ADP i uwolnieniem poprawnie zwiniętego białka. Wraz ze wzrostem siły jonowej udział tej konformacji HSP90 spada. Natomiast trudno jest orzec jaka może być funkcja konformacji bardziej zamkniętej. Odnosząc się do wspomnianej powyżej sugestii, jakoby bardziej otwarta konformacja sprzyjałaby związaniu substratu poprzez zwiększenie dostępności do hydrofobowej części białka HSP90, bardziej zamknięta konformacja sprzyjałaby silniejszemu wiązaniu substratu i zbliżała do siebie domeny N-terminalne odpowiedzialne za wiązanie ATP. Wiązanie substratu przez białko HtpG powoduje przesunięcie równowagi konformacyjnej w stronę konformacji zamkniętej, przypominającej kształtem literę V (Street i wsp., 2011). Czynniki takie jak pH, temperatura i siła jonowa zmieniają delikatną równowagę między różnymi stanami konformacyjnymi, pomiędzy którymi bariera energetyczna jest niewielka. Korzystne energetycznie wiązanie nukleotydów lub substratu powodowałoby zmiany konformacyjne, które prowadziłyby w rezultacie do hydrolizy ATP i fałdowania się substratu. Warunki stresowe mogłyby obniżać barierę energetyczną pomiędzy kolejnymi stanami konformacyjnymi, a w rezultacie zapewnić jak najbardziej wydajne fałdowanie się białek substratów.

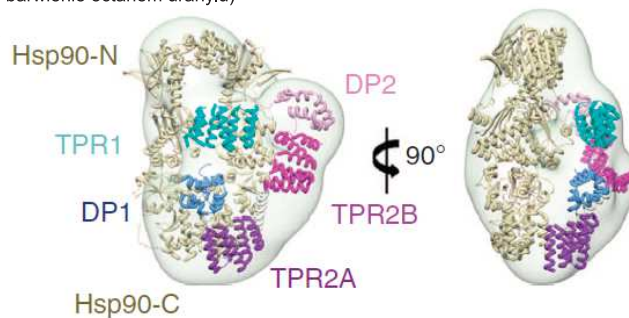
Z porównania modeli otrzymanych na podstawie danych SAXS dla białka HSP90 z różnych organizmów przez grupę prof. Agarda (<http://www.msg.ucsf.edu/agard/pdb/hsp90.html>) do krzywych eksperymentalnych dla trzech konformacji białka HSP90 z pszenicy wynika, że jedynie dla konformacji pośredniej model dla białka HtpG otrzymany w wysokim pH pasuje do danych eksperymentalnych (rysunek 5.3). Samo dopasowanie charakteryzuje się wartością wskaźnika rozbieżności χ równą 1,949. Modele dla białka HSP90 α (Krukenber i wsp., 2009b), Grp94 (Krukenber i wsp., 2009b) i HtpG (Krukenber i wsp., 2009a) (konformacja w niskim pH) mają znacznie gorsze dopasowanie do danych eksperymentalnych jak widać na rysunku 5.3. Porównując kształt krzywej $P(r)$ dla konformacji pośredniej i krzywej $P(r)$ dla białka HtpG w wysokim pH (Krukenber i wsp., 2009a) można zauważyć spore podobieństwo, a w szczególności kształt maksimum, które jest mniej symetryczne niż w przypadku krzywych $P(r)$ dla dwóch pozostałych konformacji białka HSP90 z pszenicy. Dla dwóch

pozostałych konformacji białka HSP90 z pszenicy dopasowanie modeli SAXS (HSP90 α , Grp94, HtpG wysokie i niskie pH) daje wysokie wartości χ . Tak, więc dwie główne konformacje białka HSP90 z pszenicy są inne niż te dla białek HSP90 z innych organizmów. Ponieważ bakterie i drożdże są dużo mniej złożonymi organizmami niż rośliny i ludzie, rola białek HSP90 może się różnić pomiędzy tymi grupami organizmów. W bakteriach i drożdżach białko HSP90 byłoby głównie odpowiedzialne za fałdowanie się białek w warunkach stresowych. Natomiast u roślin i ludzi, gdzie białka są bardziej złożone, a ich działanie podlega większej regulacji, białko HSP90 brałoby udział w stabilizowaniu ich struktury w warunkach natywnych regulując wiele szlaków sygnalizacyjnych w komórce. Dlatego też dynamika i wpływ różnych czynników zewnętrznych na konformację białek HSP90 z różnych organizmów mogą być inne. Bakteryjne i drożdżowe białka HSP90 zmieniają konformację pod wpływem wiązania analogu ATP na zamkniętą (Krukenberg i wsp., 2008, Cunningham i wsp., 2012). Natomiast ludzkie białko HSP90 jedynie w bardzo małym stopniu zmienia swoją konformację pod wpływem AMPNP. Pod tym względem białko HSP90 z pszenicy jest bardziej podobne do białka ludzkiego. Podobne zachowanie zaobserwowano dla białka HSP90 z pomarańczy (Mendonça i wsp., 2009), które wykazywało bardzo małe zmiany pod wpływem wiązania innego analogu ATP - ATP γ S.

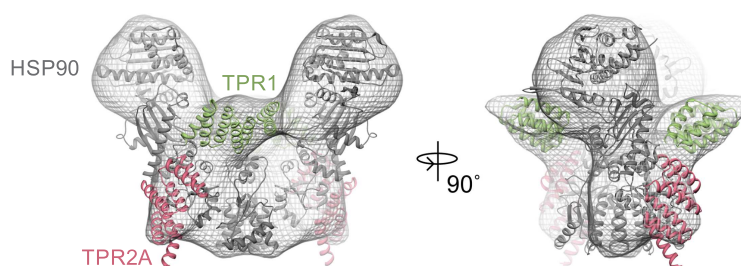
5.5 Stechiometria i struktura kompleksu białek HSP90 i SGT1 z ADP

Białko HSP90, żeby pełnić swoją funkcję w komórce oddziałuje z białkami ko-chaperonami, które zaangażowane są w regulację jego aktywność ATPazy oraz oddziaływania z innymi białkami (Taipale i wsp., 2010). Każdy z ko-chaperonów oddziałuje z białkiem HSP90 w skoordynowany sposób w trakcie cyklu fałdowania białka-klienta przy udziale białka HSP90. Wykazano, że ko-chaperony HSP90 w skoordynowany sposób tworzą asymetryczne kompleksy (Li i wsp., 2011, Ebong i wsp., 2011). W przypadku oddziaływania białka HOP i ko-chaperonów z rodziny izomeras peptydyloprolinowych (PPI), białko to tworzy asymetryczny kompleks, w skład którego wchodzi dimer białka HSP90 i po jednej cząsteczce białka HOP i jednej cząsteczki białka z rodziny PPI: XAP2 (białko PPI) lub FKBP51 lub FKBP52. Samo białko HOP w kompleksie z białkiem HSP90 może występować w postaci asymetrycznego kompleksu zawierającego jedną cząsteczkę białka HOP i dimer białka HSP90 a także, w mniejszej ilości w postaci symetrycznego dimeru dwóch podjednostek białka HOP i dimeru białka HSP90. Wydają się więc że, asymetryczne formowanie się kompleksu jest ważne dla działania kompleksu białka HSP90. Białko HOP oddziałuje z białkiem HSP90 oraz HSP70 i uczestniczy w przekazywaniu klientów pomiędzy tymi dwoma kompleksami. Za pomocą mikroskopii elektronowej określono strukturę kompleksów białka HOP z białkiem HSP90 i białkiem HSP70 z ADP (Alvira i wsp., 2014). Model dla kompleksu białek HSP90-HOP z ADP zaprezentowany jest na rysunku 5.4. Białko HOP oddziałuje z C-terminalnym motywem MEEVD poprzez domenę TPR2A oraz z domeną środkową poprzez domeny TPR1, DP1 i DP2. W kompleksie tym białko HSP90 ma zamkniętą konformację. Dodatkowo domena DP2 białka HOP oddziałuje z fragmentem domeny N-terminalnej białka HSP90. W modelu otrzymanym przez grupę Agarda (Southworth i wsp., 2011) dla kompleksu białek HSP90 i HOP stabilizowanego wiązaniami S-S występują dwie podjednostki białka HOP, a samo białko HSP90 posiada otwartą konformację (rysunek 5.4).

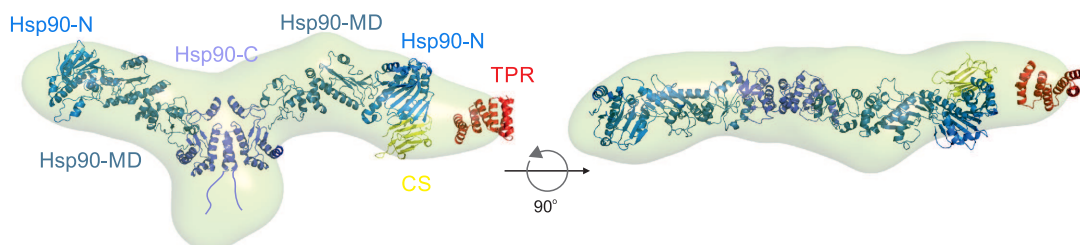
HSP90-HOP-ADP (model TEM barwienie octanem uranylu)



HSP90-HOP (model cryo-TEM)



HSP90-SGT1 Δ SGS-ADP (model SAXS)



Rysunek 5.4 Porównanie struktury kompleksu białek HSP90-HOP z ADP (Alvira i wsp., 2014) w konformacji zamkniętej otrzymanego za pomocą mikroskopii elektronowej (barwienie octanem uranylu), modelu kompleksu białek HSP90-HOP otrzymanego za pomocą techniki cryo-TEM (Southworth i wsp., 2011) z modelem kompleksu HSP90-SGT1 Δ SGS z ADP otrzymanego za pomocą techniki SAXS (Poszczególne domeny białka HSP90 oraz SGT1 zostały nałożone ręcznie na model otrzymany techniką SAXS, nieuporządkowany fragment na C-końcu został dorysowany w programie Corel Draw).

Z naszych badań wynika, że białko SGT1 Δ SGS oddziałuje z białkiem HSP90 w takiej samej stechiometrii jak w strukturze białka HSP90-HOP z ADP otrzymanej za pomocą mikroskopii TEM. W kompleksie HSP90-SGT1 Δ SGS jedna cząsteczka białka SGT1 oddziałuje z dimerem białka HSP90. W odróżnieniu jednak od białka HOP, które występuje jako monomer, białko SGT1 oraz SGT1 Δ SGS występują w postaci dimeru. Tak więc wiązanie białka SGT1 do białka HSP90 powoduje dysocjacje

dimeru białka SGT1. Nadmiar białka SGT1 nad białkiem HSP90 nie powodował powstawania symetrycznego kompleksu o stechiometrii: $(\text{HSP90})_2(\text{SGT1})_2$ lub $(\text{HSP90})_2(\text{SGT1})_4$. Tak więc białko SGT1, najprawdopodobniej tak jak białko HOP, występuje w postaci asymetrycznego kompleksu z białkiem HSP90. Jest to kolejne, oprócz dynamicznego charakteru, modułowej budowy domenowej oraz wiązaniu białka HSP70, podobieństwo pomiędzy białkami HOP i SGT1. Jednakże, w kompleksie z białkiem SGT1 Δ SGS, konformacja białka HSP90 różni się od tej zaobserwowanej w kompleksie z białkiem HOP. Jak już wspomniano wyżej w kompleksie z ADP i białkiem HOP, białko HSP90 ma zamkniętą konformację. W przypadku białka HSP90 z pszenicy w kompleksie z białkiem SGT1 Δ SGS i ADP białko HSP90 ma otwartą konformację. Ponieważ nie stwierdzono oddziaływania domeny SGS z białkiem HSP90, konformacja pełnej długości białka SGT1 powinna być bardzo podobna. Domena TRP białka SGT1 nie oddziałuje z białkiem HSP90 i skierowana jest w przeciwną stronę niż wolny, nie oddziałujący z białkiem SGT1, N-koniec białka HSP90. Być może, zawada sferyczna w postaci domeny N-terminalnej białka HSP90, zaburza oddziaływania pomiędzy domenami TRP w dimerze białka SGT1. Wiązanie białka SGT1 Δ SGS do białka HSP90 oprócz dysocjacji dimeru nie wpływa na jego dynamiczną strukturę co jest odmienne niż w przypadku białka HOP, które oddziałuje z białkiem HSP90 prawie na całej swojej długości.

Uważa się także, że białko SGT1 jest zaangażowane w fałdowanie się białek zawierających domenę LRR. Domena SGS białka SGT1 jest odpowiedzialna za oddziaływanie z domenami LRR. Co ciekawe, jednocześnie wykazano, że domena SGS oddziałuje z białkiem z rodziny HSP70 - HSC70 (ang. heat shock cognate 70 kDa) u *Arabidopsis thaliana* (Noël i wsp., 2007). Z kolei Stuttmann i współpracownicy (2008) zasugerowali, że podobnie jak białko HOP jest zaangażowane w fałdowanie się domeny wiążącej ligand jądrowych receptorów steroidowych, tak białko SGT1 może pełnić analogiczną funkcję w przypadku fałdowania się białek z domeną LRR. Do tej pory nie zidentyfikowano innych białek opiekuńczych mogących oddziaływać z białkiem SGT1 i pośredniczyć w procesie fałdowania się białek z domenami LRR. Białko SGT1 w komórkach drożdży występuje również jako dimer, a dimeryzacja białka SGT1 jest potrzebna do formowania się kinetochoru (Bansal i wsp., 2009a). Dimeryzacja białka SGT1 *in-vivo* nie została potwierdzona w przypadku komórek roślinnych, aczkolwiek nie jest ona wymagana do pełnienia funkcji w odpowiedzi przeciwko patogenom (Botër i wsp., 2007). Nie jest też wiadomo czy sama dimeryzacja jest potrzebna do fałdowania się białek z domeną LRR.

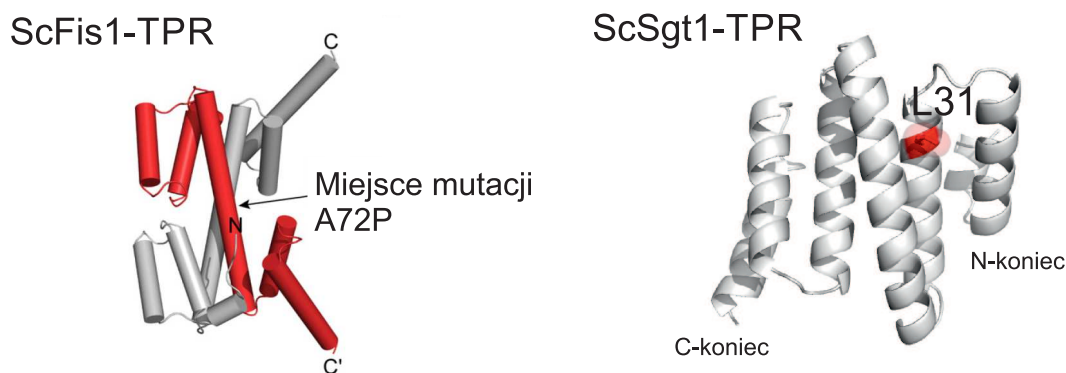
W celu sprawdzenia czy dimeryzacja białka HSP90 potrzebna jest do dysocjacji białka SGT1 po jego związaniu przeprowadzono badania SAXS dla białka HSP90 bez domeny C terminalnej. Białko HSP90 Δ C jest monomerem w roztworze i podobnie jak białko pełnej długości wiąże tylko jedną podjednostkę białka SGT1 Δ SGS, również kiedy białko SGT1 Δ SGS jest w nadmiarze. Wynika więc, że dimeryzacja białka HSP90 nie jest potrzebna do wiązania monomeru białka SGT1 Δ SGS. Wątpliwe jest również, że samo ADP wpływa na dysocjację białka SGT1. Po pierwsze, białko SGT1 Δ SGS w takich samych warunkach jak te użyte do badania kompleksu z pełnej długości białkiem HSP90 jest dimerem. Również w analizie MCR-ALS i SVD wykryto istnienie tylko trzech komponentów: kompleksu oraz nieskompleksowanych białek HSP90/HSP90 Δ C i SGT1 Δ SGS. Otwartą kwestią pozostaje czy sama domena N-terminalna białka HSP90 jest wystarczająca do spowodowania dysocjacji dimeru białek SGT1 i SGT1 Δ SGS. Porównując modele pełnej długości białka HSP90 i białka HSP90 Δ C z

modelami kompleksu białek HSP90-SGT1 Δ SGS, HSP90 Δ C-SGT1 Δ SGS oraz HSP90 Δ C-SGT1-CS z ADP można wywnioskować, że wiązanie domeny CS przez N-terminalną domenę białka HSP90 nie zmienia konformacji samego białka HSP90.

5.6 Analiza dimeryzacji domeny TPR białka SGT1 z różnych organizmów

Powtórzenia TPR są odpowiedzialne są za oddziaływania białko-białko. Najbardziej znanym przykładem oddziaływania jest wiązanie motywu EEVD występującego na C-końcu białek HSP70 i HSP90 przez domeny TPR białka HOP. Innym przykładem oddziaływań z udziałem białek z rodziny TPR jest ich dimeryzacja. Przykładem homooligomeryzacji może być oddziaływanie dwóch białek Cul9 z kompleksu promującego anafazę APC (ang. anaphase promoting complex) czy też dimeryzacja konsensusowej domeny TPR (rysunek 5.5) (Zhang i wsp., 2010b, Krachler i wsp., 2010). Innym ciekawym przykładem asocjacji domeny TPR, zależnej w tym przypadku od temperatury, jest dimeryzacja domeny TPR białka Fis1 (ang. Fision 1) (Lees i wsp., 2012). Białko Fis1 jest zaangażowane w proces dzielenia się mitochondriów i peroksysomów. Wykazano, że domena TPR białka Fis1 pod wpływem temperatury oraz denaturacji po powrocie do natywnych warunków ulega dimeryzacji. Natomiast izolowany dimer po denaturacji powraca do stanu monomerycznego. Dimeryzacja białka Fis1 zachodzi poprzez wymianę helis 1-3 oraz 4-6 pomiędzy podjednostkami. Równowaga monomer-dimer jest istotna dla pełnienia funkcji przez białko Fis1 ponieważ mutanty białka Fis1 (wyłącznie monomer lub dimer) są нефункционалне.

W przypadku domeny TPR z białka SGT1 z jęczmienia, *Arabidopsis thaliana* i drożdży piekarskich dimeryzacja domeny TPR jest zależna od siły jonowej i potencjału redoks. Ludzkie białko SGT1 nie ulega dimeryzacji. Do tej pory nie udało się ustalić struktury przestrzennej dimeru domeny TPR białka SGT1. Za pomocą metody SAXS zbadano zachowanie się w roztworze domeny TPR białka SGT1 z czterech organizmów: jęczmienia, *Arabidopsis thaliana*, drożdży piekarskich oraz człowieka. Dla roślinnych białek TPR dimeryzacja zachodzi w roztworach o niskiej sile jonowej, a równowaga monomer-dimer przesuwa się w stronę monomeru w roztworach o niskiej sile jonowej. Wyniki te są zgodne z wynikami otrzymanymi w pracy Nyarko i współpracowników (Nyarko i wsp., 2007) oraz z wynikami otrzymanymi dla pełnej długości białka SGT1 z jęczmienia. Domena TPR ludzkiego białka SGT1 nie ulegała dimeryzacji, co jest również zgodne z danymi literaturowymi. Dla domeny TPR białka SGT1 z drożdży piekarskich zaobserwowano inne zachowanie. Przede wszystkim w roztworze o niskiej sile jonowej domena TPR występowała w postaci tetrameru o wydłużonym kształcie. Frakcja tetrameru podobnie jak frakcje dimeru w przypadku domen TPR z roślinnych białek SGT1 malała wraz ze wzrostem siły jonowej. Jednocześnie poczynając od 100 mM KCl drożdżowa domena TPR występuje w około 50 % w postaci dimeru. Co ciekawe frakcja dimeru pozostaje na stałym poziomie w zakresie siły jonowej od 100 mM KCl do 800 mM KCl. Dimeryzacja domeny TPR nie jest zależna od siły jonowej.



Rysunek 5.5 Model dimeru domeny TPR białka Fis1 (Lees i wsp., 2012) wraz z miejscem mutacji A72P blokującej dimeryzację oraz model domeny TPR białka ScSGT1 wraz z zaznaczoną leucyną 31, której mutacja L31P powoduje zahamowanie dimeryzacji białka SGT1 *in-vivo*.

Wraz ze wzrostem siły jonowej wzrasta udział wkładu monomeru domeny TPR drożdżowego białka SGT1. Jednak monomer tej domeny najprawdopodobniej jest częściowo nieuporządkowany, jak wynika z wykonanych przeze mnie eksperymentów SAXS. Odczyn roztworu użytego w eksperymentach SAXS był obojętny ($\text{pH} = 7,5$) i był on bardzo bliski punktowi izoelektrycznemu obliczonemu na podstawie sekwencji aminokwasowej dla konstruktów domeny TPR z drożdżowego białka SGT1. Być może domena TPR w tych warunkach jest bardziej niestabilna niż domeny TPR z białka SGT1 z innych badanych przeze mnie organizmów. Konstrukt domeny TPR nie zawierał też pierwszych 6 aminokwasów (sześciu alanin), które nie są zakonserwowane pomiędzy białkami SGT1 z różnych organizmów (Azevedo i wsp., 2002). Dlatego też, delekcja tych aminokwasów mogła wpływać negatywnie na zachowanie się drożdżowej domeny TPR w roztworze. Jednocześnie częściowe rozplecenie domeny TPR umożliwiałyby podobną wymianę domen jak w przypadku białka Fis1. Co ciekawe mutacja leucyny w pozycji 31 powodująca jej zmianę w prolinę jest odpowiedzialna za zahamowanie dimeryzacji drożdżowego białka SGT1 (Bansal i wsp., 2009a). Podobnie, mutacja powodująca zamianę alaniny 72 na prolinę w obrębie domeny TPR białka Fis1 powodowała zahamowanie dimeryzacji i wymiany domen. Na rysunku 5.5 przedstawiony jest model dimeru białka Fis1 oraz monomeru domeny TPR białka ScSGT1 wraz z zaznaczonymi miejscami mutacji zaburzającymi dimeryzację.

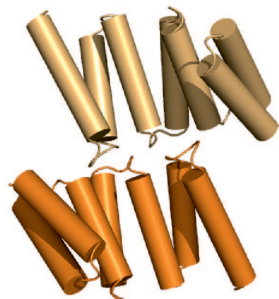
5.6.1 Modelowanie dimeru domeny TPR białka SGT1

Niskorozdzielcze modele otrzymane na podstawie danych SAXS dla monomeru domeny TPR białka SGT1 z jęczmienia, *Arabidopsis thaliana* i człowieka mają bardzo podobny kształt, tak samo jak modele otrzymane z programów do modelowania bioinformatycznego. Także, struktura monomeru jest zachowana pomiędzy badanymi organizmami. Oczywiście mogą występować różnice, których nie da się zauważyć za pomocą techniki SAXS. Jedynie dla drożdży piekarskich domena TPR ma wydłużony kształt i najprawdopodobniej ma nieuporządkowany charakter.

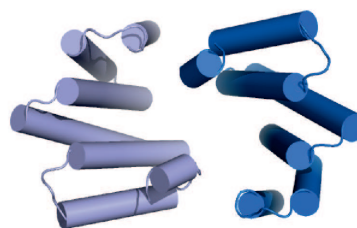
Jak dotąd brak jest dokładnych danych strukturalnych odnośnie dimeryzacji domeny TPR białka SGT1. Wiadomo jedynie z eksperymentów spektrometrii masowej w połączeniu z chemicznym sieci-

waniem, że w dimeryzacji przypuszczalnie bierze udział C-terminalna część domeny TPR. Wykazano również, że oksydacja pary cystein, która także znajduje się w C-terminalnej części roślinnych białek SGT1 zaburza dimeryzację. Na podstawie danych SAXS udało się otrzymać modele trójwymiarowe dimeru domeny TPR białka SGT1 z jęczmienia, *Arabidopsis thaliana* i drożdży piekarskich. Dla domeny TPR z jęczmiennego białka TPR udało się otrzymać tylko jeden model, który jest zgodny z doświadczeniami ze spektrometrii masowej. Natomiast dla białek z *Arabidopsis* oraz drożdży piekarskich otrzymano kilka modeli, których struktura różni się między sobą choć wszystkie mają bardzo podobne krzywe SAXS i dopasowanie do danych eksperymentalnych. Modele otrzymane w programie DAMMIN dla różnych organizmów są bardzo podobne do siebie co potwierdza, że najprawdopodobniej struktura dimeru może być zachowana pomiędzy organizmami, choć jak było wspomniane wyżej istnieją różnice w zachowaniu się dimeru pomiędzy roślinami a drożdżami. W przypadku domeny TPR z *Arabidopsis* otrzymane modele dimeryzują boczną stroną domeny TPR poprzez pętle łączące helisy w powtórzeniu TPR. Dla białka drożdżowego otrzymano modele, które oddziałują wklęsłą stroną białka (tworzą strukturę podobną do chińskiego znaku Yin i Yang). W obu typach modeli C-terminalna część białka bardzo często bierze udział w dimeryzacji. Niestety rozdzielczość danych SAXS nie pozwala jednoznacznie wyznaczyć struktury dimeru białka SGT1. Nie mniej jednak znacznie przybliżyła nas do poznania dokładnego mechanizmu dimeryzacji. Strukturę dimeru domen TPR podobną do modelu dla drożdżowej domeny TPR zaobserwowano dla białka Cdc23 z drożdży *Schizosaccharomyces pombe*. Dwie podjednostki N-terminalnej domeny tworzą także strukturę podobną do znaku Yin i Yang (Zhang i wsp., 2013). Z drugiej strony dla domeny TPR białka BUB1 określono strukturę dimeru, w której to dimeryzacja zachodzi poprzez oddziaływanie pętli pomiędzy helisami z powtórzeń TPR (Bolanos-Garcia i wsp., 2009). Także, porównując struktury dimerów innych białek TPR można wywnioskować, że oba modele dimeryzacji domeny TPR białka SGT1 są równo prawdopodobne. Na rysunku 5.6 przedstawione jest porównanie modeli dla dimeru domeny TPR białka SGT1 z białkami BUB1 i Cdc23. Co ciekawe otrzymany niezależnie model tetrameru domeny TPR drożdżowego białka SGT1 zawiera rdzeń w postaci dimeru domeny TPR. Rdzeń ten ma taką samą strukturę jak otrzymany wcześniej, na podstawie krzywej SAXS, model dla dimeru domeny TPR. Także, bardzo możliwe jest, że dimer domeny TPR stanowi integralną część tetrameru i wraz ze wzrostem siły jonowej, tetramer ulega dysocjacji na dimer oraz dwie podjednostki monomeryczne.

Model SAXS domeny TPR białka AtSGT1a



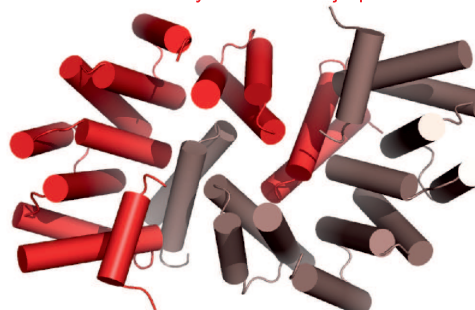
Model SAXS domeny TPR białka ScSGT1a



Struktura domeny TPR białka BUB1



Struktura dimeru domeny N-terminalnej SpCdc23



Rysunek 5.6 Porównanie modeli dla dimeru domeny TPR białka SGT1 z *Arabidopsis thaliana* i drożdży piekarskich z strukturami dimeru domeny TPR białka BUB1 z drożdży piekarskich (kod pdb: 3ESL, Bolanos-Garcia i wsp., 2009) oraz Cdc23 (kod pdb: 3ZN3, Zhang i wsp., 2013) z drożdży *Schizosaccharomyces pombe*.



6 Podsumowanie

Otrzymane dotychczas wyniki prowadzonych przeze mnie prac badawczych w ramach pracy doktorskiej można podsumować w postaci przedstawionych poniżej wniosków:

- Białko SGT1 z jęczmienia, zarówno w formie monomerycznej jak i dimerycznej posiada dynamiczną strukturę, w której regiony zmienne (VR1 i VR2) pomiędzy domenami TPR, CS i SGS są nieuporządkowane.
- Białko RAR1 podobnie jak białko SGT1 ma wydłużony kształt i dynamiczną strukturę, w której domeny CHORD I i CHORD II zachowują się jak bryły sztywne a fragment pomiędzy nimi jest nieuporządkowany.
- Niskorozdzielcza struktura w roztworze monomeru domeny TPR białka SGT1 z *Arabidopsis thaliana*, jęczmienia i ludzkiego homologu jest bardzo podobna. Z modelowania bioinformatycznego wynika, że struktura domeny TPR białka SGT1 jest bardzo podobna do struktury domeny TPR występujących w innych białkach.
- Domena TPR z roślinnych białek SGT1 występuje w roztworze w równowadze monomer-dimer, a równowaga ta zależna jest od siły jonowej. W modelu dimeru domeny TPR jęczmiennego białka SGT1 w dimeryzacji bierze udział C-terminalny fragment białka, co zgadza się z wynikami spektrometrii mas i sieciowania chemicznego. Dla domeny TPR z białka SGT1a z *Arabidopsis* otrzymano na podstawie danych SAXS model dimeru, w którym dwie podjednostki domeny TPR oddziałują poprzez pętle pomiędzy helisami α powtórzeń TPR znajdujące się na bocznej krawędzi modelu domeny TPR. Dla domeny TPR z *Arabidopsis* otrzymano również model, w którym w dimeryzacji bierze udział C-terminalny fragment. Dla obu typów otrzymanych modeli wartość dopasowania χ jest bardzo podobna.
- Domena TPR drożdżowego białka SGT1 w roztworze występuje w równowadze monomer-dimer-tetramer. Frakcja tetrameru przeważa w roztworach o niskiej sile jonowej. Wraz ze wzrostem siły jonowej rośnie wkład frakcji monomerycznej. Natomiast frakcja dimeru w zakresie stężenia KCl od 100 mM do 800 mM występuje na stałym poziomie 50 %. Tak więc dimeryzacja domeny TPR nie zależy od siły jonowej. Niskorozdzielczy model dimeru domeny TPR ma kształt bardzo podobny do modeli otrzymanych dla domeny TPR z białka SGT1 z *Arabidopsis thaliana* i jęczmienia.
- Niskorozdzielczy model tetrameru domeny TPR białka SGT1 z drożdży piekarskich ma wydłużony kształt. Rdzeń tetrameru stanowi dimer domeny TPR o strukturze takiej samej jak model otrzymany dla dimeru domeny TPR.
- Monomer domeny TPR białka SGT1 z drożdży piekarskich jest częściowo nieuporządkowany. Nieuporządkowany monomer podobnie jak białko Fis1 z drożdży piekarskich mógłby dimery-

zować poprzez częściową wymianę domen, co prowadziłyby do powstania dimeru odpornego na wysokie stężenie soli. W modelu równowagi oligomerycznej domeny TPR białka SGT1 z drożdży piekarskich tetramer składa się z dimeru najprawdopodobniej powstałego poprzez wymianę domen oraz dwóch podjednostek monomerycznych asocjowanych z dimerem. Wzrost siły jonowej powoduje oddysocjowanie dwóch podjednostek monomerycznych od dimeru domen TPR.

- Białko HSP90 z jęczmienia ma niskorozdzielczą strukturę w roztworze bardzo podobną do tej zaobserwowanej dla ludzkiego i drożdżowego białka HSP90. Białko HSP90 występuje w roztworze głównie postaci konformacji otwartej a wiązanie nukleotydów ADP lub AMPPNP nie zmienia zasadniczo równowagi konformacyjnej.
- Temperatura i siła jonowa znacząco wpływają na dynamikę konformacyjną białka HSP90 z pszenicy. W roztworze białko HSP90 z pszenicy występuje w równowadze trzech konformacji: otwartej, zamkniętej i pośredniej. Zarówno wysoka siła jonowa jak i temperatura przesuwać równowagę konformacyjną w stronę konformacji otwartej.
- Konformacja pośrednia białka HSP90 z pszenicy jest bardzo podobna do struktury białka HtpG z *E.coli* zaobserwowanej w roztworach o wysokim odczynie pH. Pozostałe stany konformacje białka HSP90 z pszenicy są podobne do struktury białek z rodziny HSP90 w roztworze lecz wykazują wyraźne różnice.
- Czynniki zewnętrzne takie jak pH, temperatura i siła jonowa wykazują podobny wpływ na równowagę konformacyjną białek z rodziny HSP90 z różnych organizmów i powodują przesunięcie tej równowagi w stronę konformacji otwartej.
- Za pomocą analizy SVD i MCR-ALS zbadano kompleks białek HSP90 i SGT1 z ADP z jęczmienia. Za pomocą techniki MCR-ALS otrzymano krzywe rozpraszania dla kompleksu białek HSP90-SGT1 Δ SGS, HSP90 Δ C-SGT1 Δ SGS oraz HSP90 Δ C-SGT1-CS.
- Białko SGT1 Δ SGS w kompleksie z pełnej długości białkiem HSP90 występuje w formie monomerycznej i tworzy asymetryczny kompleks o stechiometrii 2:1. W kompleksie z białkiem SGT1 Δ SGS białko HSP90 występuje w otwartej konformacji.
- Dimeryzacja białka HSP90 nie jest potrzebna do dysocjacji dimeru białka SGT1 Δ SGS podczas jego przyłączania. Białko HSP90 Δ C jest monomerem w roztworze i wiąże się z białkiem SGT1 Δ SGS w stosunku molowym 1:1.
- Domena CS białka SGT1 oddziałuje z N-terminalną domeną białka HSP90 Δ C. Wiązanie domeny CS nie zmienia ogólnej konformacji białka HSP90 Δ C.

Uzyskane wyniki pozwalają w chwili obecnej na zaplanowanie kolejnych badań w szczególności z wykorzystaniem mikroskopii elektronowej czy spektrometrii mas, które w przyszłości będą mogły zaowocować pełnym opisem mechanizmu tworzenia kompleksu białek HSP90-SGT1-RAR1.



7 Literatura

A

- Alberts B, Johnson A., Lewis J., Raff M., Roberts K., Walter P. *Molecular Biology of the Cell*. 2002. Garland Science. New York.
- Ali MM, Roe SM, Vaughan CK, Meyer P, Panaretou B, Piper PW, Prodromou C, Pearl LH. Crystal structure of an Hsp90-nucleotide-p23/Sba1 closed chaperone complex. *Nature*. 2006. 440, 1013-1017.
- Alvira S, Cuéllar J, Röhl A, Yamamoto S, Itoh H, Alfonso C, Rivas G, Buchner J, Valpuesta JM. Structural characterization of the substrate transfer mechanism in Hsp70/Hsp90 folding machinery mediated by Hop. *Nat Commun*. 2014. 5, 5484.
- Andrade MA, Chacón P, Merelo JJ, Morán F. Evaluation of secondary structure of proteins from UV circular dichroism using an unsupervised learning neural network. *Prot. Engineering* 1993. 6, 383-390.
- Austin MJ, Muskett P, Kahn K, Feys BJ, Jones JD, Parker JE. Regulatory role of SGT1 in early R gene-mediated plant defenses. *Science*. 2002. 295, 2077-2080.
- Azevedo C, Sadanandom A, Kitagawa K, Freialdenhoven A, Shirasu K, Schulze-Lefert P. The RAR1 interactor SGT1, an essential component of R gene-triggered disease resistance. *Science*. 2002. 295, 2073-2076.
- Azevedo C, Betsuyaku S, Peart J, Takahashi A, Noël L, Sadanandom A, Casais C, Parker J, Shirasu K. Role of SGT1 in resistance protein accumulation in plant immunity. *EMBO J*. 2006. 25, 2007-2016.

B

- Ballatore C, Lee VM, Trojanowski JQ. Tau-mediated neurodegeneration in Alzheimer's disease and related disorders. *Nat Rev Neurosci*. 2007. 8, 663-672.
- Bansal PK, Nourse A, Abdulle R, Kitagawa K. Sgt1 dimerization is required for yeast kinetochore assembly. *J Biol Chem*. 2009a. 284, 3586-3592.
- Bansal PK, Mishra A, High AA, Abdulle R, Kitagawa K. Sgt1 dimerization is negatively regulated by protein kinase CK2-mediated phosphorylation at Ser361. *J Biol Chem*. 2009b. 284, 18692-18698.
- Bernadó P, Mylonas E, Petoukhov MV, Blackledge M, Svergun DI. Structural characterization of flexible proteins using small-angle X-ray scattering. *J Am Chem Soc*. 2007, 129, 5656-5664.
- Bernoux M, Ve T, Williams S, Warren C, Hatters D, Valkov E, Zhang X, Ellis JG, Kobe B, Dodds PN. Structural and functional analysis of a plant resistance protein TIR domain reveals interfaces for self-association, signaling, and autoregulation. *Cell Host Microbe*. 2011. 9, 200-211.
- Bieri S, Mauch S, Shen QH, Peart J, Devoto A, Casais C, Ceron F, Schulze S, Steinbiss HH, Shirasu K, Schulze-Lefert P. RAR1 positively controls steady state levels of barley MLA resistance proteins and enables sufficient MLA6 accumulation for effective resistance. *Plant Cell*. 2004. 16, 3480-3495.
- Blakemore R. Magnetotactic bacteria. *Science*. 1975. 190, 377-379.
- Blobel J, Bernadó P, Svergun DI, Tauler R, Pons M. Low-resolution structures of transient protein-protein complexes using small-angle X-ray scattering. *J Am Chem Soc*. 2009, 131, 4378-4386.
- Bolanos-Garcia VM, Kiyomitsu T, D'Arcy S, Chirgadze DY, Grossmann JG, Matak-Vinkovic D, Venkitaraman AR, Yanagida M, Robinson CV, Blundell TL. The crystal structure of the N-terminal region of BUB1 provides insight into the mechanism of BUB1 recruitment to kinetochores. *Structure*. 2009. 17, 105-116.
- Borkovich KA, Farrelly FW, Finkelstein DB, Taulien J, Lindquist S. Hsp82 is an essential

protein that is required in higher concentrations for growth of cells at higher temperatures. *Mol Cell Biol.* 1989. 9, 3919-3930.

- Botër M, Amigues B, Peart J, Breuer C, Kadota Y, Casais C, Moore G, Kleanthous C, Ochsenbein F, Shirasu K, Guerois R. Structural and functional analysis of SGT1 reveals that its interaction with HSP90 is required for the accumulation of Rx, an R protein involved in plant immunity. *Plant Cell.* 2007. 19, 3791-3804.
- Bron P, Giudice E, Rolland JP, Buey RM, Barbier P, Díaz JF, Peyrot V, Thomas D, Garnier C. Apo-Hsp90 coexists in two open conformational states in solution. *Biol Cell.* 2008. 100, 413-25.
- Bragg WL. The Diffraction of Short Electromagnetic Waves by a Crystal. *Proceedings of the Cambridge Philosophical Society*, 1913. 17, 43–57.
- Brancaccio M, Menini N, Bongioanni D, Ferretti R, De Acetis M, Silengo L, Tarone G. Chp-1 and melusin, two CHORD containing proteins in vertebrates. *FEBS Lett.* 2003. 551, 47-52.

C

- Cantor CR, Schimmel PR. *Biophysical Chemistry. Part II: techniques for the study of biological structure and function.* 1998. WH Freeman and company, New York.
- Caplan J, Padmanabhan M, Dinesh-Kumar SP. Plant NB-LRR immune receptors: from recognition to transcriptional reprogramming. *Cell Host Microbe.* 2008. 3, 126-135.
- Carlisle CH. Crowfoot D. The Crystal Structure of Cholesteryl Iodide. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences.* 1945. 184, 64-83.
- Catlett, M. G., and Kaplan, K. B. (2006) Sgt1p Is a Unique Cochaperone That Acts as a Client Adaptor to Link Hsp90 to Skp1p, *J. Biol. Chem.* 281, 33739-33748.
- Cavanagh J, Fairbrother WJ, Palmer AG, Rance M, NJ Skeleton. *Protein NMR spectroscopy: principles and practice.* Academic Press. Londyn. 2007.
- Chacón P, Morán F, Díaz JF, Pantos E, Andreu JM. Low- Resolution Structures of Proteins in Solution Retrieved from X-Ray Scattering with a Genetic Algorithm, *Biophys. J.* 1998, 74, 2760-2775.
- Chadli A, Ladjimi MM, Baulieu EE, Catelli MG. Heat-induced oligomerization of the molecular chaperone Hsp90. Inhibition by ATP and geldanamycin and activation by transition metal oxyanions. *J Biol Chem.* 1999. 274, 4133-4139.
- Chan SL, Mukasa T, Santelli E, Low LY, Pascual J. The crystal structure of a TIR domain from *Arabidopsis thaliana* reveals a conserved helical region unique to plants. *Protein Sci.* 2010. 19, 55-161.
- Chinchilla D, Bauer Z, Regenass M, Boller T, Felix G: The *Arabidopsis* receptor kinase FLS2 binds flg22 and determines the specificity of flagellin perception. *Plant Cell* 2006. 18, 465-476.
- Collins NC, Thordal-Christensen H, Lipka V, Bau S, Kombrink E, Qiu JL, Hüchelhoven R, Stein M, Freialdenhoven A, Somerville SC, Schulze-Lefert P. SNARE-protein-mediated disease resistance at the plant cell wall. *Nature.* 2003. 425, 973-977.
- Compton LA, Johnson WC Jr. Analysis of protein circular dichroism spectra for secondary structure using a simple matrix multiplication. *Anal Biochem.* 1986. 155, 155-67.
- Crowfoot D. Bunn CW. Rogers-Low BW. Turner-Jones A. X-ray crystallographic investigation of the structure of penicillin. 1949. *Chemistry of Penicillin.* Princeton University Press. 310–367.
- Cunningham CN, Southworth DR, Krukenberg KA, Agard DA. The conserved arginine 380 of Hsp90 is not a catalytic residue, but stabilizes the closed conformation required for ATP hydrolysis. *Protein Sci.* 2012. 21, 1162-1171.

D

- Dangl JL, Jones JD. Plant pathogens and integrated defence responses to infection. *Nature.* 2001. 411, 826-833.

- D'Andrea LD, Regan L. TPR proteins: the versatile helix. *Trends Biochem Sci.* 2003. 28, 655-662.
- Das AK, Cohen PW, Barford D. The structure of the tetratricopeptide repeats of protein phosphatase 5: implications for TPR-mediated protein-protein interactions. *EMBO J.* 1998. 17, 1192-1199.
- da Silva Correia J, Miranda Y, Leonard N, Ulevitch R. SGT1 is essential for Nod1 activation. *Proc Natl Acad Sci U S A.* 2007. 104, 6764-6769.
- Debye P. Zerstreung von Röntgenstrahlen. *Ann. Phys. (Leipzig)* 1915. 46, 809-823.
- Debye P. Molecular weight determination by light scattering. *J.Phys.Colloid.Chem.* 1947. 51, 18-32
- Debye P, Anderson JHR, Brumberger H. *Journal of Applied Physics.* 1957. 28, 679-683.
- de Juan A, Jaumotb J, Tauler R. Multivariate Curve Resolution (MCR). Solving the mixture analysis problem. *Anal. Methods*, 2014, 6, 4964-4976.
- Deng D, Yan C, Pan X, Mahfouz M, Wang J, Zhu JK, Shi Y, Yan N. Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science.* 2012. 335, 720-723.
- Deslandes L, Rivas S. Catch me if you can: bacterial effectors and plant targets. *Trends Plant Sci.* 2012. 17, 644-655.
- Dickinson RG. Raymond AL, The crystal structure of hexamethylene-tetramine. *J. Amer. Chem. Soc.* 1923, 45, 22-29.
- Dittmar KD, Banach M, Galigniana MD, Pratt WB. The role of DnaJ-like proteins in glucocorticoid receptor.hsp90 heterocomplex assembly by the reconstituted hsp90.p60.hsp70 foldosome complex. *J Biol Chem.* 1998. 273, 7358-7366.
- Dollins DE, Warren JJ, Immormino RM, Gewirth DT. Structures of GRP94-nucleotide complexes reveal mechanistic differences between the hsp90 chaperones. *Mol Cell.* 2007. 28, 41-56.
- Dubacq C, Guerois R, Courbeyrette R, Kitagawa K, Mann C. Sgt1p contributes to cyclic AMP pathway activity and physically interacts with the adenylyl cyclase Cyr1p/Cdc35p in budding yeast. *Eukaryot Cell.* 2002. 1, 568-582.
- Durrant WE, Dong X. Systemic Acquired Resistance. *Annu. Rev. Phytopathol.* 2004. 42, 185-209.
- Dutta R, Inouye M. GHKL, an emergent ATPase/kinase superfamily. *Trends Biochem Sci.* 2000. 25, 24-28.

E

- Enninga J, Rosenshine I. Imaging the assembly, structure and activity of type III secretion systems. *Cell Microbiol.* 2009. 11, 1462-1470.

F

- Feigin LA, Svergun DI. (1987) *Structure Analysis by Small-Angle X-Ray and Neutron Scattering.* Plenum Press. Nowy York. 1987.
- Feng X, Zhang Q, Conga P, Zhu Z. Determination of the paracetamol degradation process with online UV spectroscopic and multivariate curve resolution-alternating least squares methods: comparative validation by HPLC. *Anal. Methods.* 2013, 5, 5286-5293.
- Finkelstein R. Abscisic Acid Synthesis and Response. *The Arabidopsis Book.* 2013. 11: e0166.
- Fournet G, Bull. Soc. Fr. Mineral. Crist. 1951. 74, 39-113.
- Franke J, Eichner S, Zeilinger C, Kirschning A. Targeting heat-shock-protein 90 (Hsp90) by natural products: geldanamycin, a show case in cancer therapy. *Nat Prod Rep.* 2013. 30, 1299-1323.
- Friedrich W, P. Knipping P, Laue M. Interferenzerscheinungen bei Röntgenstrahlen. *Annalen der Physik* 1913. 346, 971-988.
- Fu ZQ, Dong X. Systemic acquired resistance: turning local infection into global defense. *Annu Rev Plant Biol.* 2013. 64, 839-863.

G

- Galantini L, Leggio C, Konarev PV, Pavel NV. Human serum albumin binding ibuprofen: a 3D description of the unfolding pathway in urea. *Biophys Chem.* 2010, 147, 111-122.
- Gay NJ, Symmons MF, Gangloff M, Bryant CE. Assembly and localization of Toll-like receptor signalling complexes. *Nat Rev Immunol.* 2014. 14, 546-558.
- Goel AK, Lundberg D, Torres MA, Matthews R, Akimoto-Tomiyama C, Farmer L, Dangel JL, Grant SR. The *Pseudomonas syringae* type III effector HopAM1 enhances virulence on water-stressed plants. *Mol Plant Microbe Interact.* 2008. 21, 361-370.
- Gołaś E, Maisuradze GG, Senet P, Ołdziej S, Czaplewski C, Scheraga HA, Liwo A. Simulation of the opening and closing of Hsp70 chaperones by coarse-grained molecular dynamics. *J Chem Theory Comput.* 2012. 8, 1750-1764.
- Göhre V, Robatzek S. Breaking the barriers: microbial effector molecules subvert plant immunity. *Annu Rev Phytopathol.* 2008. 46, 189-215.
- Gray WM, Muskett PR, Chuang HW, Parker JE. Arabidopsis SGT1b is required for SCF(TIR1)-mediated auxin response. *Plant Cell.* 2003. 15, 1310-1319.
- Grishaev A, Wu J, Trehwella J, Bax A. Refinement of multidomain protein structures by combination of solution small-angle X-ray scattering and NMR data. *J Am Chem Soc.* 2005, 127,16621-16628.
- Grove TZ, Osuji CO, Forster JD, Dufresne ER, Regan L. Stimuli-responsive smart gels realized via modular protein design. *J Am Chem Soc.* 2010. 132, 14024-14026.
- Grove TZ, Forster J, Pimienta G, Dufresne E, Regan L. A modular approach to the design of protein-based smart gels. *Biopolymers.* 2012. 97, 508-517.
- Guinier A, Fournet G. Small-angle scattering of X-rays. John Wiley and Sons, Inc. New York. 1955.

H

- Hagn F, Lagleder S, Retzlaff M, Rohrberg J, Demmer O, Richter K, Buchner J, Kessler H. Structural analysis of the interaction between Hsp90 and the tumor suppressor protein p53. *Nat Struct Mol Biol.* 2011. 18, 1086-1093.
- Heise CT, Le Duff CS, Boter M, Casais C, Airey JE, Leech AP, Amigues B, Guerois R, Moore GR, Shirasu K, Kleanthous C. Biochemical characterization of RAR1 cysteine- and histidine-rich domains (CHORDs): a novel class of zinc-dependent protein-protein interaction modules. *Biochemistry.* 2007. 46, 1612-1623.
- Hennessey JP Jr, Johnson WC Jr. Information content in the circular dichroism of proteins. *Biochemistry.* 1981. 20, 1085-1094.
- Hills RD Jr, Brooks CL 3rd. Insights from coarse-grained Go models for protein folding and dynamics. *Int J Mol Sci.* 2009, 10, 889-905.
- Hodgkin DC, Pickworth J, Robertson JH, Trueblood KN, Prosen RJ, White JG. Structure of Vitamin B12 : The Crystal Structure of the Hexacarboxylic Acid derived from B12 and the Molecular Structure of the Vitamin. *Nature* 1955. 176, 325-328.
- Hong TJ, Kim S, Wi AR, Lee P, Kang M, Jeong JH, Hahn JS. Dynamic nucleotide-dependent interactions of cysteine- and histidine-rich domain (CHORD)-containing Hsp90 cochaperones Chp-1 and melusin with cochaperones PP5 and Sgt1. *J Biol Chem.* 2013. 288, 215-222.
- Hoser R, Zurczak M, Lichočka M, Zuzga S, Dadlez M, Samuel MA, Ellis BE, Stuttmann J, Parker JE, Hennig J, Krzymowska M. Nucleocytoplasmic partitioning of tobacco N receptor is modulated by SGT1. *New Phytol.* 2013. 200, 158-171.
- Hu Z, Yan C, Liu P, Huang Z, Ma R, Zhang C, Wang R, Zhang Y, Martinon F, Miao D, Deng H, Wang J, Chang J, Chai J. Crystal structure of NLRC4 reveals its autoinhibition mechanism. *Science.* 2013. 341, 172-175.
- Hutagalung AH, Landsverk ML, Price MG, Epstein HF. The UCS family of myosin chaperones.

J. Cell Sci. 2002. 115, 3983–3990.

I

- Itoh H, Ogura M, Komatsuda A, Wakui H, Miura AB, Tashima Y. A novel chaperone-activity-reducing mechanism of the 90-kDa molecular chaperone HSP90. *Biochem J.* 1999. 343, 697-703.

J

- Jacob F, Vernaldi S, Maekawa T. Evolution and Conservation of Plant NLR Functions. *Front Immunol.* 2013. 4, 297.
- Jaumot J, Escaja N, Gargallo R, González C, Pedroso E, Tauler R. Multivariate curve resolution: a powerful tool for the analysis of conformational transitions in nucleic acids. *Nucleic Acids Res.* 2002 Sep 1;30(17):e92.
- Johnson WC. Analyzing protein circular dichroism spectra for accurate secondary structures. *Proteins.* 1999. 35, 307-312.
- Johansson ON, Fantozzi E, Fahlberg P, Nilsson AK, Buhot N, Tör M, Andersson MX. Role of the penetration-resistance genes PEN1, PEN2 and PEN3 in the hypersensitive response and race-specific resistance in *Arabidopsis thaliana*. *Plant J.* 2014. 79, 466-76.
- Jones JD, Dangl JL. The plant immune system. *Nature.* 2007. 444, 323-329.

K

- Kadota Y, Amigues B, Ducassou L, Madaoui H, Ochsenbein F, Guerois R, Shirasu K. Structural and functional analysis of SGT1-HSP90 core complex required for innate immunity in plants. *EMBO Rep.* 2008. 9, 1209-1215.
- Kadota Y, Shirasu K, Guerois R. NLR sensors meet at the SGT1-HSP90 crossroad. *Trends Biochem Sci.* 2010, 4,199-207.
- Kajander T, Cortajarena AL, Mochrie S, Regan L. Structure and stability of designed TPR protein superhelices: unusual crystal packing and implications for natural TPR proteins. *Acta Crystallogr D Biol Crystallogr.* 2007. 63, 800-811.
- Kaku H, Nishizawa Y, Ishii-Minami N, Akimoto-Tomiya C, Dohmae N, Takio K, Minami E, Shibuya N. Plant cells recognize chitin fragments for defense signaling through a plasma membrane receptor. *Proc Natl Acad Sci U S A.* 2006. 103, 11086-11091.
- Karagöz GE, Duarte AM, Akoury E, Ippel H, Biernat J, Morán Luengo T, Radli M, Didenko T, Nordhues BA, Veprintsev DB, Dickey CA, Mandelkow E, Zweckstetter M, Boelens R, Madl T, Rüdiger SG. Hsp90-Tau complex reveals molecular basis for specificity in chaperone action. *Cell.* 2014. 156, 963-974.
- Karanicolas J, Brooks CL 3rd. The origins of asymmetry in the folding transition states of protein L and protein G. *Protein Sci.* 2002. 11, 2351-2361.
- Kendrew JC, Bodo G, Dintzis HM, Parrish RG, Wyckoff H, Phillips DC. A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis. *Nature.* 1958, 181, 662–666.
- Kim YJ, Lin NC, Martin GB. Two distinct *Pseudomonas* effector proteins interact with the Pto kinase and activate plant immunity. *Cell.* 2002. 109, 589-598.
- Kim YC, Tang C, Clore GM, Hummer G. Replica exchange simulations of transient encounter complexes in protein-protein association. *Proc Natl Acad Sci U S A.* 2008. 105, 12855-12860.
- Kim KH, Muniyappan S, Oang KY, Kim JG, Nozawa S, Sato T, Koshihara SY, Henning R, Kosheleva I, Ki H, Kim Y, Kim TW, Kim J, Adachi S, Ihee H. Direct observation of cooperative protein structural dynamics of homodimeric hemoglobin from 100 ps to 10 ms with pump-probe X-ray solution scattering. *J Am Chem Soc.* 2012, 134, 7001-7008.
- Kirkpatrick S, Gelatt Jr CD, Vecchi MP. Optimization by Simulated Annealing. *Science.* 1983, 220, 671–680.
- Kitagawa K, Skowrya D, Elledge SJ, Harper JW, Hieter P. SGT1 encodes an essential com-

ponent of the yeast kinetochore assembly pathway and a novel subunit of the SCF ubiquitin ligase complex. *Mol Cell*. 1999. 4, 21-33.

- Klug A, Caspar DL. The structure of small viruses. *Adv Virus Res*. 1960, 7, 225-325.
- Kobe B, Deisenhofer J. Crystal structure of porcine ribonuclease inhibitor, a protein with leucine-rich repeats. *Nature*. 1993. 366, 751-756.
- Kolinski A. Protein modeling and structure prediction with a reduced representation. *Acta Biochim Pol*. 2004. 51, 349-371.
- Kozak M, Lewandowska A, Oldziej S, Rodziewicz-Motowidło S, Liwo A. Combination of SAXS and NMR Techniques as a Tool for the Determination of Peptide Structure in Solution. *J. Phys. Chem. Lett.*, 2010. 1, 3128–3131
- Krachler AM, Sharma A, Kleanthous C. Self-association of TPR domains: Lessons learned from a designed, consensus-based TPR oligomer. *Proteins*. 2010. 78, 2131-2143.
- Krukenberg KA, Förster F, Rice LM, Sali A, Agard DA. Multiple conformations of *E. coli* Hsp90 in solution: insights into the conformational dynamics of Hsp90. *Structure*. 2008. 16, 755-765.
- Krukenberg KA, Southworth DR, Street TO, Agard DA. pH-dependent conformational changes in bacterial Hsp90 reveal a Grp94-like conformation at pH 6 that is highly active in suppression of citrate synthase aggregation. *J Mol Biol*. 2009a. 390, 278-291.
- Krukenberg KA, Böttcher UM, Southworth DR, Agard DA. Grp94, the endoplasmic reticulum Hsp90, has a similar solution conformation to cytosolic Hsp90 in the absence of nucleotide. *Protein Sci*. 2009b. 18, 1815-1827.
- Krukenberg KA, Street TO, Lavery LA, Agard DA. Conformational dynamics of the molecular chaperone Hsp90. *Q Rev Biophys*. 2011. 44, 229-255.

L

- Landau LD, Lifshitz EM. *Classical Theory of Fields*. Pergamon Press. Oxford. 1985.
- Lavery LA, Partridge JR, Ramelot TA, Elnatan D, Kennedy MA, Agard DA. Structural asymmetry in the closed state of mitochondrial Hsp90 (TRAP1) supports a two-step ATP hydrolysis mechanism. *Mol Cell*. 2014. 53, 330-343.
- Lee YT, Jacob J, Michowski W, Nowotny M, Kuznicki J, Chazin WJ. Human Sgt1 binds HSP90 through the CHORD-Sgt1 domain and not the tetratricopeptide repeat domain. *J Biol Chem*. 2004. 279, 16511-16517.
- Lee YT, Dimitrova YN, Schneider G, Ridenour WB, Bhattacharya S, Soss SE, Caprioli RM, Filipek A, Chazin WJ. Structure of the S100A6 complex with a fragment from the C-terminal domain of Siah-1 interacting protein: a novel mode for S100 protein target recognition. *Biochemistry*. 2008. 47, 10921-10932.
- Lees JP, Manlandro CM, Picton LK, Tan AZ, Casares S, Flanagan JM, Fleming KG, Hill RB. A designed point mutant in Fis1 disrupts dimerization and mitochondrial fission. *J Mol Biol*. 2012. 423, 143-158.
- Leipe DD, Koonin EV, Aravind L. STAND, a class of P-loop NTPases including animal and plant regulators of programmed cell death: multiple, complex domain architectures, unusual phyletic patterns, and evolution by horizontal gene transfer. *J Mol Biol*. 2004. 343, 1-28.
- Leister RT, Dahlbeck D, Day B, Li Y, Chesnokova O, Staskawicz BJ. Molecular genetic evidence for the role of SGT1 in the intramolecular complementation of Bs2 protein activity in *Nicotiana benthamiana*. *Plant Cell*. 2005. 17, 1268-1278.
- Levitt MH. *Spin Dynamics: Basics of Nuclear Magnetic Resonance*. Second edition. John Wiley and Sons. 2008.
- Li J, Richter K, Buchner J. Mixed Hsp90-cochaperone complexes are important for the progression of the reaction cycle. *Nat Struct Mol Biol*. 2011. 18, 61-66.
- Li J, Soroka J, Buchner J. The Hsp90 chaperone machinery: conformational dynamics and regulation by co-chaperones. *Biochim Biophys Acta*. 2012. 1823, 624-635.

- Li J, Buchner J. Structure, function and regulation of the hsp90 machinery. *Biomed J.* 2013a. 36, 106-117.
- Li Y, Li S, Bi D, Cheng YT, Li X, Zhang Y. SRFR1 negatively regulates plant NB-LRR resistance protein accumulation to prevent autoimmunity. *PLoS Pathog.* 2010. 6:e1001111.
- Li Z, Hartl FU, Bracher A. Structure and function of Hip, an attenuator of the Hsp70 chaperone cycle. *Nat Struct Mol Biol.* 2013b. 20, 929-935.
- Liu H, Morris RJ, Hexemer A, Grandison S, Zwart PH. Computation of small-angle scattering profiles with three-dimensional Zernike polynomials. *Acta Crystallogr A.* 2012a 68, 278-285.
- Liu H, Hexemer A, Zwart PH. The Small Angle Scattering ToolBox (SASTBX): an open-source software for biomolecular small-angle scattering. *Acta Crystallogr A.* 2012b 45, 587-593.
- Liu Y, Schiff M, Marathe R, Dinesh-Kumar SP. Tobacco Rar1, EDS1 and NPR1/NIM1 like genes are required for N-mediated resistance to tobacco mosaic virus. *Plant J.* 2002a. 30, 415-429.
- Liu Y, Schiff M, Serino G, Deng XW, Dinesh-Kumar SP. Role of SCF ubiquitin-ligase and the COP9 signalosome in the N gene-mediated resistance response to Tobacco mosaic virus. *Plant Cell.* 2002b. 14, 1483-1496.
- Liu Y, Burch-Smith T, Schiff M, Feng S, Dinesh-Kumar SP. Molecular chaperone Hsp90 associates with resistance protein N and its signaling proteins SGT1 and Rar1 to modulate an innate immune response in plants. *J Biol Chem.* 2004. 279, 2101-2108.
- Liwo A, Baranowski M, Czaplowski C, Gołaś E, He Y, Jagieła D, Krupa P, Maciejczyk M, Makowski M, Mozolewska MA, Niadzedtski A, Ołdziej S, Scheraga HA, Sieradzan AK, Slusarz R, Wirecki T, Yin Y, Zaborowski B. A unified coarse-grained model of biological macromolecules based on mean-field multipole-multipole interactions. *J Mol Model.* 2014, 20, 2306.
- Lorenz OR, Freiburger L, Rutz DA, Krause M, Zierer BK, Alvira S, Cuéllar J, Valpuesta JM, Madl T, Sattler M, Buchner J. Modulation of the Hsp90 chaperone cycle by a stringent client protein. *Mol Cell.* 2014. 53, 941-953.
- Ludwig AA, Saitoh H, Felix G, Freymark G, Miersch O, Wasternack C, Boller T, Jones JD, Romeis T. Ethylene-mediated cross-talk between calcium-dependent protein kinase and MAPK signaling controls stress responses in plants. *Proc Natl Acad Sci U S A.* 2005. 102, 10736-10741.
- Luna E, Pastor V, Robert J, Flors V, Mauch-Mani B, Ton J. Callose deposition: a multifaceted plant defense response. *Mol Plant Microbe Interact.* 2011. 24, 183-193.

M

- Macho AP, Zipfel C. Plant PRRs and the activation of innate immune signaling. *Mol Cell.* 2014. 54, 263-72.
- Maekawa T, Cheng W, Spiridon LN, Töller A, Lukasik E, Saijo Y, Liu P, Shen QH, Micluta MA, Somssich IE, Takken FL, Petrescu AJ, Chai J, Schulze-Lefert P. Coiled-coil domain-dependent homodimerization of intracellular barley immune receptors defines a minimal functional module for triggering cell death. *Cell Host Microbe.* 2011. 9, 187-199.
- Main ER, Xiong Y, Cocco MJ, D'Andrea L, Regan L. Design of stable alpha-helical arrays from an idealized TPR motif. *Structure.* 2003. 11, 497-508.
- Mak AN, Bradley P, Cernadas RA, Bogdanove AJ, Stoddard BL. The crystal structure of TAL effector PthXo1 bound to its DNA target. *Science.* 2012. 335, 716-719.
- Manavalan P, Johnson WC. Variable selection method improves the prediction of protein secondary structure from circular dichroism spectra. *Anal. Biochem.* 1987. 167, 76-85.
- Marcu MG, Chadli A, Bouhouche I, Catelli M, Neckers LM. The heat shock protein 90 antagonist novobiocin interacts with a previously unrecognized ATP-binding domain in the carboxyl terminus of the chaperone. *J Biol Chem.* 2000. 275, 37181-37186.
- Martin GS. The hunting of the Src. *Nat Rev Mol Cell Biol.* 2001, 2, 467-475.
- Matsunaga T, Okamura Y. Genes and proteins involved in bacterial magnetic particle formation. *Trends Microbiol.* 2003. 11, 536-541.

- Mayor A, Martinon F, De Smedt T, Pétrilli V, Tschopp J. A crucial function of SGT1 and HSP90 in inflammasome activity links mammalian and plant innate immune responses. *Nat Immunol.* 2007. 8, 497-503.
- McHale L, Tan X, Koehl P, Michelmore RW. Plant NBS-LRR proteins: adaptable guards. *Genome Biol.* 2006.7, 212.
- McLaughlin SH, Sobott F, Yao ZP, Zhang W, Nielsen PR, Grossmann JG, Laue ED, Robinson CV, Jackson SE. The co-chaperone p23 arrests the Hsp90 ATPase cycle to trap client proteins. *J Mol Biol.* 2006. 356, 746-758.
- Melotto M, Underwood W, Koczan J, Nomura K, He SY. Plant stomata function in innate immunity against bacterial invasion. *Cell.* 2006. 126, 969-980.
- Mendonça YM, Aparicio R, Ramos CHI. Structural analysis of the sweet orange Hsp90 by SAXS. LNLS 2009 Activity Report. (<http://lnls.cnpem.br/ar2009/PDF/1597.pdf>)
- Mertens HD, Svergun DI. Structural characterization of proteins and complexes using small-angle X-ray solution scattering. *J Struct Biol.* 2010, 172, 128-141.
- Mestre P, Baulcombe DC. Elicitor-mediated oligomerization of the tobacco N disease resistance protein. *Plant Cell.* 2006. 18, 491-501.
- Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics* 1953, 21, 1087.
- Meyers BC, Kozik A, Griego A, Kuang H, Michelmore RW. Genome-wide analysis of NBS-LRR-encoding genes in Arabidopsis. *Plant Cell.* 2003. 15, 809-834.
- Michowski W, Ferretti R, Wisniewska MB, Ambrozkiwicz M, Beresewicz M, Fusella F, Skibinska-Kijek A, Zablocka B, Brancaccio M, Tarone G, Kuznicki J. Morgana/CHP-1 is a novel chaperone able to protect cells from stress. *Biochim Biophys Acta.* 2010. 1803, 1043-1049.
- Miller JC, Tan S, Qiao G, Barlow KA, Wang J, Xia DF, Meng X, Paschon DE, Leung E, Hinkley SJ, Dulay GP, Hua KL, Ankoudinova I, Cost GJ, Urnov FD, Zhang HS, Holmes MC, Zhang L, Gregory PD, Rebar EJ. A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol.* 2011. 29, 143-148.
- Minh DD, Makowski L. Wide-angle X-ray solution scattering for protein-ligand binding: multivariate curve resolution with Bayesian confidence intervals. *Biophys J.* 2013, 104, 873-883.
- Murata S, Minami Y, Minami M, Chiba T, Tanaka K. CHIP is a chaperone-dependent E3 ligase that ubiquitylates unfolded protein. *EMBO Rep.* 2001. 2, 1133-1138.
- Muskett PR, Kahn K, Austin MJ, Moisan LJ, Sadanandom A, Shirasu K, Jones JD, Parker JE. Arabidopsis RAR1 exerts rate-limiting control of R gene-mediated defenses against multiple pathogens. *Plant Cell.* 2002. 14, 979-992.

N

- Navarro L, Dunoyer P, Jay F, Arnold B, Dharmasiri N, Estelle M, Voinnet O, Jones JD. A plant miRNA contributes to antibacterial resistance by repressing auxin signaling. *Science.* 2006. 312, 436-439.
- Nemoto T, Sato N. Oligomeric forms of the 90-kDa heat shock protein. *Biochem J.* 1998. 330, 989-995.
- Noël LD, Cagna G, Stuttmann J, Wirthmüller L, Betsuyaku S, Witte CP, Bhat R, Pochon N, Colby T, Parker JE. Interaction between SGT1 and cytosolic/nuclear HSC70 chaperones regulates Arabidopsis immune responses. *Plant Cell.* 2007. 19, 4061-4076.
- Nowotny M, Bhattacharya S, Filipek A, Krezel AM, Chazin W, Kuznicki J. Characterization of the interaction of calcyclin (S100A6) and calcyclin-binding protein. *J Biol Chem.* 2000. 275, 31178-31182.
- Novotni M, Klein R. Proceedings of the Eighth ACM Symposium on Solid Modeling and Applications. 2003 New York: ACM.
- Nyarko A, Mosbahi K, Rowe AJ, Leech A, Boter M, Shirasu K, Kleanthous C. TPR-Mediated self-association of plant SGT1. *Biochemistry.* 2007. 46, 11331-11341.

P

- Padmanabhan MS, Ma S, Burch-Smith TM, Czymbek K, Huijser P, Dinesh-Kumar SP. Novel positive regulatory role for the SPL6 transcription factor in the N TIR-NB-LRR receptor-mediated plant innate immunity. *PLoS Pathog.* 2013. 9, e1003235.
- Parson WW. *Modern Optical Spectroscopy.* 2009. Springer-Verlag.
- Pearl LH, Prodromou C. Structure and in vivo function of Hsp90. *Curr Opin Struct Biol.* 2000. 10, 46-51.
- Perutz MF, Rossmann MG, Cullis AF, Muirhead H, Will G, North AC. Structure of haemoglobin: a three-dimensional Fourier synthesis at 5.5-Å resolution, obtained by X-ray analysis. *Nature.* 1960, 185, 416-422.
- Peters JM. The anaphase promoting complex/cyclosome: a machine designed to destroy. *Nat Rev Mol Cell Biol.* 2006. 7, 644-656.
- Petoukhov MV, Svergun DI. Global rigid body modelling of macromolecular complexes against small-angle scattering data. *Biophys. J.* 2005, 89, 1237-1250.
- Porod G. Die Röntgenkleinwinkelstreuung von dichtgepackten kolloiden Systemen. *Kolloid Z. Z. Polym.* 1951. 124, 83-114.
- Prodromou C, Roe SM, O'Brien R, Ladbury JE, Piper PW, Pearl LH. Identification and structural characterization of the ATP/ADP-binding site in the Hsp90 molecular chaperone. *Cell.* 1997. 90, 65-75.

R

- Ratajczak T, Ward BK, Minchin RF. Immunophilin chaperones in steroid receptor signalling. *Curr Top Med Chem.* 2003. 3, 1348-1357.
- Retzlaff M, Hagn F, Mitschke L, Hessling M, Gugel F, Kessler H, Richter K, Buchner J. Asymmetric activation of the hsp90 dimer by its cochaperone aha1. *Mol Cell.* 2010. 37, 344-354.
- Richmond TJ, Finch JT, Rushton B, Rhodes D, Klug A. Structure of the nucleosome core particle at 7 Å resolution. *Nature.* 1984, 311, 532-537.
- Riedl SJ, Salvesen GS. The apoptosome: signalling platform of cell death. *Nat Rev Mol Cell Biol.* 2007. 8, 405-413.
- Robertson MP, Joyce GF. The origins of the RNA world. *Cold Spring Harb Perspect Biol.* 2012, 4:a003608.
- Ron M, Avni A. The receptor for the fungal elicitor ethylene-inducing xylanase is a member of a resistance-like gene family in tomato. *Plant Cell.* 2004. 16, 1604-1615.
- Rosenfeld L, *Z. Physik* 1928. 52, 161.
- Röhl A, Rohrberg J, Buchner J. The chaperone Hsp90: changing partners for demanding clients. *Trends Biochem Sci.* 2013. 38, 253-262.
- Różycki B, Kim YC, Hummer G. SAXS ensemble refinement of ESCRT-III CHMP3 conformational transitions. *Structure.* 2011, 19, 109-116.

Q

- Qi D, Innes RW. Recent Advances in Plant NLR Structure, Function, Localization, and Signaling. *Front Immunol.* 2013. 4, 48.

S

- Sampathkumar P, Roach C, Michels PA, Hol WG. Structural insights into the recognition of peroxisomal targeting signal 1 by *Trypanosoma brucei* peroxin 5. *J. Mol Biol.* 2008. 381, 867-880.
- Sbroggiò M, Ferretti R, Percivalle E, Gutkowska M, Zylicz A, Michowski W, Kuznicki J, Ac-

- cornero F, Pacchioni B, Lanfranchi G, Hamm J, Turco E, Silengo L, Tarone G, Brancaccio M. The mammalian CHORD-containing protein melusin is a stress response protein interacting with Hsp90 and Sgt1. *FEBS Lett.* 2008. 582, 1788-1794.
- Scheufler C, Brinker A, Bourenkov G, Pegoraro S, Moroder L, Bartunik H, Hartl FU, Moarefi I. Structure of TPR domain-peptide complexes: critical elements in the assembly of the Hsp70-Hsp90 multichaperone machine. *Cell.* 2000. 101, 199-210.
 - Schleiff E, Becker T. Common ground for protein translocation: access control for mitochondria and chloroplasts. *Nat Rev Mol Cell Biol.* 2011. 12, 48-59.
 - Schwieters CD, Clore GM. Using small angle solution scattering data in Xplor-NIH structure calculations. *Prog Nucl Magn Reson Spectrosc.* 2014. 80, 1-11.
 - Seo YS, Lee SK, Song MY, Suh JP, Hahn TR, Roland P, Jeon JS. The Hsp90-Sgt1-Rar1 molecular chaperone complex: a core modulator in plant immunity. *Journal of Plant Biol.* 2008. 51, 1-10.
 - Shen QH, Saijo Y, Mauch S, Biskup C, Bieri S, Keller B, Seki H, Ulker B, Somssich IE, Schulze-Lefert P. Nuclear activity of MLA immune receptors links isolate-specific and basal disease-resistance responses. *Science.* 2007, 315, 1098-1103.
 - Shiau AK, Harris SF, Southworth DR, Agard DA. Structural Analysis of *E. coli* hsp90 reveals dramatic nucleotide-dependent conformational rearrangements. *Cell.* 2006. 127, 329-340.
 - Shibuya N, Minami E. Oligosaccharide signalling for defence responses in plant. *Physiological and Molecular Plant Pathology.* 2001. 59, 223-233.
 - Shirasu K, Lahaye T, Tan MW, Zhou F, Azevedo C, Schulze-Lefert P. A novel class of eukaryotic zinc-binding proteins is required for disease resistance signaling in barley and development in *C. elegans*. *Cell.* 1999. 99, 355-366.
 - Shirasu K. The HSP90-SGT1 chaperone complex for NLR immune sensors. *Annu Rev Plant Biol.* 2009. 60, 139-164.
 - Sobkowiak L, Szarzynska B, Szweykowska Kulińska Z. Plant microRNA biogenesis. *Postepy Biochem.* 2008. 54, 308-316.
 - Southworth DR, Agard DA. Species-dependent ensembles of conserved conformational states define the Hsp90 chaperone ATPase cycle. *Mol Cell.* 2008. 32, 631-640.
 - Southworth DR, Agard DA. Client-loading conformation of the Hsp90 molecular chaperone revealed in the cryo-EM structure of the human Hsp90:Hop complex. *Mol Cell.* 2011. 42, 771-781.
 - Söti C, Pál C, Papp B, Csermely P. Molecular chaperones as regulatory elements of cellular networks. *Curr Opin Cell Biol.* 2005. 17, 210-215.
 - Söti C, Rácz A, Csermely P. A Nucleotide-dependent molecular switch controls ATP binding at the C-terminal domain of Hsp90. N-terminal nucleotide binding unmasks a C-terminal binding pocket. *J Biol Chem.* 2002. 277, 7066-7075.
 - Sreerama N, Woody RW. A self-consistent method for the analysis of protein secondary structure from circular dichroism. *Anal Biochem.* 1993. 209, 32-44.
 - Sreerama N, Woody RW. Computation and analysis of protein circular dichroism spectra. *Methods Enzymol.* 2004. 383, 318-351.
 - Stanley WA, Filipp FV, Kursula P, Schüller N, Erdmann R, Schliebs W, Sattler M, Wilmanns M. Recognition of a functional peroxisome type 1 target by the dynamic import receptor pex5p. *Mol Cell.* 2006. 24, 653-663.
 - Stein M, Dittgen J, Sánchez-Rodríguez C, Hou BH, Molina A, Schulze-Lefert P, Lipka V, Somerville S. Arabidopsis PEN3/PDR8, an ATP binding cassette transporter, contributes to nonhost resistance to inappropriate pathogens that enter by direct penetration. *Plant Cell.* 2006, 731-46.
 - Street TO, Krukenberg KA, Rosgen J, Bolen DW, Agard DA. Osmolyte-induced conformational changes in the Hsp90 molecular chaperone. *Protein Sci.* 2010. 19, 57-65.
 - Street TO, Lavery LA, Agard DA. Substrate binding drives large-scale conformational changes in the Hsp90 molecular chaperone. *Mol Cell.* 2011. 42, 96-105.

- Stuhrmann HB. Interpretation of small-angle scattering functions of dilute solutions and gases. A representation of the structures related to a one-particle scattering function. *Acta Cryst. A* 1970, 26, 297-306.
- Stuttmann J, Parker JE, Noël LD. Staying in the fold: The SGT1/chaperone machinery in maintenance and evolution of leucine-rich repeat proteins. *Plant Signal Behav.* 2008. 3,283-285.
- Svergun DI. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* 1992. 25, 495-503.
- Svergun DI, Barberato C, Koch MHJ. CRY SOL - a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates *J. Appl. Cryst.* 1995, 28, 768-773.
- Svergun DI, Volkov VV, Kozin MB, Stuhrmann HB. New developments in direct shape determination from small-angle scattering. 2. Uniqueness. *Acta Crystallogr. A.* 1996. 52, 419-426.
- Svergun DI, Volkov VV, Kozin MB, Stuhrmann HB, Barberato C, Koch MHJ. Shape Determination from Solution Scattering of Biopolymers. *J. Appl. Crystallogr.* 1997, 30, 798-802.
- Svergun DI. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys J.* 1999, 76, 2879-2886.
- Svergun DI, Nierhaus KH. A map of protein-rRNA distribution in the 70 S Escherichia coli ribosome. *J Biol Chem.* 2000. 275, 14432-14439.
- Svergun DI, Petoukhov MV, Koch MHJ. Determination of domain structure of proteins from X-ray solution scattering. *Biophys. J.* 2001, 80, 2946-2953.

T

- Taipale M, Jarosz DF, Lindquist S. HSP90 at the hub of protein homeostasis: emerging mechanistic insights. *Nat Rev Mol Cell Biol.* 2010. 11, 515-528.
- Takahashi A, Casais C, Ichimura K, Shirasu K. HSP90 interacts with RAR1 and SGT1 and is essential for RPS2-mediated disease resistance in Arabidopsis. *Proc Natl Acad Sci U S A.* 2003. 100, 11777-11782.
- Takken FL, Goverse A. How to build a pathogen detector: structural basis of NB-LRR function. *Curr Opin Plant Biol.* 2012. 15, 375-384.
- Taoka A, Asada R, Sasaki H, Anzawa K, Wu LF, Fukumori Y. Spatial localizations of Mam22 and Mam12 in the magnetosomes of *Magnetospirillum magnetotacticum*. *J Bacteriol.* 2006. 188, 3805-3812.
- Tauler 2005. http://www.cid.csic.es/homes/rtaqam/tmp/WEB_MCR/download/pdf/MCR.2005.pdf
- Thao NP, Chen L, Nakashima A, Hara S, Umemura K, Takahashi A, Shirasu K, Kawasaki T, Shimamoto K. RAR1 and HSP90 form a complex with Rac/Rop GTPase and function in innate-immune responses in rice. *Plant Cell.* 2007. 19, 4035-4045.
- Thomashow LS, Reeves S, Thomashow MF. Crown gall oncogenesis: evidence that a T-DNA gene from the *Agrobacterium* Ti plasmid pTiA6 encodes an enzyme that catalyzes synthesis of indoleacetic acid. *Proc Natl Acad Sci U S A.* 1984. 81, 5071-5075.
- Tinoco I, Woody RW, Bradley DF. Absorption and rotation of light by helical polymers: the effect of chain length. *J Chem Phys* 1963. 38, 1317-1325.

U

- Ueeda Y, Taketomi H, Go N: Studies on protein folding, unfolding and fluctuations by computer simulation. A three-dimensional lattice model of lysozyme. *Biopolymers* 1978. 17, 1531-1548.
- Umemoto N, Kakitani M, Iwamatsu A, Yoshikawa M, Yamaoka N, Ishida I. The structure and function of a soybean beta-glucan-elicitor-binding protein. *Proc Natl Acad Sci U S A.* 1997. 94, 1029-1034.

W

- Wang L, Liu YT, Hao R, Chen L, Chang Z, Wang HR, Wang ZX, Wu JW. Molecular mechanism of the negative regulation of Smad1/5 protein by carboxyl terminus of Hsc70-interacting protein (CHIP). *J Biol Chem.* 2011. 286, 15883-15894.
- Wang Y, Li J, Hou S, Wang X, Li Y, Ren D, Chen S, Tang X, Zhou JM. A *Pseudomonas syringae* ADP-ribosyltransferase inhibits Arabidopsis mitogen-activated protein kinase kinases. *Plant Cell.* 2010a. 22, 2033-2044.
- Wang YX, Zuo X, Wang J, Yu P, Butcher SE. Rapid global structure determination of large RNA and RNA complexes using NMR and small-angle X-ray scattering. *Methods.* 2010b, 52, 180-191.
- Watson JD, Crick FHC. A Structure for Deoxyribose Nucleic Acid. *Nature.* 1953, 171, 737–738.
- Weaver AJ, Sullivan WP, Felts SJ, Owen BA, Toft DO. Crystal structure and activity of human p23, a heat shock protein 90 co-chaperone. *J Biol Chem.* 2000. 275, 23045-23052.
- Whitesell L, Lindquist SL. HSP90 and the chaperoning of cancer. *Nat Rev Cancer.* 2005. 5, 61-772.
- Whitmore L, Wallace BA. DICHROWEB: an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Research.* 2004. 32, W668-673.
- Wiedemann H. Synchrotron Radiation. Springer-Verlag. Berlin Heidelberg. 2003.
- Williamson TE, Craig BA, Kondrashkina E, Bailey-Kellogg C, Friedman AM. Analysis of self-associating proteins by singular value decomposition of solution scattering data. *Biophys J.* 2008. 94, 4906-4923.
- Woody RW, Tinoco I. Optical rotation of oriented helices. III. Calculation of the rotatory dispersion and circular dichroism of the alpha and 310-helix. *J Chem Phys.* 1967. 46, 4927–4945.
- Wu J, Luo S, Jiang H, Li H. Mammalian CHORD-containing protein 1 is a novel heat shock protein 90-interacting protein. *FEBS Lett.* 2005. 579, 421-426.

V

- Van der Biezen EA, Jones JD. The NB-ARC domain: a novel signalling motif shared by plant resistance gene products and regulators of cell death in animals. *Curr Biol.* 1998. 8, 226-227.
- Van Etten HD, Mansfield JW, Bailey JA, Farmer EE. Two Classes of Plant Antibiotics: Phytoalexins versus “Phytoanticipins”. *Plant Cell.* 1994. 6, 1191-1192.
- Van Stokkum IHM, Spoelder HJW, Bloemendal M, Van Grondelle R, Groen FCA. Estimation of protein secondary structure and error analysis from CD spectra. *Anal. Biochem.* 1990. 191, 110-118.

X

- Xiang T, Zong N, Zou Y, Wu Y, Zhang J, Xing W, Li Y, Tang X, Zhu L, Chai J, Zhou JM. *Pseudomonas syringae* effector AvrPto blocks innate immunity by targeting receptor kinases. *Curr Biol.* 2008. 18, 74-80.
- Xu D, Zhang Y., Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field. *Proteins.* 2012, 80, 1715-1735.

Y

- Yang S, Park S, Makowski L, Roux B. A rapid coarse residue-based computational method for x-ray solution scattering characterization of protein folds and multiple conformational states of large protein complexes. *Biophys J.* 2009, 96, 4449-4463.
- Yang S, Blachowicz L, Makowski L, Roux B. Multidomain assembled states of Hck tyrosine kinase in solution. *Proc Natl Acad Sci U S A.* 2010, 107, 15757-15762.

Z

- Zabka M, Leśniak W, Prus W, Kuźnicki J, Filipek A. Sgt1 has co-chaperone properties and is up-regulated by heat shock. *Biochem Biophys Res Commun.* 2008. 370, 179-183.
- Zeytuni N, Ozyamak E, Ben-Harush K, Davidov G, Levin M, Gat Y, Moyal T, Brik A, Komeili A, Zarivach R. Self-recognition mechanism of MamA, a magnetosome-associated TPR-containing protein, promotes complex assembly. *Proc Natl Acad Sci U S A.* 2011. 108, E480-487.
- Zeytuni N, Zarivach R. Structural and functional discussion of the tetra-trico-peptide repeat, a protein interaction module. *Structure.* 2012. 20, 397-405.
- Zhang Z, Chang L, Yang J, Conin N, Kulkarni K, Barford D. The four canonical tpr subunits of human APC/C form related homo-dimeric structures and stack in parallel to form a TPR suprahelix. *J Mol Biol.* 2013. 425, 4236-4248.
- Zhang M, Botër M, Li K, Kadota Y, Panaretou B, Prodromou C, Shirasu K, Pearl LH. Structural and functional coupling of Hsp90- and Sgt1-centred multi-protein complexes. *EMBO J.* 2008. 27, 2789-2798.
- Zhang M, Kadota Y, Prodromou C, Shirasu K, Pearl LH. Structural basis for assembly of Hsp90-Sgt1-CHORD protein complexes: implications for chaperoning of NLR innate immunity receptors. *Mol Cell.* 2010a. 39, 269-281.
- Zhang Z, Kulkarni K, Hanrahan SJ, Thompson AJ, and Barford D. The APC/C subunit Cdc16/Cut9 is a contiguous tetratricopeptide repeat superhelix with a homo-dimer interface similar to Cdc27. *EMBO J.* 2010b. 29, 3733-3744.
- Zipfel C. Pattern-recognition receptors in plant innate immunity. *Curr Opin Immunol.* 2008. 20, 10-16.
- Zipfel C, Robatzek S, Navarro L, Oakeley EJ, Jones JD, Felix G, Boller T. Bacterial disease resistance in Arabidopsis through flagellin perception. *Nature.* 2004. 428, 764-767.
- Zuo X, Wang J, Foster TR, Schwieters CD, Tiede DM, Butcher SE, Wang YX. Global molecular structure and interfaces: refining an RNA:RNA complex structure using solution X-ray scattering data. *J Am Chem Soc.* 2008, 130, 3292-3293.