

## **Autoreferat**

# **Metodyczne aspekty wysokorozdzielczej krystalografii makromolekuł**

w związku z wnioskiem o przeprowadzenie przewodu habilitacyjnego

**dr Mirosław Gilski**  
**Wydział Chemii**  
**Uniwersytet im. A. Mickiewicza**



**w Poznaniu**

**Poznań, 2016r.**

1. *Imię i Nazwisko:* **Mirosław Gilski**

2. *Posiadane dyplomy, stopnie naukowe/ artystyczne – z podaniem nazwy, miejsca i roku ich uzyskania oraz tytułu rozprawy doktorskiej.*

Doktor nauk chemicznych	2001r.	Uniwersytet im. Adama Mickiewicza w Poznaniu Wydział Chemii Rozprawa doktorska pt. „Struktura molekularna i właściwości asocjacyjne kationu cytydyniowego i deoksycytydyniowego w kryształach wybranych soli. Badania rentgenograficzne” Promotor: prof. dr hab. Mariusz Jaskólski
Magister fizyki	1987r.	Uniwersytet im. Adama Mickiewicza w Poznaniu Wydział Matematyki i Fizyki Praca magisterska pt. „Analiza konformacji cząsteczek organicznych metodą minimalizacji energii potencjalnej” Promotor: prof. dr hab. Jerzy Pietrzak

3. *Informacje o dotychczasowym zatrudnieniu w jednostkach naukowych.*

Pracownik inżyniero-techniczny	1987-1994r.	Uniwersytet im. Adama Mickiewicza w Poznaniu Wydział Chemii
Asystent	1994-2001r.	Uniwersytet im. Adama Mickiewicza w Poznaniu Wydział Chemii
Starszy specjalista (½ etatu)	od 1994r.	Instytut Chemii Bioorganicznej PAN Centrum Badań Biokrytalograficznych
Adiunkt	od 2002r.	Uniwersytet im. Adama Mickiewicza w Poznaniu Wydział Chemii

4. *Wskazanie osiągnięć stanowiących podstawę postępowania habilitacyjnego*

Osiągnięciami naukowymi wynikającymi z art. 16 ust. 2 ustawy z dnia 14 marca 2003 o stopniach naukowych i tytułach naukowych oraz o stopniach naukowych i tytułach naukowych w zakresie sztuki (Dz. U. Nr 65, poz. 595 ze zm.) jest cykl powiązanych tematycznie publikacji naukowych pt.:

**„Metodyczne aspekty wysokorozdzielczej krystalografii makromolekuł”**

4.1 Wykaz powiżanych tematycznie artykułów naukowych **H1-H11** stanowiących podstawę postępowania habilitacyjnego

Sumaryczny Impact Factor: **47.907**

Liczba cytowa : **204**

Liczba prac: **11**

---

**H1. Gilski, M.** (2008).

Automation and remote synchrotron data collection.

*Acta Phys. Pol.* **A114**, 331-338.

*Mój wkład w powstanie tej pracy polegał na: przygotowaniu koncepcji pracy, opracowaniu protokołów zdalnej rejestracji danych synchrotronowych, przygotowaniu i przeprowadzeniu kilku zdalnych sesji pomiarowych w ośrodku synchrotronowym ESRF w Grenoble, przeprowadzeniu testów systemu EDNA, procesowaniu zarejestrowanych danych dyfrakcyjnych oraz przygotowaniu publikacji.*

*Mój udział procentowy: 100%*

IF(2008)=**0.321**, IF5=0.497, liczba cytowa : **2**

---

**H2. Gilski, M.** (2012).

Data Processing Programs for Analysis of Diffraction Images of Macromolecular Crystals Recorded using Synchrotron Radiation.

*Acta Phys. Pol.* **A121**, 871-875.

*Mój wkład w powstanie tej pracy polegał na: opracowaniu koncepcji pracy, selekcji zestawów dyfrakcyjnych danych synchrotronowych odpowiednich do testów, opracowanie dedykowanych protokołów obróbki dla poszczególnych zestawów danych, przetwarzanie surowych danych za pomocą różnych programów i algorytmów, analizie otrzymanych wyników oraz przygotowaniu publikacji.*

*Mój udział procentowy: 100%*

IF(2012)=**0.531**, IF5=0.497, liczba cytowa : **1**

---

**H3. Khatib, F., Dimaio, F., Cooper, S., Kazmierczyk, M., Gilski, M., Krzywda, S., Zabranska, H., Pichova, I., Thompson, J., Popovic, Z., Jaskolski, M. & Baker, D.** (2011).

Crystal structure of a monomeric retroviral protease solved by protein folding game players.

*Nature Struct. Mol. Biol.*, **18**(10), 1175-1177.

*Mój wkład w powstanie tej pracy polegał na: testowaniu różnych tradycyjnych metod rozwiązania struktury, opracowaniu niestandardowych protokołów dla metody podstawienia cząsteczkowego, wyborze i optymalnej obróbce zestawu obrazów dyfrakcyjnych, opracowaniu strategii udokładniania oraz udziale w udokładnianiu modelu struktury, analizie wyników i przygotowaniu publikacji.*

*Mój udział procentowy szacuję na: 25%*

IF(2011)=**12.712**, IF5=12.479, liczba cytowa : **134**

---

**H4. Gilski, M., Kazmierczyk, M., Krzywda, S., Zabranska, H., Cooper, S., Popovic, Z., Khatib, F., DiMaio, F., Thompson, J., Baker, D., Pichova, I. & Jaskolski, M.** (2011).

High-resolution structure of a retroviral protease folded as a monomer.

*Acta Cryst.* **D67**, 907-914, IF5=9.574, cyt. 5

*Mój wkład w powstanie tej pracy polegał na: wyborze i optymalnej obróbce zestawu obrazów dyfrakcyjnych, manualnej przebudowie wstępnego modelu struktury, opracowaniu protokołu i strategii udokładniania, obliczeniach krystalograficznych oraz udziale w udokładnianiu modelu struktury, analizie wyników i przygotowaniu publikacji.*

*Mój udział procentowy szacuję na: 40%*

IF(2011)=**12.619**, IF5=9.585, liczba cytowa : **5**

---

**H5.** Wierzbicki, M., **Gilski, M.**, Rissanen, K. Jaskolski, M. & Szumna, M. (2014).  
Experiences with applications of macromolecular tools in supramolecular crystallography.  
*CrystEngComm*. **16**, 3773–3780.

*Mój wkład w powstanie tej pracy polegał na: opracowaniu strategii rozwinięcia struktury supramolekularnej metod podstawienia cz. stezkowego, opracowaniu niestandardowych protokołów użycy w metodzie podstawienia cz. stezkowego i procesie udokładniania, przygotowanie skryptów do konwersji danych, rozwinięciu struktury i manualnej przebudowie wstępnego modelu, wykonaniu i porównaniu efektywności metody podstawienia cz. stezkowego dla różnych modeli, udział w analizie wyników i przygotowaniu publikacji.*

*Mój udział procentowy szacuję na: 50%*  
IF(2014)=**4.034**, IF5=4.022, liczba cytowa : **2**

---

**H6.** Drozdal, P., **Gilski, M.** Kierzek, R. & Jaskolski, M. (2013).  
Ultrahigh-resolution crystal structures of Z-DNA in complex with Mn<sup>2+</sup> and Zn<sup>2+</sup> ions.  
*Acta Cryst.* **D69**, 1180-1190.

*Mój wkład w powstanie tej pracy polegał na: rejestracji synchrotronowych danych dyfrakcyjnych, optymalnej obróbce danych dyfrakcyjnych, opracowaniu strategii i wysokorozdzielczych protokołów udokładniania programem SHELXL, przygotowaniu zestawu poprawnych wartości geometrycznych dla kwasów nukleinowych, nadzorowaniu procesu udokładniania i walidacji modelu, udział w analizie wyników i przygotowaniu publikacji.*

*Mój udział procentowy szacuję na: 50%*  
IF(2013)=**7.232**, IF5=9.585, liczba cytowa : **3**

---

**H7.** Drozdal, P., **Gilski, M.**, Kierzek, R., Lomozik, L. & Jaskolski, M., (2015).  
High-resolution crystal structure of Z-DNA in complex with Cr<sup>3+</sup> cations.  
*J. Biol. Inorg. Chem.* **20**, 595–602.

*Mój wkład w powstanie tej pracy polegał na: rejestracji synchrotronowych danych dyfrakcyjnych, optymalnej obróbce obrazów dyfrakcyjnych, opracowaniu strategii i wysokorozdzielczych protokołów udokładniania programem REFMAC5, przygotowaniu zestawu poprawnych wartości geometrycznych dla kwasów nukleinowych, korekcie błędnych wartości w bibliotece stereochemicznej programu REFMAC5, nadzorowaniu procesu udokładniania i walidacji modelu, udział w analizie wyników i przygotowaniu publikacji.*

*Mój udział procentowy szacuję na: 50%*  
IF(2014)=**2.538**, IF5=2.894, liczba cytowa : **1**

---

**H8.** **Gilski, M.**, Drozdal, P., Kierzek, R. & Jaskolski, M. (2016).  
Atomic-resolution structure of a chimeric DNA-RNA Z-type duplex in complex with Ba<sup>2+</sup> ions:  
a case of complicated multi-domain twinning.  
*Acta Cryst.* **D72**, 211-223.

*Mój wkład w powstanie tej pracy polegał na: rejestracji synchrotronowych danych dyfrakcyjnych, obróbce wszystkich zestawów surowych danych dyfrakcyjnych, opracowaniu optymalnych protokołów detekcji zbliżonego niaczenia, pseudosymetrii oraz procedur do jednoznacznego określenia grupy przestrzennej, przygotowaniu dedykowanych skryptów do transformacji współrzędnych i skalowania refleksów między różnymi układami krystalograficznymi i grupami przestrzennymi, opracowaniu wysokorozdzielczych strategii i protokołów udokładniania programami SHELXL i REFMAC5 dla skomplikowanych przypadków ze zbliżonym niaczeniem i pseudosymetrią, przygotowaniu zestawu poprawnych wartości geometrycznych dla kwasów nukleinowych, korekcie błędnych wartości w bibliotece stereochemicznej programu REFMAC5, nadzorowaniu wszystkich etapów procesu udokładniania i walidacji modelu, udział w analizie wyników i przygotowaniu publikacji.*

*Mój udział procentowy szacuję na: 80%*  
IF(2014)=**2.680**, IF5=9.585, liczba cytowa : **0**

---

**H9.** Jaskolski, M., **Gilski, M.**, Dauter, Z. & Włodawer, A. (2007).  
Stereochemical restraints revisited: how accurate are refinement targets and how much should protein structures be allowed to deviate from them?  
*Acta Cryst.* **D63**, 611-620.

*Mój wkład w powstanie tej pracy polegał na: udziale w zainicjowaniu projektu, analizie statystycznej parametrów*

*geometrycznych aminokwasów na podstawie strukturalnych baz danych CSD i PDB, porównaniu otrzymanych wyników ze standardowymi bibliotekami stereochemicznymi, obliczenia statystyczne, udziale w dyskusji ze współautorami o treść artykułu oraz udziale w analizie wyników i przygotowaniu publikacji.*

*Mój udział procentowy szacuję na: 40%*

**IF(2007)=2.620, IF5=9.585, liczba cytowa : 48**

---

**H10. Jaskolski, M., Gilski, M., Dauter, Z. & Włodawer, A. (2007).**

Numerology versus reality: a voice in a recent dispute.

*Acta Cryst. D* **63**, 1282-1283.

*Mój wkład w powstanie tej pracy polegał na: udziale w zainicjowaniu projektu, udziale w dyskusji ze współautorami o treść artykułu oraz udziale w analizie wyników i przygotowaniu publikacji.*

*Mój udział procentowy szacuję na: 40%*

**IF(2007)=2.620, IF5=9.585, liczba cytowa : 8**

---

**H11. Gilski, M. (2014).**

Wysokorozdzielcza krystalografia makromolekuł.

*Wiad. Chem.* **68**, 587-607, IF5=0, cyt. 0

*Mój wkład w powstanie tej pracy polegał na: opracowaniu koncepcji pracy (publikacja przegładowana na zaproszenie edytorów), wybór projektów ilustrujących poszczególne etapy określania wysokorozdzielczej struktury krystalicznej, przygotowanie publikacji.*

*Mój udział procentowy: 100%*

**IF(2014)=0, IF5=0, liczba cytowa : 0**

---

**Suma dla prac H1-H11**

**IF = 47.907, IF5 = 68.314, liczba cytowa : 204**

W tabeli podano wartości parametru Impact Factor (IF) zgodne z rokiem publikacji (dla artykułów opublikowanych w roku 2014 i poniżej podano wartości z roku 2014).

Symbol IF5 oznacza aktualny pięcioletni Impact Factor.

Cytowania podane są zgodnie z bazą ISI Web of Science (11.06.2016r.).

## Metodyczne aspekty wysokorozdzielczej krystalografii makromolekuł

W prezentowanym autoreferacie zawarty jest opis osiągnięć naukowych udokumentowanego cyklem powiązanych tematycznie prac (H1-H11), dotyczących nowych lub usprawnionych metod stosowanych na kolejnych, różnych etapach określania trójwymiarowej struktury krystalicznej makromolekuł, od momentu otrzymania kryształu poprzez rejestrację i obróbkę danych dyfrakcyjnych, rozwiązanie i ułożenie struktury, a także końcowej walidacji uzyskanego modelu struktury. W każdym z tych etapów wykorzystano nowe lub usprawnione metody lub protokoły postępowania. Oprócz tego niezwykle ważnym poznawczym efektem tych działań są wyniki strukturalne, przedstawione w pracach (H2-H8). Wszystkie opisane poniżej struktury były do tej pory nieznanymi a ich szczegółowy opis i analiza stanowi duży wkład do nauki.

### 1. Wprowadzenie

Dzięki rozwojowi metodyki, nowych zaawansowanych technologii oraz technik obliczeniowych i laboratoryjnych, w ostatnich dekadach nastąpił okres rozkwitu krystalografii makromolekuł. Dziś bank struktur białkowych PDB (Protein Data Bank; Berman *et al.*, 2000) zawiera informacji o budowie przestrzennej prawie 120 tys. makromolekuł i powiększa się w tempie ok. 10 tys. struktur rocznie.

Współczesna krystalografia makromolekuł opiera się na metodyce rozwijanej i udoskonalanej od wielu dziesięcioleci, która jest stosowana rutynowo, poprzez procedury implementowane w wielu programach i pakietach oprogramowania krystalograficznego. Metody te są sprawdzone i przetestowane dla wielu różnych rodzajów przypadków. Wiele takich „pakietów” posiada dedykowane interfejsy użytkownika ułatwiające użycie programów dla standardowych przypadków, jednak z reguły nie pozwalają one na wykorzystanie wszystkich zaawansowanych możliwości ukrytego pod nimi oprogramowania. Jednakże w miarę wzrostu liczby struktur, które stają się potencjalnymi obiektami badań, coraz częściej zdarzają się przypadki, kiedy „standardowa” metodyka nie jest wystarczająca do rozwiązania złożonego problemu, jakim jest rozwikłanie przestrzennej struktury białek lub kwasów nukleinowych. W takich przypadkach, w zależności od stopnia skomplikowania problemu, konieczne jest zastosowanie niestandardowego podejścia polegającego na łączeniu wielu różnych metod, zarówno już istniejących, ich indywidualnych modyfikacji, jak również nowych, opracowanych w celu rozwiązania konkretnego zagadnienia.

Proces określenia struktury krystalicznej biomolekuły można wyobrazić sobie jako sekwencję operacji takich jak (i) produkcja białka, (ii) krystalizacja, (iii) rejestracja i obróbka danych dyfrakcyjnych, (iv) rozwiązanie struktury, (v) budowa modelu i udokładnianie oraz (vi) walidacja i deponowanie wyników, (**H1**; Stevens & Wilson, 2001). W celu przyspieszenia całego procesu próbowano opracować powtarzalne protokoły zmierzające do usprawnienia i częściowej automatyzacji niektórych procedur. Pomimo wielu różnych prób, do tej pory nie udało się stworzyć jednolitego systemu sterującego wszystkimi operacjami składającymi się na proces prowadzący do określenia struktury, który byłby powszechnie używany. Zestawy programów opisywane jako systemy do automatycznego określenia struktury (“automated structure-determination pipelines”) takie jak *ACRs* (Brunzelle *et al.*, 2003); *ELVES* (Holton & Alber, 2004) i *SGXpro* (Fu *et al.*, 2005) są próbami połączenia i przetwarzania potokowego kilku etapów procesu określenia struktury. Podobnie jak w przypadku systemów rozwijanych w ramach projektów Genomiki Strukturalnej (GS), które startują od surowych danych dyfrakcyjnych, np. *HKL3000* (Minor *et al.*, 2006), lub bazujących na danych wstępnie przetworzonych, jak w *Auto-Rickshaw* (Panjikar *et al.*, 2005) lub *autoSHARP* (Vonrhein *et al.*, 2007), próbują w sposób częściowo automatyczny rozwiązać strukturę, zbudować model i wstępnie udokładnić model makromolekuły. Niestety sukces tego rodzaju systemów jest ograniczony do przypadków typowych charakteryzujących się dobrze jako ci dane eksperymentalne lub, jak to jest również w przypadku systemu *phenix.ligand\_pipeline* (Echols *et al.*, 2014), do przypadków dla których istnieje dobry model makromolekuły, który może być użyty w metodzie podstawienia cząsteczkowego (MR). Ze względu na złożoność całego procesu określenia struktur makromolekuł, stworzenie jednolitego, uniwersalnego i nie wymagającego interwencji użytkownika systemu jest raczej niemożliwe, lecz niektóre poszczególne etapy tego procesu mogłyby w dużej części zautomatyzowane. Przykładem może być użycie systemu eksperckiego *EDNA* (Incardona *et al.*, 2009) i zestawu zaawansowanych robotów do przeprowadzenia zdalnie kontrolowanej rejestracji synchrotronowych danych dyfrakcyjnych (**H1**; Jaskolski & Gilski, 2007).

Z drugiej strony coraz częściej pojawiają się przypadki wanych projektów strukturalnych, których płynny przebieg zostaje zatrzymany na pewnym etapie z powodu nieoczekiwanych trudności z pokonaniem niektórych problemów. Ilustracją takich przypadków może być np. (i) bardzo istotna z medycznego punktu widzenia struktura monomeru proteazy retrowirusa (**H3**; **H4**); (ii) struktura chimery DNA-RNA (**H8**); (iii) struktura supramolekularnej kapsuły, złożonej z dwóch czasz utworzonych z kaliks[4]rezorcynarenu z aminokwasowymi ramionami (**H5**); (iv)

struktura L-asparaginazy z bakterii *Rhizobium etli*, potencjalnego leku przebiełaczkowego; czy (v) struktura rolinnej cystatyny z *Dactylis glomerata*. Struktury (i)-(iii) zostały rozwinięte i udokumentowane dzięki zastosowaniu niestandardowych metod (pierwsza po 10, druga po 2 latach badań); prace nad rozwikłaniem struktur (iv) i (v) nie są jeszcze zakończone.

Innymi problemami są aspekty metodyczne związane np. (a) z optymalną obróbką surowych danych dyfrakcyjnych (**H2**); (b) ze zgodnością modelu z regułami stereochemii (**H9**; **H10**); czy (c) z optymalnymi protokołami udokładniania wysokorozdzielczego (**H6**; **H7**; **H11**).

## 2. Rejestracja i przetwarzanie dyfrakcyjnych danych synchrotronowych

Rejestracja dyfrakcji promieniowania rentgenowskiego na badanym kryształcie jest w zasadzie ostatnim eksperymentalnym etapem procesu mającego na celu ustalenie trójwymiarowego modelu makromolekuły. Jednocześnie nie jest to kluczowy etap, w wyniku którego otrzymujemy zestaw danych (wskaźniki  $h$ ,  $k$ ,  $l$  refleksów oraz ich intensywności), które są podstawą do wszystkich dalszych etapów obliczeniowych. Dlatego też sposób rejestracji danych dyfrakcyjnych oraz ich przetwarzania mają istotny wpływ na jakość i wiarygodność wyznaczonej ostatecznie struktury. W ostatnich latach powstało wiele lokalnych i międzynarodowych projektów Genomiki Strukturalnej mających na celu określenie w jak najkrótszym czasie maksymalnej liczby struktur makromolekuł. Jest to ściśle powiązane z wysoko-wydajnym sekwencjonowaniem całych genomów kolejnych organizmów. Zdecydowana większość danych dyfrakcyjnych dla kryształów białek i kwasów nukleinowych jest obecnie rejestrowana przy użyciu promieniowania synchrotronowego dostępnego w kilkunastu wielkich ośrodkach synchrotronowych. Dlatego też w projektach GS, które w dużym stopniu zależą od dostępu do tych ośrodków, kładzie się ogromny nacisk na jak najwikszą automatyzację całego procesu określania struktury.

### 2.1 Zdalna rejestracja danych dyfrakcyjnych

Znaczne sukcesy automatyzacji można zaobserwować na etapie produkcji i krystalizacji białek, gdzie obecnie na szeroką skalę wykorzystywane są wyspecjalizowane roboty, które potrafią np. przygotować tysiące próbek krystalizacyjnych w czasie godzin lub minut (Zhu *et al.*, 2014). Roboty i automatyka stały się również powszechne na wszystkich makromolekularnych liniach pomiarowych działających w ośrodkach synchrotronowych. W ramach wielu europejskich projektów, takich jak np. Structural Proteomics In Europe (SPINE) (Betova *et al.*, 2006),



Biocrystallography (X) on a Highly Integrated Technology Platform for European Structural Genomics (BIOXHIT; <http://www.bioxhit.org>), e-Science Resource for High-Throughput Protein Crystallography (e-HTPX; <http://www.e-htpx.ac.uk>) oraz Automated Collection of Data (DNA; Leslie *et al.*, 2002) opracowywane były standardy oraz rozwijano oprogramowanie służące do automatyzacji rejestracji danych, umożliwiające zdalną kontrolę nad eksperymentem synchrotronowym. Ogniwem, które integruje wszystkie kolejne procedury konieczne do automatycznej i zdalnej rejestracji danych dyfrakcyjnych jest system *EDNA* (poprzednia nazwa *DNA*) (Incardona *et al.*, 2009; **H1**; Jaskolski & Gilski, 2007). Jest to system ekspercki, który na podstawie wstępnych danych dyfrakcyjnych jest w stanie podejmować decyzje mające na celu rejestrację danych w sposób optymalny i efektywny. *EDNA* zawiera moduły pozwalające na automatyczne montowanie/usuwanie oraz centrowanie badanego kryształu w wiązce promieniowania rentgenowskiego oraz, po ustaleniu parametrów i strategii pomiaru, dokonanie rejestracji oraz wstępnej obróbki kompletnego zestawu obrazów dyfrakcyjnych. Okazało się, że cały ten proces może być łatwo kontrolowany zdalnie, przy użyciu sieci Internet i komputera oddalonego nawet o tysiące kilometrów od linii pomiarowej (**H1**; Jaskolski & Gilski, 2007; Gilski, 2007). System *EDNA* jest obecnie podstawowym modułem do wstępnej charakterystyki oraz planowania strategii pomiarów dyfrakcyjnych używanym przez większość europejskich ośrodków synchrotronowych (Svensson, 2016). Stało się to możliwe dzięki dokładnym testom, którym został poddany cały system. Pierwsze, zakończone sukcesem testy zdalnego użycia systemu *EDNA* zostały przeprowadzone ponad 10 lat temu w poznańskim Centrum TID (Training, Implementation and Dissemination Center) działającym w Centrum Badań Biokrytalograficznych (CBB) Instytutu Chemii Bioorganicznej PAN w ramach europejskiego programu BIOXHIT (**H1**; Jaskolski & Gilski, 2007; Gilski, 2007). Podczas jednego z wykładów, prowadzonych w czasie regularnych corocznych warsztatów biokrytalograficznych dla doktorantów i studentów z całej Europy organizowanych przez poznańskie Centrum TID, przeprowadziłem na żywo zdalnie kontrolowaną sesję rejestracji danych dyfrakcyjnych na linii ID29 w ośrodku synchrotronowym ESRF (European Synchrotron Radiation Facility) w Grenoble. W trakcie sesji zaprezentowałem następujące możliwości: (i) zdalnego sterowania robotem umieszczającym, zamrozone w ciekłym azocie, kryształy w wiązce promieniowania rentgenowskiego; (ii) centrowania kryształu; (iii) jego wstępnej charakterystyki przy użyciu systemu *EDNA*; (iv) ustawienia optymalnych parametrów linii pomiarowej ID29; (v) rejestracji kompletnego zestawu danych dyfrakcyjnych; oraz (vi) transferem wstępnie przetworzonych danych oraz surowych obrazów dyfrakcyjnych na lokalny komputer wykładowcy w Poznaniu

(H1; Jaskolski & Gilski, 2007; Gilski, 2007). Wydarzenie to, które powtórzyłem podczas kilku innych wykładów, wzbudziło duże zainteresowanie środowiska naukowego, czego przejawem była rejestracja wykładu przez Uniwersyteckie Studio Filmowe UAM oraz emisja na antenie TV WTK w „Magazynie akademickim - Życia Uniwersytetu”. Zostało ono również opisane (w polskiej i angielskiej wersji językowej; Jaskolski & Gilski, 2007) w wydawanym przez PAN, popularnonaukowym kwartalniku *Academia*, promującym w kraju i za granicami osiągnięcia polskich badaczy. Ten sposób rejestracji danych został w późniejszym okresie wykorzystany do rutynowych pomiarów związanych z projektami badawczymi realizowanymi w CBB w Poznaniu.

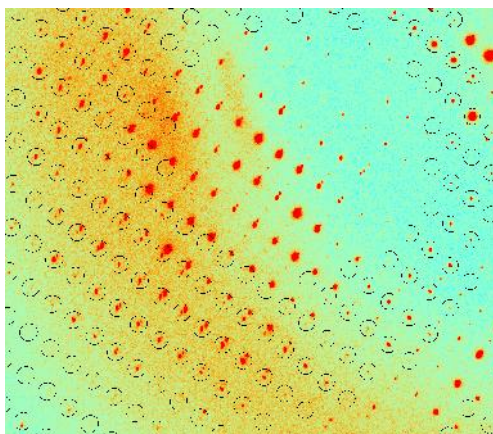
## 2.2 Przetwarzanie danych dyfrakcyjnych

Niezależnie od tego, czy obrazy dyfrakcyjne zostały zarejestrowane automatycznie, zdalnie czy pod osobistym nadzorem badacza, kolejnym krokiem musi być prawidłowa i optymalna procedura obróbki danych polegająca na skalowaniu, integracji i skalowaniu intensywności zarejestrowanych refleksów. Również w tym przypadku, szczególnie w ostatnich latach, pojawiają się próby częściowej automatyzacji. Jednym z bardziej udanych rozwiązań jest system *XDSAPP* (Krug *et al.*, 2012) oparty na rozwijanym i testowanym od lat 80-tych XX. wieku, pakiecie programów *XDS* (Kabsch, 2010). *XDSAPP* jest w zasadzie interfejsem graficznym programu *XDS* z elementami systemu eksperckiego, który w jednoznacznych przypadkach jest w stanie przetworzyć dane dyfrakcyjne przy niewielkim udziale użytkownika. Jednakże, jak wskazuje praktyka, nawet w takich przypadkach manualna optymalizacja parametrów dokonana przez doświadczonego badacza pozwala uzyskać jakoś i wiarygodnościowych wyników (H2).

Do grupy najczęściej obecnie używanych programów przeznaczonych do obróbki surowych obrazów dyfrakcyjnych należą: *HKL2000* (Otwinowski & Minor, 1997), *MOSFLM* (Leslie, 1992), *d\*TREK* (Pflugrath, 1999) oraz wspomniany już pakiet *XDS*. Programy te są rozwijane i testowane od wielu lat, oparte są na podobnych algorytmach i dla większości typowych wyników pomiaru dyfrakcyjnego dają podobny wynik końcowy. Różnice jednak uwypuklają się w przypadkach trudnych lub nietypowych. Jak wykazałem w pracy H2, przetwarzanie niektórych danych dyfrakcyjnych, szczególnie tych najciekawszych, z reguły nie jest trywialne i większość automatycznych czy półautomatycznych procedur zawodzi. W takich przypadkach niezbędnym jest dobór odpowiedniego narzędzia przetwarzającego dane oraz ręczne zoptymalizowanie wielu skorelowanych parametrów kontrolnych, a niekiedy zastosowanie

eksperymentalnych skryptów, do czego konieczne jest doświadczenie oraz dogłębna znajomość i zrozumienie całego procesu i sposobu działania programu.

Przykładem takiego przypadku mogą być synchrotronowe dane dyfrakcyjne (NRAD, rozdzielczość 1.58 Å) kryształu białka RadA z archeonta *Pyrococcus horikoshii* (Włodawer, 2007; Oeemig, 2012; Rys. 1). Wizualna ocena obrazów dyfrakcyjnych natychmiast wykazała ewidentne niemeroedryczne zbliniaczenie kryształu u tego do pomiarów. Jednakże początkowo wszystkie wyżej wymienione programy, które używałem do obróbki danych w trybie standardowym dały w miarę dobrej jakości dane oparte tylko na refleksach pochodzących od dominującej domeny. Dalsza manualna optymalizacja i powtórna obróbka danych przy użyciu programu *XDS* i opracowanych protokołów i skryptów zakładających wielodomenowe zbliniaczenie (**H2**), umożliwiła mi otrzymanie dwóch niezależnych i kompletnych zestawów danych, pochodzących od dwóch domen kryształu, które oczywiście reprezentowały dwie różne formy krystaliczne z parametrami komórki elementarnej różniącymi się o ponad 5%. Oprócz tego, uzyskane przeze mnie zestawy danych charakteryzowały się lepszymi parametrami określającymi ich jako , np. wskaźnik  $R_{\text{merge}}$  zmniejszył się z 8.6 do 7.5% (dla dominującej domeny), uzyskano znacznie lepszą rozdzielczość (1.45 zamiast 1.58 Å) oraz o ponad 6000 większą liczbę niezależnych refleksów (**H2**).



Rys. 1. Fragment obrazu dyfrakcyjnego kryształu NRAD z nałożonymi pozycjami refleksów (kółka) przewidywanymi dla danej komórki i grupy przestrzennej. Nie wszystkie pozycje refleksów są zgodne z przewidywaniami, co jest związane z niemeroedrycznym zbliniaczeniem (**H2**).

Innym przykładem niestandardowego przetwarzania danych jest ponowna obróbka obrazów dyfrakcyjnych kryształów heksameru Z-DNA sekwencji d(CGCGCG), na podstawie których określono strukturę kwasu nukleinowego o rekordowo wysokiej rozdzielczości 0.55 Å (Brzezinski *et al.*, 2011). Po ponownym przeprocesowaniu danych przy użyciu pakietu *XDS*

uzyskałem ostateczny zestaw danych o rozdzielczości 0.53 Å o wysokiej jakości i kompletności, który zawierał ponad 1700 refleksów w tej pierwotny zestaw danych (**H2**). Dane te są obecnie wykorzystane do badań deformacji gęstości elektronowej ukazującej odchylenia od powszechnie stosowanego podczas udokładniania modelu atomu sferycznego. Struktura Z-DNA została ponownie udokładniona z wykorzystaniem modelu multipolowego atomu. Otrzymany w ten sposób model struktury charakteryzuje się niezwykle precyzyjnie, uwidaczniając elementy struktury np. podwójne konformacje, których nie można było zaobserwować w strukturze udokładnionej w klasyczny sposób. Jest to pierwsza struktura kwasu nukleinowego udokładniona multipolowo (Guillot *et al.*, 2016; dane niepublikowane, publikacja w trakcie przygotowania, mój wkład do tej pracy szacuję na ok. 30%).

Z powyższych rozważań wynika, że rejestracja i przetwarzanie dyfrakcyjnych danych synchrotronowych mogłyby w dużej mierze wykonywane zdalnie oraz w sposób czysto automatyczny. Specjalistyczne systemy pozwalają na zdalne wykonanie eksperymentu, co związane jest też w istotny sposób z kosztami prowadzonych badań. Jednakże coraz częściej krytalografowie muszą poradzić sobie również z przypadkami, w których rutynowe użycie zastosowanych w programach procedur jest niewystarczające, co zostało opisane przeze mnie w pracach **H1** i **H2**.

### **3. Niestandardowe metody rozwiązywania struktur krystalicznych**

Dysponując zestawem eksperymentalnych amplitud czynników struktury (lub intensywności refleksów), otrzymanych w wyniku rejestracji oraz przetwarzania danych dyfrakcyjnych, można podjąć kolejny krok prowadzący do rozwiązania struktury. Etap ten jest z matematycznego punktu widzenia jednym z najtrudniejszych oraz ma kluczowe znaczenie dla dalszego przebiegu całego procesu. Głównym jego celem jest rozwiązanie jednego z fundamentalnych problemów krytalografii rentgenowskiej, tzw. „problemu fazowego”. We współczesnej krytalografii makromolekuł głównymi metodami rozwiązania problemu fazowego są podstawienie cząsteczkowe (Molecular Replacement, MR; Rossmann & Blow, 1962) oraz metoda polegająca na wprowadzeniu do struktury atomów o specjalnych właściwościach dyfrakcyjnych. W uproszczeniu, podstawienie cząsteczkowe polega na użyciu, jako przybliżonego modelu, makrocząsteczki o podobnej strukturze, np. jej mutantu lub homologu, oraz porównania dyfrakcji obserwowanej eksperymentalnie z transformat Fouriera obliczoną teoretycznie. Druga metoda rozwiązania problemu fazowego polega na wprowadzeniu do kryształu (np. poprzez

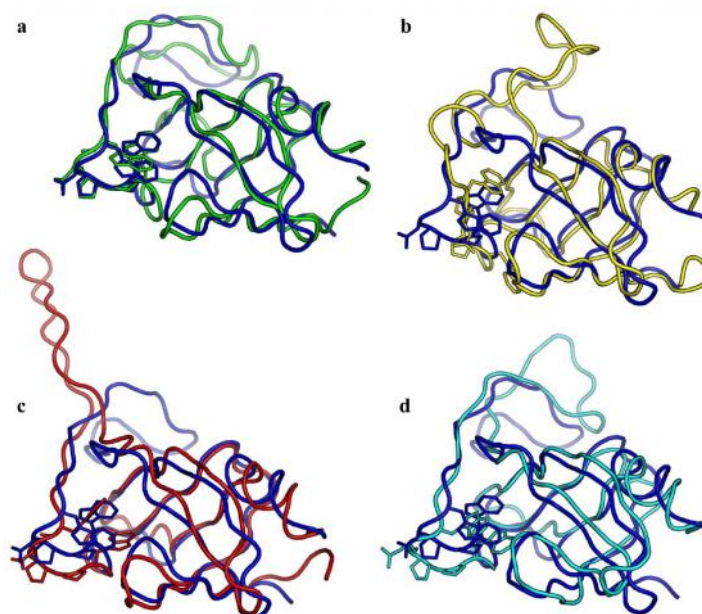
nas czanie, lub modyfikacji niektórych aminokwasów), bez zaburzenia jego struktury, atomu o dużej liczbie elektronów („ciężkiego”) dającego silną dyfrakcję rentgenowską lub takiego, dla którego można zarejestrować dyfrakcję anomalną i na tej podstawie wyznaczyć współczynniki fazy (Dauter & Jaskolski, 2011).

Jak wspomniano wyżej, metody rozwijane i uwyślane przez ostatnie dziesięciolecie rozkwitu krystalografii białek i kwasów nukleinowych nie zawsze jednak pozwalają rozwiązywać nietypowe problemy napotymane na drodze prowadzącej do rozszyfrowania struktury makromolekuła.

### 3.1 Struktura proteazy wirusa M-PMV

Doskonałym przykładem tego typu problemów może być struktura monomeru proteazy (PR) małpiego wirusa Masona–Pfizera (M-PMV) powodującego zespół nabytego niedoboru odporności (Simian Acquired Immunodeficiency Syndrome, SAIDS) u azjatyckich małp z rodzaju *Macaca* (**H3**; **H4**). Ze względu na fakt, że reowirus M-PMV jest bliskim kuzynem wirusa HIV, odpowiedzialnego za chorobę AIDS u ludzi, poznanie struktury monomerycznej formy tej PR może pozwolić na opracowanie nowych leków stosowanych w zwalczaniu innych infekcji reowirusowych, również takich jak w przypadku AIDS.

Pomimo udanych eksperymentów krystalizacyjnych, w wyniku których otrzymano kilka form krystalicznych proteazy M-PMV w postaci monomeru, a także rejestracji wielu zestawów wysokiej jakości danych dyfrakcyjnych i przeprowadzeniu optymalnej obróbki tych danych, przez ponad dziesięć lat nie udało się rozwiązać struktury tego białka. Z pozoru był to idealny przypadek dla użycia klasycznej metody podstawienia cząsteczkowego – obiektem badania jest małe białko o masie cząsteczkowej 13 kDa (114 reszt aminokwasowych); istnieje wiele znanych struktur krystalicznych proteaz reowirusowych z rodziny retropepsyn w formie dimerycznej (Włodawer & Gutchina, 2000), które mogły służyć jako modele w procedurze MR; znany jest monomeryczny model PR M-PMV otrzymany metodą rezonansu magnetycznego NMR (Veverka *et al.*, 2003); dane dyfrakcyjne charakteryzują się wysoką rozdzielczością i bardzo dobrą jakością. Jednakże wszystkie próby zastosowania metody MR wraz z modelami tak w formie dimerycznej, jak i pojedynczych podjednostek oraz modelu monomeru PR (otrzymanego metodą NMR) zawiodły. Również zastosowane przez mnie wszystkie znane zaawansowane protokoły wykorzystujące metodę MR i jej różnorodne modyfikacje nie przyniosło spodziewanych wyników.



Rys. 2. Nałożenie różnych modeli stosowanych podczas prób rozwinięcia struktury M-PMV PR metod podstawienia cz. steczkowego. Ostatecznie określona struktura krystaliczna jest pokazana w kolorze niebieskim. (a) Na zielono: najlepszy model stworzony przez graczy gry Foldit. (b) Na żółto: najlepszy model zbudowany przy użyciu protokołu „rebuild-and-refine” programu Rosetta startujący z modelu NMR. (c) Na czerwono: model z protokołu „relax” programu Rosetta. (d) W kolorze turkusowym: model z modułu CS-Rosetta stosującego przesunięcia chemiczne jako wiązki (H3).

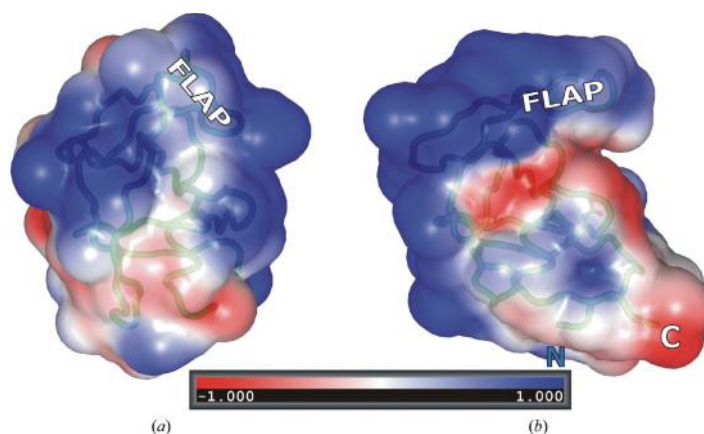
Następnie podjęto próby zbudowania odpowiedniego modelu, na podstawie dostępnych podobnych struktur i metod bioinformatycznych. Zbudowanie modelu, który byłby podobny do rzeczywistej struktury monomeru proteazy M-PMV, a następnie znalezienie wstępnych faz mogłoby doprowadzić do rozwinięcia struktury. Jedną z takich metod jest nowy algorytm *mr\_rosetta* (DiMaio *et al.*, 2011), który jest pojęciem, bioinformatycznego programu *Rosetta* (Rohl *et al.*, 2004), najbardziej skutecznego (Bonneau *et al.*, 2001; Bradley *et al.*, 2003; Lesk *et al.*, 2001) systemu do przewidywania *in silico* trójwymiarowych struktur białek, z algorytmem podstawienia cz. steczkowego. *Mr\_rosetta* próbuje zbudować nowy model na podstawie znanych podobnych struktur wykorzystując jednocześnie eksperymentalne dane dyfrakcyjne jako wiązki. Po bardzo czasochłonnych obliczeniach, wykorzystujących m.in. różne protokoły i moduły systemu *Rosetta*, metoda *mr\_rosetta*, która została użyta wcześniej do rozwinięcia wielu trudnych struktur makromolekuł (DiMaio *et al.*, 2011), wygenerowała co prawda zestaw najbardziej korzystnych energetycznie modeli białka, lecz niestety żaden z nich nie był wystarczająco dobry, by posłużyć do rozwinięcia struktury (H3; H4; Rys. 2).

Moduły systemu *Rosetta* wykorzystywane są w wielu narzędziach bioinformatycznych

(Kaufmann, 2010) służących np. do dokowania białek i małych cząsteczek w strukturze, modelowania proteiny oraz projektowania białek. Jednym z ciekawszych zastosowań funkcji szacowania energii pakietu *Rosetta* jest użycie jej w internetowej grze komputerowej *Foldit* (Cooper *et al.*, 2010), w której tysiące graczy z całego świata próbuje pomóc w rozwikłaniu niektórych problemów związanych z przewidywaniem struktur makromolekularnych cząsteczek. Moduły z systemu *Rosetta* służą do wyznaczania energii konformacyjnej modeli zbudowanych przez graczy, co jest podstawą punktacji w grze oraz elementem współzawodnictwa.

W grze tej gracze z całego świata (kilkaset tysięcy zarejestrowanych) samodzielnie, lub tworząc nieformalne grupy, próbują rozwiązać zadania postawione przez autorów gry. Zadania te polegają na takim przestrzennym manipulowaniu otrzymanym wirtualnym modelem makromolekuły, aby uzyskać najbardziej korzystną energetycznie konformację.

Do puli zadań dostępnych dla graczy *Foldit* dodaliśmy w 2011 roku nowy cel: stworzenie takiego modelu proteazy M-PMV (**H3**; **H4**), który pozwoliłby uzyskać wirtualne fazy i ostatecznie rozwiązać strukturę krystaliczną metodą MR. Startując ze współrzędnych atomowych struktury NMR Inso (Veverka *et al.*, 2003) w czasie trzech tygodni gracze, którzy w większości nie mieli żadnych związków z nauką i nie znali w zaawansowanym stopniu biochemii i biologii strukturalnej, wygenerowali ponad milion różnych modeli proteazy M-PMV (**H3**; **H4**). Okazało się, że jeden z tych modeli, po użyciu go w metodzie podstawienia cząsteczkowego, pozwolił wygenerować wiarygodne rozwiązanie struktury (**H3**). To rozwiązanie umożliwiło nam przejście do kolejnych etapów, takich jak udokładnianie, analiza i szczegółowy opis struktury (**H4**). Udokładniony model struktury krystalicznej proteazy wirusa M-PMV charakteryzuje się wysoką jakością i wiarygodnością, co pozwala na jednoznaczny analizę wszystkich aspektów strukturalnych opisanych szczegółowo w pracy **H4**.



Rys. 3. Powierzchnia potencjału elektrostatycznego protomerów proteazy retrowirusowej. Monomer M-PMV PR (a) jest pokazany w tej samej skali i orientacji jak protomer HIV-1 PR (b) wyodrębniony z dimeru (kod PDB: 3hvp; **H4**).

Jest to pierwsza struktura proteazy retrowirusowej w formie monomerycznej (**H3**; **H4**). Na szczególną uwagę w tej strukturze retropepsyny zasługuje powierzchnia białka, która normalnie w cząsteczce homodimerycznej tworzy powierzchnię styku (interfejs) podjednostek dimeru, Rys. 3. W formie monomerycznej można zauważyć znaczące przesłabienie łata cucha głównego w obrębie pętli (tzw. „flaps”) i miejsca aktywnego oraz kompletnie nieuporządkowane N- i C-końce (Rys. 2). Te cechy dają możliwość zaprojektowania nowej generacji leków antyretrowirusowych (w tym również przeciw wirusowi HIV). Cząsteczka leku przyłączona do powierzchni wyeksponowanej w monomerze, lecz ukrytej w formie dimerycznej, mogłaby przesunąć równowagę dimeryzacji w kierunku monomeru, który jest katalitycznie nieaktywny, i umożliwiłby tym samym dojrzenie nowych wirionów (**H4**).

Rozwinięcie struktury monomeru proteazy M-PMV pokazuje ogromne możliwości gier internetowych, które wykorzystując zbiorową intuicję i umiejętności ludzi, takie jak wyobraźnia przestrzenna, mogą doprowadzić do rozwiązania nowych problemów naukowych. Choć ostatnio poświęca się sporo uwagi tego typu rozwiązaniom, określanym ogólnie pojęciem „crowd sourcing” (Kiss *et al.*, 2013), to opisane w pracach **H3** i **H4** zastosowanie takiej metody było pierwszym przypadkiem, gdy gracze internetowej gry komputerowej rozwiali rzeczywisty, bardzo ważny i przez wiele lat nierozwiązalny problem naukowy. Uzyskane rezultaty ilustrują olbrzymi potencjał integracji gier komputerowych z odbywającym się w rzeczywistym świecie procesem badań naukowych. Kreatywność i pomysłowość tysięcy graczy jest potężną siłą, która odpowiednio skierowana, może być użyta do rozwiązania wielu różnorodnych problemów naukowych. Publikacja artykułów **H3** oraz **H4** spotkała się z dużym zainteresowaniem medialnym a informacje o tym odkryciu były prezentowane w serwisach prasowych, radiowych i telewizyjnych w Polsce i zagranicą.

### 3.2 Nietypowe zastosowanie metod makromolekularnych

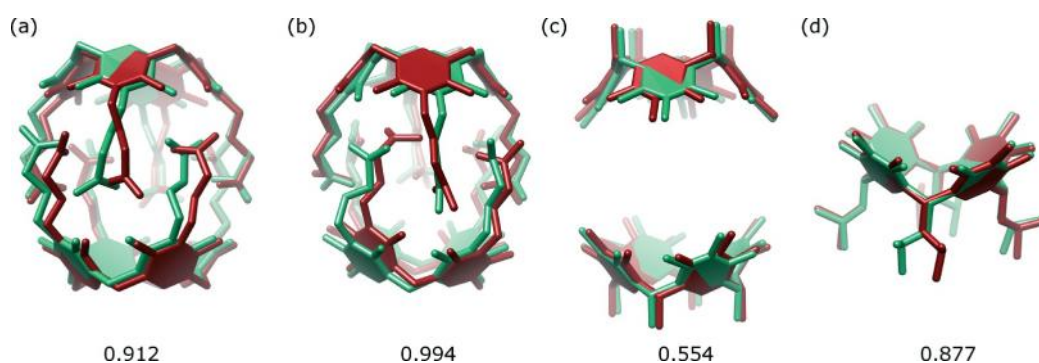
W większości przypadków w krystalografii makromolekuł niemożliwe jest wykorzystanie metod bezpośrednich np. *SHELXS* (Sheldrick, 2008), *SIR* (Burla *et al.*, 2012), używanych powszechnie od wielu lat do rozwijania struktur małych i średnich cząsteczek organicznych (o wielkości do kilkuset atomów). Metody te, należące do grupy metod *ab initio*, która obejmuje zarówno klasyczne metody bezpośrednie jak również metody typu „dual-space”, np. *SHELXD* (Sheldrick, 2010), *SUPERFLIP* (Palatinus & Chapuis, 2007) czy *Shake-and-Bake* (Miller *et al.*, 1994),



wymagaj danych o wysokiej jakości i rzadko uzyskiwanej dla kryształów makromolekuł rozdzielczości lepszej niż  $1.2 \text{ \AA}$  (**H5**; **H11**). Poza tym typowe cząsteczki makromolekuł, których dane strukturalne są zdeponowane w bazie PDB, zbudowane są z reguły z kilku tysięcy atomów (średnia liczba atomów w modelach makromolekuł w bazie PDB to  $\sim 7000$ ).

Pomimo tego, te podstawy teoretyczne metod używanych w krytalografii makromolekuł i krytalografii małych i średnich cząsteczek organicznych są w zasadzie takie same, to eksperymentalne i obliczeniowe aspekty różnią się znacznie, co częściowo wynika z tradycji, jak również z odmienności obu dziedzin. Metody wykorzystywane w oprogramowaniu sprawdzają się doskonale w przypadkach, gdy używane narzędzie jest dobrze dopasowane do badanego problemu. Jednak ta sytuacja staje się kłopotliwa, gdy trudno jednoznacznie określić czy badany obiekt jest wielką cząsteczką czy też małym makromolekułem. Wtedy może się okazać, że metody krytalografii małych cząsteczek już nie działają, a metody zaprojektowane dla makromolekułów są nie do końca adekwatne. Do tego typu projektów należą badania obiektów supramolekularnych, których wielkość (od kilkuset do tysiąca atomów niewodorowych) jest pomiędzy małymi cząsteczkami a makromolekułami. Sytuacja jest szczególnie trudna jeżeli należą one do grupy dużych niekwalencyjnych układów, takich jak chiralne kapsuły zbudowane z lekkich atomów (**H5**; Szymanski *et al.* 2016). W takich przypadkach metody przeznaczone do rozwijania struktury małych cząsteczek często zawodzą, głównie z powodu specyficznej budowy kapsuł supramolekularnych. Takie cechy jak występowanie w strukturze luk i pustych przestrzeni o dużej objętości, które często są wypełnione nieuporządkowanymi cząsteczkami organicznego rozpuszczalnika używanego podczas krystalizacji, mogą wpływać negatywnie na jakość dyfrakcji. Jednakże inną cechą strukturalną kapsuł supramolekularnych, mianowicie fakt, że są one zbudowane z dobrze zdefiniowanych, powtarzalnych fragmentów molekularnych, umożliwia budowanie struktur krystalicznych z „klocków” molekularnych o znanej strukturze, co z kolei stanowi istotną metodę podstawienia molekularnego (MR), powszechnie stosowanej do rozwijania struktur makromolekularnych. Metoda ta za pomocą rotacji i translacji przybliżonego modelu struktury (lub jej fragmentu) próbuje umieścić go w asymetrycznej komórce elementarnej badanego kryształu w sposób najbardziej zgodny z danymi eksperymentalnymi. Programy używane metod MR są rozwijane i stosowane w krytalografii makromolekuł od kilku dziesięcioleci. Skuteczność tej metody zależy głównie od dwóch czynników: jakości asymetrycznej komórki elementarnej badanego kryształu oraz od stopnia zgodności modelu z docelową strukturą, określonego średnim kwadratowym odchyleniem (r.m.s.d., root mean square deviation) otrzymanym po optymalnym nałożeniu modelu i

fragmentu docelowej struktury. Dostanie dobrego modelu jest niezbędnym warunkiem powodzenia, jednak wybór odpowiedniej strategii może być równie decydujący o sukcesie metody. Czynnikiem komplikującym rozwiązanie problemu fazowego metod MR jest wysoka symetria, translacyjna symetria niekrystalograficzna (tNCS), gęste upakowanie kryształu, oraz utrudnienie podczas przeszukiwania kilku modeli jednocześnie, które odpowiadają różnym fragmentom komórki asymetrycznej. Wyżej wymienione aspekty opisaliśmy szczegółowo w pracy **H5**. Na przykładzie supramolekularnej chiralnej struktury kapsuły organicznej zbudowanej z dwóch niekowalencyjnie złożonych podjednostek (czasz), utworzonych przez sprężenie szkieletu kaliks[4]rezorcynarenu z czterema ramionami zbudowanymi z L-alaniny (Rys. 4), opracowaliśmy protokół postępowania w przypadku utrudnienia metodyki krystalografii makromolekuł do rozwiązania struktury supramolekularnej. Dane dyfrakcyjne, zarejestrowane przy użyciu tradycyjnego źródła promieniowania rentgenowskiego, charakteryzowały się wysoką rozdzielczością (0.90 Å) i jakością ( $R_{\text{int}}=0.044$ ) i wydawały się odpowiednie do zastosowania metod *ab initio* w celu rozwiązania struktury (**H5**). Pomimo utrudnienia różnych metod dostępnych w programach takich jak: *SHELXS*, *SIR*, *SUPERFLIP*, niezależnie przeprowadzone w laboratoriach w Polsce (Szumna) oraz w Finlandii (Rissanen) próby rozwiązania struktury nie powiodły się. Ponieważ wszystkie metody używane w krystalografii małych i średnich cząsteczek zawiodły, postanowiliśmy zastosować makromolekularną metodę podstawienia cząsteczkowego (MR). Jako model wstępny posłużył rdzeń kapsuły z poznanej już struktury analogu z L-feniloalaniną (Kuberski & Szumna, 2009; Rys. 4a). Pomimo tego, że ramiona oligopeptydowe są bardzo giętkie, motyw asocjacyjny kapsuły był na tyle zachowany, że udało się ostatecznie rozwiązać strukturę.



Rys. 4. Natężenie odpowiadających sobie fragmentów rozwiązanej struktury kapsuły supramolekularnej (kolor zielony) i różnych modeli używanych w metodzie MR (kolor czerwony). (a) szkielet struktury analogu z L-feniloalaniną; (b) zbudowany na podstawie mechaniki molekularnej; (c) szkielety dwóch czasz kapsuły; (d) szkielet jednej czaszy z giętkimi łańcuchami alkilowymi. Wartości r.m.s.d. w Å podane są dla każdego przypadku poniżej modelu (**H5**).

Następnie przeprowadziliśmy szereg testów u różnych zestawów modeli o różnej wielkości jak również pochodzeniu (np. model zbudowany na podstawie mechaniki molekularnej i informacji ze spektroskopii NMR). Szczegółowo przeanalizowaliśmy możliwość i wydajność metody MR opartej na najwskazanej wiarygodności wykorzystywanej w programie *PHASER* (McCoy *et al.*, 2007). W ten sposób pozytywnie oceniono przydatność tej metody do rozwiązywania trudnych struktur supramolekularnych (Rys. 4a-d). Na tej podstawie opracowaliśmy protokół postępowania w takich nietypowych przypadkach, który może być bezpośrednio zastosowany przez innych badaczy. Protokół opisany w pracy **H5** został wykorzystany do rozwiązania wielu podobnych struktur supramolekularnych (Szymanski *et al.* 2016; Sinigharoy *et al.*, 2015), które jednocześnie nie pozwoliły na jego dalszą optymalizację.

Niejakim ubocznym efektem zainteresowania badaczy układów supramolekularnych metodami krystalografii makromolekuł, są coraz częstsze próby rejestracji danych dyfrakcyjnych przy użyciu promieniowania synchrotronowego. Otwiera to przed nimi nowe możliwości poznania struktur krystalicznych również w przypadkach, gdy zdolność kryształu do rozpraszania promieniowania rentgenowskiego jest niewielka. Połączenie danych synchrotronowych (niekoniecznie bardzo wysokorozdzielczych) z metodą podstawienia cząsteczkowego, która, jak to wykazaliśmy (**H5**; Szymanski *et al.* 2016), jest skuteczna w szerokim zakresie jakości i wielkości modelu, może pozwolić na rozwiązanie problemów, które niekiedy uniemożliwiałyby poznanie struktury.

#### 4. Udokładnianie wysokorozdzielczych struktur makromolekuł

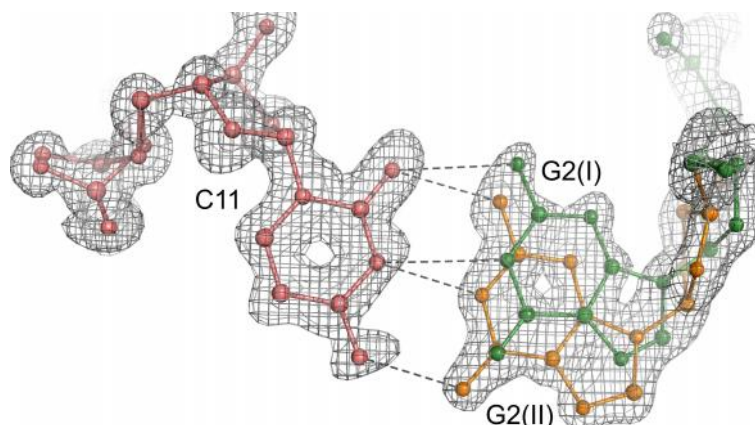
Wszystkie kroki prowadzące do wyznaczenia trójwymiarowej struktury krystalicznej są ważne, jednak udokładnianie jest tym etapem, na który należy zwrócić szczególną uwagę, przede wszystkim w przypadku udokładniania wysokorozdzielczych struktur makromolekuł. Prawidłowe przeprowadzenie procesu udokładniania ma dominujący wpływ na jakość i precyzję uzyskanego modelu oraz jego zgodność z rzeczywistymi strukturami (**H11**). Opracowano wiele programów komputerowych przeznaczonych do udokładniania modeli struktur makromolekuł, wśród których do najczęściej używanych należą *REFMAC5* (Murshudov *et al.*, 2011), *phenix.refine* (Adams *et al.*, 2010) oraz *SHELXL* (Sheldrick, 2008). Dwa pierwsze oparte są głównie na metodzie najwskazanej wiarygodności (ang. *maximum likelihood*, *ML*), trzeci natomiast na metodzie najmniejszych kwadratów (ang. *least squares*, *LSQ*). Programy te powstały w różnych okresach i oferują użytkownikom ogromną ilość możliwości dopasowania

parametrów do konkretnego przypadku. Ponieważ stopień skomplikowania procesu udokładniania rośnie wraz ze wzrostem rozdzielczości danych dyfrakcyjnych, udokładnianie struktur wysokorozdzielczych wymaga zastosowania szczególnych, często niestandardowych protokołów udokładniania (**H11**). Przykładem zastosowania takich protokołów mogą być struktury lewoskrętnych dupleksów Z-DNA z autokomplementarną sekwencją d(CGCGCG), które mają nadzwyczajny potencjał do tworzenia kryształów rozpraszających promieniowanie rentgenowskie do bardzo wysokiej rozdzielczości. Zostało to potwierdzone wieloma badaniami krystalograficznymi, włącznie ze wspomnianymi strukturami o rekordowo wysokiej rozdzielczości 0.55 Å (Brzezinski *et al.* 2011; Guillot *et al.*, 2016).

#### 4.1 Wysokorozdzielcze struktury kwasów nukleinowych

Seria wysokorozdzielczych struktur kompleksów Z-DNA z różnymi dwu- i trójwartościowymi kationami metali i protonowanym tetrakationem sperminy (**H6**; **H7**; **H8**; Drozdal *et al.*, 2016) pozwoliła nam nie tylko w bardzo precyzyjny i szczegółowy sposób opisać wszystkie detale strukturalne, lecz również opracować optymalne procedury krystalograficzne w takich przypadkach. Struktury kompleksów Z-DNA z jonami  $Mn^{2+}$  i  $Zn^{2+}$  o rozdzielczości 0.75 i 0.85 Å (**H6**) mają najwyższą rozdzielczość i precyzję wśród wszystkich kompleksów Z-DNA z kationami metali przejściowych zdeponowanymi w bazie PDB. Struktury obu kompleksów zostały określone z ultra-wysoką rozdzielczością z zastosowaniem na ostatnim etapie pełnomacierzowego udokładnianiem metod najmniejszych kwadratów, co dostarczyło całego bogactwa precyzyjnych danych strukturalnych wraz z ich odchyleniami standardowymi. Takie dane pozwalają w statystycznie poprawny sposób dyskutować niuanse w geometrii makromolekuły oraz jednoznacznie zinterpretować mapę gęstości elektronowej, np. precyzyjnie określić sposób nieuporządkowania struktury. W kompleksie  $Zn^{2+}$  można zauważyć, że łańcuch główny DNA jest zdeformowany do tego stopnia, że przyjmuje w wielu segmentach podwójną konformację. Nieuporządkowanie łańcucha głównego DNA zostało również zauważone w strukturze kompleksu chimery DNA-RNA z kationami baru (**H8**), we wszystkich czterech niezależnych dupleksach w asymetrycznej części komórki. Oprócz nieuporządkowanego szkieletu fosforanowo-cukrowego łańcucha głównego Z-DNA w kompleksie  $d(CGCGCG)_2-Cr^{3+}$  (**H7**), zaobserwowaliśmy również podwójną konformację pierścienia guaniny, co było pierwszym takim przypadkiem opisanym w literaturze w znanych do tej pory strukturach Z-DNA. Jednakże alternatywne położenia pierścienia zasady guaniny G2 nie wpłynęły na orientację komplementarnej zasady cytozyny C11, co można jednoznacznie stwierdzić analizując

wysokiej jako ci map  $mF_o-DF_c$  (Rys. 5).



Rys. 5. Fragment struktury kompleksu  $d(CGCGCG)_2-Cr^{3+}$  pokazujący podwójną konformację (I, zielony; II pomarańczowy) nukleotydu G2 i sparowany z nim nukleotyd C11 (czerwony) wraz z mapą  $2mF_o-DF_c$  skonturowaną na poziomie 1.0 (H7).

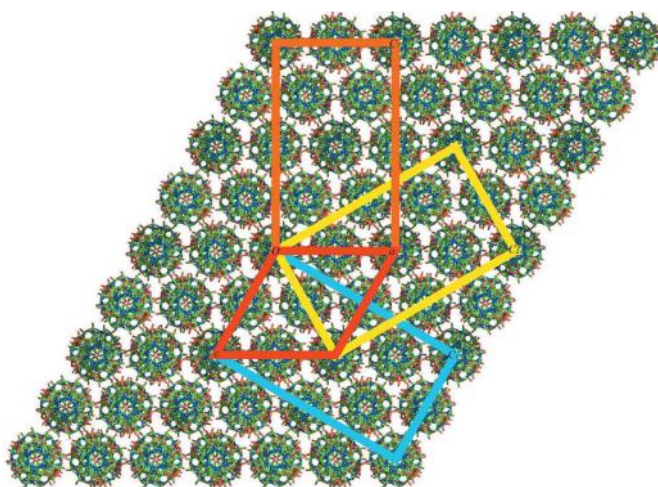
Tego typu nieporządek zaobserwowaliśmy również w strukturze D/L-DNA/Mg<sup>2+</sup>, (racemicznej mieszaniny D- i L-oligonukleotydów o różnej chiralności) określonej z rozdzielczością 0.78 Å (Drozdal *et al.*, 2016), która wykazuje najwyższy stopień nieuporządkowania wśród wszystkich, poznanych do tej pory wysokorozdzielczych struktur Z-DNA. Pomimo (trochę przedwcześnie) sugerowanej przez Brzezickiego *et al.* (2011) strukturalnej sztywności Z-DNA, nasze kolejne precyzyjne badania (H6; H7; H8; Drozdal *et al.*, 2016) skłaniają jednak do wniosku, że wewnętrzna elastyczność dupleksów Z-DNA jest naturalnym stanem dla tej formy DNA (Malinina *et al.*, 1998).

#### 4.2 Struktura chimery DNA-RNA – skomplikowane wielodomenowe zbliniaczenie

Dodatkowym wyzwaniem, w procesie określania struktury krystalicznej makromolekuł jest występowanie różnorodnych patologii kryształu, takich jak zbliniaczenie czy pseudosymetria. Jak wykazano (H8; Lebedev *et al.*, 2006; Dauter, 2003) z reguły nie wpływają one w znaczący sposób na możliwość rozwinięcia struktury za pomocą metody MR (np. w wersji zastosowanej w programie PHASER).

Problematyczne i bardzo pracochłonne może okazać się ustalenie prawdziwej grupy przestrzennej oraz udoładnianie (H8). Kryształy kompleksu chimery DNA-RNA z kationami Ba<sup>2+</sup> (jak również Sr<sup>2+</sup>) wykazują skomplikowane, wielodomenowe zbliniaczenie połączone z translacyjną i rotacyjną pseudosymetrią, co w znacznym stopniu utrudnia określenie struktury (H8). Pomimo faktu, że zarejestrowane dane dyfrakcyjne charakteryzowały się wysoką

rozdzielczości (1.09 Å) i jako ci ( $R_{\text{merge}}=6.2\%$ ) pierwszym napotkanym problemem okazała się niejednoznaczność określenia grupy przestrzennej. Zestaw danych eksperymentalnych można było wskaźnikowa, lecz nie zawsze skalowa, w wielu różnych grupach przestrzennych a prymitywna komórka elementarna miała metryki typowe dla sieci heksagonalnej z dwoma parametrami równymi i jednym kątem  $\sim 120^\circ$ . Chociaż wskaźnikowanie obrazów dyfrakcyjnych przy pomocy programu *XDS* w heksagonalnej grupie  $P6_322$  było możliwe, skalowanie zakończyło się niepowodzeniem ( $R_{\text{merge}}=26.5\%$ ) sugerując, że nie jest to właściwa grupa przestrzenna. Komórka heksagonalna może być w prosty sposób przekształcona do komórki centrowanej typu  $C$ , tzw. komórki „ortoheksagonalnej”. Okazało się, że trzy różne możliwości wyboru takiej komórki (Rys. 6.) nie są równoważne, ponieważ tylko dla jednej z nich statystyka intensywności refleksów wykonana programem *POINTLESS* (Evans, 2006) była akceptowalna. Komórka ta została następnie użyta podczas obróbki obrazów dyfrakcyjnych w rombowej grupie przestrzennej ( $C222_1$ ). Jednak mimo różnej wartości wskaźnika  $R_{\text{merge}}=6.8\%$ , wiele różnych prób rozwiązania i udokładnienia struktury w tej grupie przestrzennej nie dało pozytywnych efektów.



Rys. 6. Prymitywna jednostkowa komórka  $P$  (czerwona) z  $a=c$  i  $\alpha=120^\circ$  pokazana wraz z trzema możliwymi komórkami z centrowaniem  $C$  (niebieska, żółta i pomarańczowa) oraz upakowaniem dupleksów Z-DNA w płaszczyźnie  $ac$ , leżących na osiach  $2_1$  (**H8**). Rysunek został umieszczony na okładce lutowego (2016) numeru *Acta Crystallographica D*.

W celu rozwikłania problemu niejednoznaczności grupy przestrzennej, dogłębnie przebadaliśmy wszystkie aspekty związane ze skomplikowanym trybem zbliżeniowym oraz z translacyjnymi i rotacyjnymi pseudosymetriami, będącymi następstwem (pozornej) heksagonalnej metryki sieci. W tym celu przeanalizowaliśmy szczegółowo: (i) wszystkie statystyki intensywności refleksów (Yeates, 1997; Padilla & Yeates, 2003); (ii) natywne mapy Pattersona; (iii) wykresy funkcji

autorotacji (self-rotation function plots). Przeprowadziliśmy te wiele prób rozwiązania i udokładnienia struktury, zakładając różne możliwe grupy przestrzenne, stosując różne modele w procedurze podstawienia cząsteczkowego oraz tworząc skrypty do transformacji współrzędnych potencjalnych rozwiązań do innych grup przestrzennych i do zmiany sposobu wskazania refleksów (**H8**).

W wyniku powyższej analizy okazało się, że prawdziwą grupą przestrzenną kompleksu chimery DNA-RNA z jonami baru jest grupa  $P2_1$  z układu jednoskośnego z pseudoheksagonalną komórką elementarną z równymi parametrami  $a$  i  $c$  oraz kątem  $\sim 120^\circ$ , co w efekcie doprowadziło do bardzo rzadkiego i skomplikowanego sześciodomenowego zbliniaczenia. Trzy domeny zbliniaczenia wygenerowane są przez trójrotny operator zbliniaczenia  $l$ ,  $k$ ,  $-h-l$  (zastosowany dwukrotnie) z sumą ich (nierównych) ułamków zbliniaczenia wynoszącą  $\sim 50\%$ . Kombinacja z dodatkowym dwukrotnym pseudomeroedrycznym operatorem zbliniaczenia  $-h-l$ ,  $-k$ ,  $l$ , który generuje pozornie rombowy obraz dyfrakcyjny, daje kolejne trzy domeny, których sumaryczny wkład do dyfrakcji jest również zbliniony do 50%. Innymi słowy, kryształ zawiera sześć domen zbliniaczenia z trójrotnym i dwukrotnym operatorem zbliniaczenia, które są wzajemnie prostopadłe (**H8**). Podobny rodzaj zbliniaczenia zaobserwowano tylko w kilku przypadkach w kryształach małych cząsteczek (Fabry *et al.*, 1994, 1997; Friese *et al.*, 2003; Langer *et al.*, 2004) i w jednym kryształach makromolekuły (oksydoreduktaza akylcyanomycyny; AknOx; Sultana *et al.*, 2007). W tym ostatnim przypadku autorzy w podsumowaniu stwierdzili, że udokładnienie struktury w tak skomplikowanym przypadku nie jest możliwe a obecność wielodomenowego zbliniaczenia kryształu AknOx nie może być jednoznacznie potwierdzona.

Przecząc tej opinii nasze wyniki zaprezentowane w pracy **H8**, gdzie udowodniliśmy, że przy użyciu najnowszych technik krystalograficznych w połączeniu z opracowanymi dla takiego przypadku odrębnymi protokołami postępowania, możliwe jest jednoznaczne określenie grupy przestrzennej oraz precyzyjne wysokorozdzielcze udokładnienie ( $R_{\text{work}}=11.36\%$ ;  $R_{\text{free}}=16.91\%$ ) struktury nawet dla kryształów z bardzo skomplikowanym zbliniaczeniem (z sześcioma lub nawet dwunastoma domenami), z wieloma operatorami zbliniaczenia oraz ze skomplikowaną pseudosymetrią. Proces udokładniania przeprowadzono przy użyciu zarówno programu *SHELXL* jak i *REFMAC5* oraz dokonano porównania możliwości, zalet i niedogodności każdego z nich. Praca **H8**, opublikowana w *Acta Crystallographica D* została umieszczona na liście artykułów wyróżnionych, a jeden z rysunków z tej pracy został umieszczony na okładce lutego (2016) numeru tego czasopisma.



## 5. Walidacja i standardowe parametry stereochemiczne

W krystalografii makromolekuł rozdzielczość danych dyfrakcyjnych z reguły nie jest wysoka, szczególnie dla dużych białek lub kompleksów (średnia dla struktur zdeponowanych w bazie PDB wynosi  $\sim 2 \text{ \AA}$ ) i nierzadko liczba zarejestrowanych refleksów jest porównywalna lub niewiele większa od liczby parametrów opisujących strukturę. Aby zwiększyć liczbę obserwacji przypadających na jeden parametr, korzysta się z informacji stereochemicznych o poszczególnych znanych elementach struktury, takich jak aminokwasy w białkach czy nukleotydy w kwasach nukleinowych (Engh & Huber, 2001; Konnert & Hendrickson, 1980), traktując je jako dodatkowe obserwacje.

Mimo nieustających dyskusji (**H9**; **H10**; **H11**; Koepke *et al.*, 2003; Esposito *et al.*, 2002; Addlagatta *et al.*, 2001; Esposito *et al.*, 2000; Stec, 2007; Tickle, 2007), większe struktury makromolekuł jest udokładniania (Evans, 2007) z większymi opracowanymi na początku lat 90. w oparciu o struktury małych cząsteczek (Engh & Huber, 1991, 2001; Clowney *et al.*, 1996; Gelbin *et al.*, 1996; Parkinson *et al.*, 1996). Statystyczna analiza parametrów geometrycznych struktur o bardzo wysokiej rozdzielczości i dużej precyzji określonych w ostatnich latach, np. takich jak opisane w pracach (**H6**; **H7**; **H8**; Brzezinski *et al.*, 2011) może mieć duże znaczenie dla stworzenia bibliotek bardziej precyzyjnych wzorów, które będą lepiej odzwierciedlały specyfikę makrocząsteczek. Takie wzory mogłyby być później wykorzystywane do udokładniania struktur o niskiej lub średniej rozdzielczości oraz do walidacji geometrycznych parametrów modeli zdeponowanych w strukturalnych bazach danych. Należy przy tym zwrócić uwagę na fakt, że liczba struktur makrocząsteczek określonych z rozdzielczością atomową ( $d_{min} < 1.2 \text{ \AA}$ ) zdeponowanych w bazie PDB, w czasie ostatnich  $\sim 20$  lat, wzrosła dramatycznie od kilku na początku lat 90. do prawie 3 tys. obecnie.

Analiza danych geometrycznych (**H9**) wykazała dla niektórych parametrów istotne różnice pomiędzy obserwowanymi długościami wiązań i kątów pojedynczych aminokwasów a wartościami znajdującymi się w najczęściej używanej bibliotece białkowej (Engh & Huber, 2001). Przykłady odstępstw od planarności wiązania peptydowego, idealnych długości wiązania C–N i wartości kątów N–C–C można znaleźć w wielu pracach, w których analizowane są parametry geometryczne struktur wysokorozdzielczych (Esposito *et al.*, 2002; Addlagatta *et al.*,



2001; Esposito *et al.*, 2000). Analiza wybranych parametrów ła cucha głównego w uporządkowanych fragmentach struktur wysokorozdzielczych wykazała, że niektóre z obecnie używanych wartości geometrycznych powinny być uaktualnione lub stosowane z większą tolerancją. Np. kąt walencyjny N–C–C ma szeroki zakres wartości z bimodalnym rozkładem skorelowanym ze strukturą drugorzędową białka. Szczególnie dotyczy to struktur o najwyższej rozdzielczości, gdzie dominują cała rol mają dane dyfrakcyjne, a w uporządkowanych obszarach struktury akceptowalne są większe odchylenia od geometrii idealnej (**H9**; **H10**; **H11**). W pracy **H9** została przedstawiona korekta wartości geometrycznych stosowanych w programach do udokładniania, oparta na analizie statystycznej danych pochodzących z dobrze uporządkowanych fragmentów struktur o najwyższej rozdzielczości zgromadzonych w ostatnich latach w bazach strukturalnych PDB i CSD (Cambridge Structural Database; Allen, 2002). W pracy tej przedstawiliśmy również, na podstawie powyższej analizy i w oparciu o własne doświadczenia autorów, próby zdefiniowania racjonalnie uzasadnionych wartości średnich odchyłek kwadratowych (r.m.s.d.) udokładnianych parametrów od ich wartości idealnych z bibliotek stereochemicznych oraz stwierdziliśmy, że w wielu przypadkach stosowane wartości są zbyt restrykcyjne w dobrze określonych fragmentach makromolekuły, a fragmenty labilne i nieuporządkowane często wymagają znacznie mocniejszych wartości w celu utrzymania akceptowalnych parametrów stereochemicznych. Praca ta wywołała szeroką dyskusję w środowisku krystalograficznym, co ilustrują dwa opublikowane artykuły polemiczne (Stec, 2007; Tickle, 2007), które ukazały się w następnym zeszycie *Acta Crystallographica D* wraz z pracą **H10**, w której w bardziej szczegółowy sposób zostały wyjaśniliśmy ponownie niektóre aspekty poruszane w poprzedniej pracy. Rok później Karplus *et al.*, (2008) w pierwszym zdaniu swojego przełomowego artykułu stwierdzili, że praca **H9** „...initiated a very important discussion about the accuracy of ideal geometry targets and the appropriate stringency with which they should be obeyed at various resolutions” oraz potwierdzili postulaty zasugerowane w pracach **H9** i **H10**. Autorzy ci zaproponowali zastąpienie „twardych” wartości opartych na „pojedynczej idealnej wartości” przez wartości oparte na zależności funkcyjnej, której wartości byłyby różne w zależności od lokalnej konformacji, zgodnie z sugestiami zawartymi w pracach **H9** i **H10**. Bezpośrednim następstwem tych dyskusji było opracowanie zależności od konformacji biblioteki parametrów stereochemicznych dla ła cucha polipeptydowego (conformation-dependent library, CDL) (Berkholz *et al.*, 2009), następnie przetestowanej programami *TNT* (Tronrud *et al.*, 2010) i *SHELXL* (Tronrud *et al.*, 2011). Kolejnym krokiem było włączenie biblioteki CDL do pakietu *PHENIX* jako standardowej opcji udokładniania (Moriarty *et al.*,

2014, Moriarty *et al.*, 2016). W ten sposób częściowo zostały zrealizowane nasze postulaty sugerowane w pracach (**H9**; **H10**; **H11**).

## 6. Podsumowanie

W niniejszym autoreferacie omówiłem moje osiągnięcia naukowe zatytułowane „Aspekty metodologiczne wysokorozdzielczej krystalografii makromolekuł”, przedstawione w cyklu 11 (**H1-H11**) oryginalnych artykułów naukowych. Wszystkie prace wchodzące w skład cyklu zostały pozytywnie zrecenzowane przez niezależnych recenzentów wybranych przez edytorów poszczególnych czasopism. Powiniane tematycznie artykuły **H1-H11**, zawierają opis nowych metod, opracowanych protokołów lub optymalizacji już istniejących, które mają zastosowanie w kolejnych etapach określania trójwymiarowej struktury makromolekuł, począwszy od momentu uzyskania kryształu, poprzez rejestrację i przetwarzanie danych dyfrakcyjnych, rozwinięcie i udokładnienie struktury, aż do etapu końcowej walidacji poprawności otrzymanego modelu. Do najważniejszych moich osiągnięć opisanych w cyklu prac należą:

- Przygotowanie i przeprowadzenie pierwszej w pełni zdalnej sesji rejestracji danych dyfrakcyjnych, sterowanej z Poznania, w największym europejskim ośrodku synchrotronowym ESRF w Grenoble. W następstwie tego wydarzenia zostałem zaproszony do udziału w pracach grupy konsultantów europejskiego projektu *EDNA* oraz do wygłoszenia kilku krajowych i międzynarodowych wykładów (**H1**).
- Opracowanie bardziej korzystnych i efektywnych protokołów przetwarzania surowych danych dyfrakcyjnych, pozwalających na optymalne wskazanie i skalowanie obrazów dyfrakcyjnych dla trudnych lub niestandardowych przypadków, np. danych pochodzących od kryształów zbliżonych lub o ekstremalnie wysokiej rozdzielczości (**H2**).
- Rozwinięcie, udokładnienie i analiza pierwszej struktury monomeru proteazy retrowirusowej przy pomocy modelu zbudowanego przez tysiące graczy internetowej gry komputerowej *Foldit*. Był to pierwszy przypadek wykorzystania potencjału, intuicji i zdolności graczy gry komputerowej do rozwinięcia rzeczywistego, trudnego i ważnego problemu naukowego. Szczegółowa analiza struktury monomeru proteazy retrowirusowej daje podstawy do dalszych badań mających na celu otrzymanie nowych, bardziej skutecznych leków pomagających zwalczać infekcje

retrowirusowe, w tym AIDS (**H3**; **H4**).

- Opracowanie protokołów zastosowania metod i procedur makromolekularnych do rozwiązywania trudnych przypadków struktur supramolekularnych, dla których zawodziły metody stosowane w krystalografii małych cząsteczek. Przetestowano skuteczność metody podstawienia cząsteczkowego w zależności od wielkości i jakości tego modelu oraz opracowano szczegółowy tryb postępowania w takich przypadkach, który może być bezpośrednio zastosowany przez innych badaczy (**H5**).
- Rozwiązanie, udokładnienie i analiza czterech struktur kompleksów (z dwu- i trójwartościowymi jonami metali) lewoskrętnego heksameru Z-DNA o sekwencji d(CGCGCG) oraz chimery DNA-RNA (dCrGdCrGdCrG), charakteryzujących się wyjątkowo wysoką rozdzielczością. Pozwoliło to na opracowanie protokołów udokładniania struktur wysokorozdzielczych, umożliwiając uzyskanie najwyższej jakości i dokładności otrzymanych wyników, które jednocześnie są źródłem precyzyjnych danych stereochemicznych. Na przykładzie struktury chimery DNA-RNA pokazano, że dedykowany protokół postępowania umożliwia jednoznaczne określenie grupy przestrzennej oraz wysokorozdzielcze udokładnienie struktury nawet w przypadku bardzo skomplikowanego zbliniaczenia (z sześcioma lub nawet dwunastoma domenami), z wieloma operatorami zbliniaczenia oraz skomplikowaną pseudosymetrią (**H6**; **H7**; **H8**).
- Analiza i korekta stereochemicznych wiązań (długości wiązań i wartości kątownic walencyjnych) stosowanych podczas udokładniania struktury białek, dokonana na podstawie informacji geometrycznych dostępnych w bazach strukturalnych, pochodzących ze struktur określonych z wysoką rozdzielczością i precyzją. Wyniki tej analizy zostały włączone do biblioteki wiązań zależnych od konformacji (CDL), która powstała dwa lata później w wyniku szerokiej dyskusji, którą zainicjowały prace **H9**, **H10**. Biblioteka ta została ostatnio włączona do programów udokładniających, co było również sugerowane w pracach **H9**, **H10**.
- Oprócz aspektów metodycznych, wikszość prac wchodzi w skład cyklu (**H3**-**H8**) wnoszący również duży wkład strukturalny do naszej wiedzy o budowie przestrzennej makromolekuł biologicznych, polegający na określeniu i dogłębnej analizie szeregu nowych, nieznanych do tej pory struktur białek i kwasów nukleinowych, które stanowią ważny element poznania mechanizmów

molekularnych, bardzo istotnych np. w procesie projektowania nowych leków. W monoautorskiej przeglądowej pracy **H11**, przygotowanej na zaproszenie redakcji, podsumowałem w skrócie aspekty związane z poszczególnymi etapami wysokorozdzielczej krystalografii makromolekuł, od momentu otrzymania kryształu aż do depozytowania danych w strukturalnej bazie danych.

Wydaje się, że liczba przypadków, w których użycie standardowo zaimplementowanych w programach krystalograficznych procedur nie bierze wystarczająco do pominięcia zakłócenia procesu określania struktury, bierze cięgle wzrastała, ponieważ wiele prostych problemów zostało już rozwiązanych, a nowe będą coraz trudniejsze. W takich sytuacjach, oraz wtedy, gdy mamy do czynienia z różnymi patologiami kryształu lub zbliżamy się do granicznych przypadków, np. ekstremalnie wysokich rozdzielczości, zawsze konieczne będzie do wyłączenia i ekspercka wiedza badacza, umożliwiająca niestandardowe podejście oraz modyfikacje używanych rutynowo procedur.

## **Literatura**

Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. & Zwart, P. H. (2010). *Acta Cryst.* **D66**, 213-221.

Addlagatta, A., Krzywda, S., Czapinska, H., Otlewski, J. & Jaskolski, M. (2001). *Acta Cryst.* **D57**, 649-663.

Allen, F. H. (2002). *Acta Cryst.* **B58**, 380-388.

Berkholz, D.S., Shapovalov, M.V., Dunbrack, R.L.Jr & Karplus, P.A. (2009). *Structure.* **17**, 1316-1325.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235-242.

Betova, A., Cipriani, F., Cusack, S., Delageniere, S., Gbadinho, J., Gordon, E.J., Guijarro, M., Hall, D.R., Larsen, S., Launer, L., Lavault, C.B., Leonard, G.A., Mairs, T., McCarthy, A., McCarthy, J., Meyer, J., Mitchell, E., Monaco, S., Nurizzo, D., Pernot, P., Pieritz, R., Ravelli, R.G.B., Rey, V., Shepard, W., Spruce, D., Stuart, D.I., Svensson, O., Theveneau, P., Thibault,

- X., Turkenburg, J., Walsh, M. & McSweeney, S.M. (2006). *Acta Cryst.* **D62**, 1162-1169.
- Bonneau, R., Tsai, J., Ruczinski, I., Chivian, D., Rohl, C., Strauss, C.E.M. & Baker, D. (2001). *Proteins Struct. Funct. Genet.*, **45**, 119.
- Bradley, P., Chivian, D., Meiler, J., Misura, K.M., Rohl, C.A., Schief, W.R., Wedemeyer, W.J., Schueler-Furman, O., Murphy, P., Schonbrun, J., Strauss, C.E.M. & Baker, D. (2003). *Proteins Struct. Funct. Genet.* **53**, 57.
- Brunzelle, J. S., Shafae, P., Yang, X., Weigand, S., Ren, Z. & Anderson, W. F. (2003). *Acta Cryst.* **D59**, 1138-1144.
- Brzezinski, K., Brzuszkiewicz, A., Dauter, M., Kubicki, M., Jaskolski, M. & Dauter, Z. (2011). *Nucleic Acids Res.* **39**, 6238.
- Burla, M. C., Caliandro, R., Camalli, M., Carrozzini, B., Cascarano, G. L., Giacovazzo, C., Mallamo, M., Mazzone, A., Polidori, G. & Spagna, R. (2012). *J. Appl. Cryst.* **45**, 357-361.
- Clowney, L., Jain, S.C., Srinivasan, A.R., Westbrook, J., Olson, W.K. & Berman, H.M. (1996). *J. Am. Chem. Soc.* **118**, 509–518.
- Cooper, S., Khatib, F., Treuille, A., Barbero, J., Lee, J., Beenen, M., Leaver-Fay, A., Baker, D. & Popovici, Z. (2010). *Nature* **466**, 756-760.
- Dauter, Z. (2003). *Acta Cryst.* **D59**, 2004-2016.
- Dauter, Z. & Jaskolski, M. Promieniowanie synchrotronowe w spektroskopii i badaniach strukturalnych. Wybrane zagadnienia (B. J. Kowalski, W. Paszkowicz, E. A. Gorlich, eds.), p.305, Krakow, 2011.
- DiMaio, F., Terwilliger, T. C., Read, R. J., Wlodawer, A., Oberdorfer, G., Wagner, U., Valkov, E., Alon, A., Fass, D., Axelrod, H. L., Das, D., Vorobiev, S. M., Iwata, H., Pokkuluri, P. R. & Baker, D. (2011). *Nature* **473**, 540–543.
- Drozdal, P., Gilski, M. & Jaskolski, M. (2016). *Nucl. Acids Res.* (submitted).
- Echols, N., Moriarty, N. W., Klei, H. E., Afonine, P. V., Bunkóczi, G., Headd, J. J., McCoy, A. J., Oeffner, R. D., Read, R. J., Terwilliger, T. C. & Adams, P. D. (2014) *Acta Cryst.* **D70**, 144-154.
- Engh, R. & Huber, R. (1991). *Acta Cryst.* **A47**, 392–400.

- Engh, R. A. & Huber, R. (2001). *International Tables for Crystallography*, Vol. F, edited by M. G. Rossmann & E. Arnold, pp. 382-392. Dordrecht: Kluwer Academic Publishers.
- Esposito, L., Vitagliano, L., Zagari, A. & Mazzarella, L. (2000). *Protein Eng.* **13**, 825-828.
- Esposito, L., Vitagliano, L. & Mazzarella, L. (2002). *Protein Pept. Lett.* **9**, 95-105.
- Evans, P. (2006). *Acta Cryst.* **D62**, 72-82.
- Fabry, J., Breczewski, T. & Madariaga, G. (1994). *Acta Cryst.* **B50**, 13-22.
- Fabry, J., Petricek, V., Vanek, P. & Cisarova, I. (1997). *Acta Cryst.* **B53**, 596-603.
- Friese, K., Kienle, L., Duppel, V., Luo, H. & Lin, C. (2003). *Acta Cryst.* **B59**, 182-189.
- Fu, Z.-Q., Rose, J. & Wang, B.-C. (2005). *Acta Cryst.* **D61**, 951-959.
- Gelbin, A., Schneider, B., Clowney, L., Hsieh, S.-H., Olson, W.K. & Berman, H.M. (1996). *J. Am. Chem. Soc.* **118**, 519-529.
- Gilski, M. (2007). *Synchr. Radiat. Nat. Sci.* **6**, 95-98.
- Guillot, B., Gilski, M., Pyziak, M., Brzezinski, K., Claudot, J., Dauter, Z., Jaskolski, M. & Kubicki M. (2016). *Nucl. Acids Res.* (in preparation).
- Holton, J. & Alber, T. (2004). *Proc. Natl Acad. Sci. USA*, **101**, 1537-1542.
- Incardona, M.-F., Bourenkov, G. P., Levik, K., Pieritz, R. A., Popov, A. N. & Svensson, O. (2009). *J. Synchrotron Rad.* **16**, 872-879.
- Jaskolski, M. & Gilski, M. (2007). *Academia* **3**, 8-11.
- Kabsch, W. (2010). *Acta Cryst.* **D66**, 125-132.
- Kiss, G., Çelebi-Ölçüm, N., Moretti, R., Baker, D. & Houk, K. N. (2013). *Angew. Chem. Int. Ed.* **52**, 5700-5725.
- Koepke, J., Scharff, E.I., Lucke, C., Ruterjans, H. & Fritzsche, G. (2003). *Acta Cryst.* **D59**, 1744.
- Konnert, J. H. & Hendrickson, W. A. (1980). *Acta Cryst.* **A36**, 344-350.
- Krug M., Weiss M. S., Mueller U. & Heinemann U. (2012). *J. Appl. Cryst.* **45**, 568-572.
- Kuberski, B. & Szumna, A. (2009). *Chem. Commun.*, 1959-1961.

- Langer, V., Smrcok, L. & Masuda, Y. (2004). *Acta Cryst.* **C60**, i104-i106.
- Lebedev, A. A., Vagin, A. A. & Murshudov, G. N. (2006). *Acta Cryst.* **D62**, 83–95.
- Lesk, A.M., Lo Conte, L. & Hubbard, T.J. (2001). *Proteins Struct. Funct. Genet.*, **S5**, 98.
- Leslie, A.G.W. (1992). *Jnt CCP4/ESF-EACBM Newsl. Protein Crystallogr.* **26**.
- Leslie, A. G. W., Powell, H. R., Winter, G., Svensson, O., Spruce, D., McSweeney, S., Love, D., Kinder, S., Duke, E. & Nave, C. (2002). *Acta Cryst.* **D58**, 1924–1928.
- Malinina, L., Tereshko, V., Ivanova, E., Subirana J. A., Zarytova, V. & Nekrasov, Y. (1998). *Biophys. J.* **74**, 2482–2490.
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658-674.
- Miller, R., Gallo, S. M., Khalak, H. G. & Weeks, C. M. (1994). *J. Appl. Cryst.* **27**, 613-621.
- Minor, W., Cymborowski, M., Otwinowski, Z. & Chruszcz, M. (2006). *Acta Cryst.* **D62**, 859-866.
- Moriarty, N.W., Tronrud, D.E., Adams, P.D. & Karplus, P.A. (2014). *FEBS J.* **281**, 4061–4071.
- Moriarty, N.W., Tronrud, D.E., Adams, P.D. & Karplus, P. A. (2016). *Acta Cryst.* **D72**, 176-179.
- Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* **D67**, 355-367.
- Oeemig, J. S., Zhou, D., Kajander, T., Wlodawer, A., & Iwai, H. (2012). *J. Mol. Biol.*, **421(1)**, 85–99.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307-312.
- Padilla, J. E. & Yeates, T. O. (2003). *Acta Cryst.* **D59**, 1124-1130.
- Palatinus, L. & Chapuis, G. (2007). *J. Appl. Cryst.* **40**, 786-790.
- Panjikar, S., Parthasarathy, V., Lamzin, V. S., Weiss, M. S. & Tucker, P. A. (2005). *Acta Cryst.* **D61**, 449-457.
- Parkinson, G., Vojtechovsky, J., Clowney, L., Brünger, A.T. & Berman, H.M. (1996). *Acta Cryst.* **D52**, 57–64.

Pflugrath, J. W. (1999). *Acta Cryst.* **D55**, 1718-1725.

Rossmann M.G. & Blow, D.M. (1962). *Acta Cryst.* **15**, 24-31.

Sheldrick, G. M. (2008). *Acta Cryst.* **A64**, 112-122.

Sheldrick, G. M. (2010). *Acta Cryst.* **D66**, 479-485.

Sinigharoy, A., Venkatakrishnan, B., Liu, Y., Mayne, C.G., Lee, S., Chen, C.H., Zlotnick, A., Schulten, K., Flood, A.H. (2015). *J. Am. Chem. Soc.* **137**(27), 8810–8818.

Stec, B. (2007). *Acta Cryst.* **D63**, 1113-1114.

Stevens, R. & Wilson, I. (2001). *Science* **293**, 519-520.

Sultana, A., Alexeev, I., Kursula, I., Mantsala, P., Niemi, J. & Schneider, G. (2007). *Acta Cryst.* **D63**, 149-159.

Svensson, O. (2016). *Private communication*.

Szymanski, M., Wierzbicki, M., Gilski, M., Jedrzejewska, H., Sztylko, M., Cmoch, P., Shkurenko, O., Jaskolski, M. & Szumna A. (2016). *Chem. Eur. J.* **22**, 3148-3155.

Tickle, I. J. (2007). *Acta Cryst.* **D63**, 1274-1281.

Tronrud, D. E., Berkholtz, D. S. & Karplus, P. A. (2010). *Acta Cryst.* **D66**, 834-842.

Tronrud, D. E. & Karplus, P. A. (2011). *Acta Cryst.* **D67**, 699-706.

Veverka, V., Bauerová, H., Zábanský, A., Lang, J., Ruml, T., Pichová, I. & Hrabal, R. (2003). *J. Mol. Biol.* **333**, 771-780.

Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. (2007). *Methods Mol. Biol.* **364**, 215-230.

Wlodawer, A. (2007). *Private communication*.

Yeates, T. O. (1997). *Methods Enzymol.* **276**, 344-358.

Zhu Y., Zhu L.N., Guo R., Cui H.J., Ye S. & Fang Q. (2014). *Sci Rep.* **4**, 5046.

