

PRACOWNIA BIOFIZYKI DLA ZAAWANSOWANYCH

Ćwiczenia laboratoryjne dla studentów III roku kierunku
„Zastosowania fizyki w biologii i medycynie”
Biofizyka molekularna

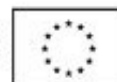
KRYSTALOGRAFIA RENTGENOWSKA WYZNACZANIE STRUKTUR BIAŁEK (ćw. 36)



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI



UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



Spis treści

1. Podstawy krystalografii rentgenowskiej.....	3
1.1. Dyfrakcja promieniowania na kryształach.....	3
1.2. Problem fazowy.....	5
1.3. Podstawienie molekularne.....	6
1.4. Mapy gęstości elektronowej.....	7
1.5. Dopasowanie modelu i udokładnianie struktury.....	7
1.6. Kontrola jakości(Validation).....	7
2. Budowa kryształów.....	8
3. Budowa dyfraktometru.....	10
3.1. Źródło.....	10
3.2. Detektor.....	11
3.3. Goniometr.....	11
4. Pomiary dyfrakcji kryształów białek.....	12
4.1. Przygotowanie kryształów do pomiaru.....	12
4.2. Zebranie danych dyfrakcyjnych.....	12
5. Rozwiązanie struktury białka – część obliczeniowa.....	13
5.1. Od obrazów dyfrakcyjnych – do czynników struktury (Data reduction).....	13
5.1.1. Indeksowanie obrazów(Indexing).....	13
5.1.2. Skalowanie i scalanie (Scaling & Merging).....	14
5.1.2.1. Pointless.....	14
5.1.2.2. Scala.....	14
5.1.2.3. cTruncate.....	15
5.2. Rozwiązanie problemu faz metodą podstawienia molekularnego.....	15
5.2.1. Tworzenie modelu (Model Building).....	15
5.2.2. Przygotowanie danych do podstawienia molekularnego.....	15
5.2.2.1. Cell Content Analysis (Matthews coefficient).....	15
5.2.2.2. Analyse data for MR.....	15
5.2.3. Podstawienie molekularne.....	16
5.3. Udokładnianie struktury.....	16
5.3.1. Refmac 5.....	16
5.3.2. Coot.....	17
5.4. Kontrola jakości dopasowania (Structure Validation).....	17
5.5. Prezentacja wyników.....	17
6. Bibliografia.....	17
7. Literatura uzupełniająca.....	18
8. Zagadnienia do kolokwium wstępnego.....	18
9. Wykonanie ćwiczenia.....	18

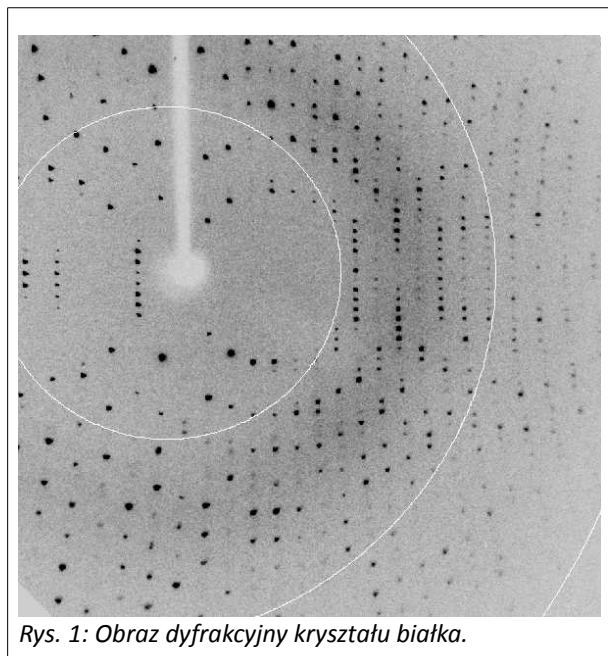
1. Podstawy krystalografii rentgenowskiej

Badanie obiektów przy pomocy promieniowania elektromagnetycznego, wymaga zastosowania fal o długościach porównywalnych z rozmiarem obserwowanego obiektu. Rozmiary atomów i długości wiązań między nimi mieszczą się w przedziale około 1 - 3,5 Å. 1Å wynosi średnica atomu wodoru, 2-4 – 3,5Å długości mają wiązania wodorowe, typowe długości wiązania C-C w cząsteczkach białka mieszczą się w przedziale 1,3-1,6Å[1]. Badanie obiektów takich rozmiarów wymaga zastosowanie promieniowania rentgenowskiego.

Rezultatem eksperymentu krystalograficznego nie jest jednak obraz atomów, lecz mapa gęstości elektronowej w cząsteczce. Dzieje się tak dlatego, że promieniowanie elektromagnetyczne oddziałuje z materią poprzez pole elektromagnetyczne. Intensywność rozproszonego promieniowania jest proporcjonalna do stosunku ładunku do masy, zatem elektrony, które są kilka tysięcy razy lżejsze od protonów i jąder atomowych oddziałują z promieniowaniem elektromagnetycznym dużo silniej. Ponieważ prędkość elektronów jest wielokrotnie większa od szybkości zmian pola elektrycznego obserwuje się nie pojedyncze elektrony, lecz uśredniony w czasie rozkład elektronów w atomie. Ponieważ elektrony są skupione wokół jąder atomowych i wiązań gęstość elektronowa dość dobrze odwzorowuje kształt cząsteczki.

Rozpraszanie promieniowania przez pojedynczą cząsteczkę jest bardzo słabe, zatem trudne do wykrycia i zmierzenia ponad poziomem szumów, pochodzących od rozpraszania przez cząsteczki wody i powietrza. W kryształach duża liczba cząsteczek jest uporządkowana w przestrzeni, więc rozproszone przez poszczególne cząsteczki promieniowanie jest zgodne w fazie, zatem zachodzi konstruktywna interferencja, i intensywność refleksów dodaje się tworząc falę o mierzalnej intensywności. Kryształ działa jak wzmacniacz.

Oczywiście jeżeli w pewnych kierunkach dojdzie do wzmocnienia intensywności promieniowania odbitego na skutek interferencji, to w innych musi dojść do wygaszenia, dlatego obraz dyfrakcji kryształu jest nieciągły i składa się z pojedynczych kropek.



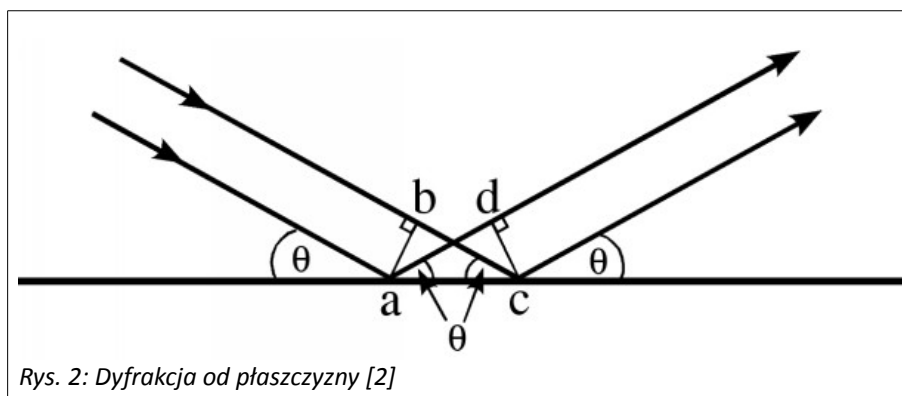
Rys. 1: Obraz dyfrakcyjny kryształu białka.

1.1. Dyfrakcja promieniowania na kryształach

Gdy fala elektromagnetyczna rozprasza się na elektronach cząsteczek tworzących kryształ fale ugięte na poszczególnych elektronach interferują ze sobą. W zależności od wzajemnego położenia elektronów w przestrzeni i kąta padania fali efektem interferencji może być wzmocnienie fali, wygaszenie fali, lub coś pomiędzy. Co się stanie zależy od całkowitej drogi, jaką fala

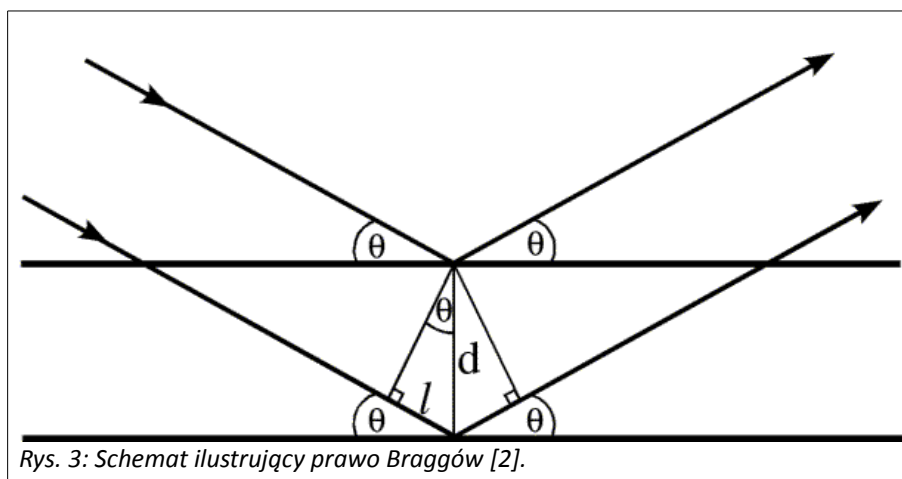
przebywa od źródła do detektora. Jeżeli różnica długości dróg dla dwóch fal rozproszonych na dwóch różnych elektronach jest wielokrotnością długości fali, wtedy fale rozproszone będą zgodne w fazie i na skutek interferencji nastąpi wzmocnienie fali, jeżeli różnica dróg będzie wielokrotnością fali plus pół długości fali, wtedy nastąpi wygaszenie fali. Warunek na konstruktywną interferencję można łatwo wyprowadzić, wyobrażając sobie, że fale objają się od płaszczyzny przechodzącej przez atomy, tę płaszczyznę nazywa się płaszczyzną Bragga, można o niej myśleć, jak o lustrze od którego odbija się promieniowanie.

Kiedy równoległa wiązka promieniowania odbija się od lustra kąt padania jest równy kątowi odbicia, to samo dotyczy płaszczyzn Bragga, a jeśli padająca wiązka jest w fazie, to wiązka odbita też jest w fazie, niezależnie od miejsca, w którym odbiły się poszczególne fale. Schemat przedstawiony na rys. 2 Wyjaśnia dlaczego.



Padająca wiązka jest w fazie, to znaczy, że w punktach a i b obie padające fale mają tę samą fazę. Odcinki bc i ad są sobie równe, ponieważ trójkąty abc i acd są przystające. Zatem promienie odbite w punktach a i c przebywają tę samą drogę, więc wiązka odbita od płaszczyzny jest zgodna w fazie.

Jeżeli fale odbite od jednej płaszczyzny muszą mieć identyczne drogi, żeby być w fazie, to fale odbite od różnych płaszczyzn, muszą mieć różne drogi, a różnica dróg musi być równa wielokrotności długości fali. Prawo Braggów mówi, w jakich odległościach od siebie muszą znajdować się płaszczyzny rozpraszania, aby rozproszona wiązka była w fazie.



Różnica długości dróg fal odbitych od różnych płaszczyzn wynosi $2l$ (rys. 3) i zależy od kąta padania, $l = d \sin \theta$, zatem prawo Braggów przyjmuje postać:

$$n\lambda = 2d \sin \theta \quad (1)$$

Z prawa Braggów wynika, że wraz ze wzrostem kąta rozpraszania, odległość między płaszczyznami Bragga, czyli d , musi maleć, żeby różnica dróg optycznych była równa długości fali. To oznacza, że im większy jest kąt rozpraszania tym mniejsze są rozmiary detali, które możemy „zobaczyć” przy pomocy pomiarów dyfrakcyjnych.

Ta odwrotna proporcjonalność powoduje, że dane dyfrakcyjne analizuje się zwykle w tak zwanej przestrzeni odwrotnej (reciprocal space). Im dalej od siebie są obiekty w przestrzeni odwrotnej, tym bliżej siebie są one w przestrzeni

rzeczywistej, pod tym większym kątem rejestruje się refleksy od tych obiektów i tym czulszy na drobne szczegóły jest obraz dyfrakcyjny.

Jak wspomniano wcześniej, kryształ działa jak wzmacniacz promieniowanie rozproszonego, dzięki zjawisku interferencji. Żeby doszło do konstruktywnej interferencji promieniowanie rozproszone na poszczególnych komórkach elementarnych kryształu musi być zgodne w fazie, zatem płaszczyzny Bragga muszą przebiegać przez te same punkty, we wszystkich komórkach elementarnych kryształu.

Jeżeli obiekty znajdujące się na płaszczyźnie Bragga rozpraszają w fazie, to obiekty znajdujące się między tymi płaszczyznami będą rozpraszać nie w fazie, a przesunięcie fazowe będzie proporcjonalne do odległości obiektu od płaszczyzny. Zatem pojedynczy eksperyment dyfrakcyjny pozwala wyznaczyć względne odległości obiektów od płaszczyzn Bragga. Jeżeli wszystkie obiekty będą znajdowały się na płaszczyznach Bragga, na obrazie dyfrakcyjnym pojawi się jedna plamka. Jeśli połowa obiektów będzie się znajdowała na płaszczyznach Bragga, a druga połowa, na płaszczyznach równoległych, dokładnie w połowie odległości między płaszczyznami Bragga, to te dwa zbiory obiektów, nie będą rozpraszały w fazie, a przesunięcie fazowe od drugiego zbioru obiektów, wyniesie dokładnie pół długości fali, zatem nastąpi destruktywna interferencja. Na obrazie dyfrakcyjnym nic się nie zmieni. Jeżeli druga połowa obiektów będzie się znajdowała w innej odległości niż $d/2$, przesunięcie fazowe rozproszonego promieniowania będzie gdzieś pomiędzy 0 a $\lambda/2$, zatem fala rozproszona wniesie swój przyczynek do interferencji, na obrazie dyfrakcyjnym zmieni się intensywność plamki.

Obraz dyfrakcyjny jest związany z obiektem, na którym promieniowanie się rozproszyło poprzez transformację Fouriera. Zatem, traktując gęstość elektronową jako funkcję matematyczną, obraz dyfrakcyjny jest transformacją Fouriera gęstości elektronowej. Większość funkcji matematycznych, ma swoje funkcje odwrotne np sinus i arcus sinus, transformata Fouriera nie jest wyjątkiem. To oznacza, że gęstość elektronowa jest odwrotną transformacją Fouriera obrazów dyfrakcyjnych. Zatem wystarczy tylko policzyć?¹

Niestety nie tak prosto. Żeby policzyć gęstość elektronową trzeba znać amplitudę i **fazę** fal rozproszonych na tej gęstości elektronowej. Ale to co jest mierzone podczas eksperymentu, to liczba fotonów padających w każdy punkt ekranu. Liczba fotonów wyznacza natężenie danego refleksu, a natężenie jest proporcjonalne do amplitudy padającej fali, ale nie istnieje, żadna bezpośrednia metoda eksperymentalna pozwalająca wyznaczyć fazy padających fal. Jest to tak zwany **problem fazowy**².

1.2. Problem fazowy

Poza uzyskaniem dobrze rozprasającego kryształu białka, problem fazowy jest często, drugim wąskim gardłem w krystalografii białek.

Żeby zrozumieć, dlaczego fazy „znikają” w czasie eksperymentu dyfrakcyjnego, nie wystarczy falowy obraz światła. Dyfrakcja, to również rozpraszanie fotonów, czyli zderzenia fotonu z elektronem. W mechanice kwantowej, prawdopodobieństwo, że foton zostanie zostanie „odbity” w określonym kierunku jest dane kwadratem amplitudy rozproszonych fal.

$$\Psi(x, t) = A \cdot e^{2\pi i(vt - x/\lambda)} \text{ - funkcja falowa fotonu}$$

$$P_{a < x < b} = \int_a^b \Psi(x, t) \cdot \Psi^*(x, t) dx = \int_a^b |\Psi(x, t)|^2 dx \text{ - prawdopodobieństwo, że foton znajduje się w obszarze } x \in (a, b)$$

$$|\Psi(x, t)|^2 = A \cdot e^{2\pi i(vt - x/\lambda)} \cdot A \cdot e^{-2\pi i(vt - x/\lambda)} = A^2$$

Dlatego na obrazie dyfrakcyjnym można zmierzyć tylko intensywność refleksu.

Skoro faz nie można zmierzyć bezpośrednio, trzeba je wydobyć z danych i pomiarów pośrednich. Istnieją dwie techniki

- 1 Żeby zobaczyć jak działa odwrotna transformacja Fouriera i co można „wydobyć” z obrazów dyfrakcyjnych można zajrzeć na stronę Kevina Cowtana: *The Interactive Structure Factor Tutorial* <http://www.ytbl.york.ac.uk/~cowtan/sfapplet/sfintro.html>
- 2 Dlaczego fazy są takie ważne można zrozumieć zaglądając na inną stronę Kevina Cowtana: *Kevin Cowtan's Book of Fourier* <http://www.ytbl.york.ac.uk/~cowtan/fourier/fourier.html>

uzyskania informacji o fazach, przez modyfikację badanej struktury i tym samym obrazu dyfrakcyjnego oraz przez zgadnięcie.

Ponieważ ciężkie atomy, a dużej liczbie elektronów silniej rozpraszają promieniowanie, podstawienie w cząsteczce białka np atomu siarki w grupie SH atomem rtęci zmienia obrazy dyfrakcyjne. Przez porównanie obrazów dyfrakcyjnych białka i jego izomorficznej pochodnej zawierającej atom lub atomy ciężkiego pierwiastka można uzyskać informacje o fazach ugiętych wiązek. Metodę tę wykorzystuje się w dwóch wariantach:

- wielokrotne podstawienie izomorficzne (Multiple Isomorphous Replacement MIR) – przebadanie kilku izomorficznych pochodnych z różnymi ciężkimi atomami

- anomalne rozproszenie w wielu długościach fali (Multiple Anomalous Dispersion MAD) – wykorzystanie jednej izomorficznej pochodnej, ale pomiary przy kilku długościach fali.

„Zgadnąć” fazy można wtedy, gdy znana jest struktura białka podobnego do białka badanego. Policzenie amplitud i faz promieniowania rozproszonego na cząsteczce o znanej strukturze nie jest problem. Policzone fazy przypisuje się badanemu białku jako początkowe fazy a następnie w rekurencyjnym procesie poprawia się je, tak aby odtworzyć zmierzoną gęstość elektronową. Metoda ta nazywana jest podstawieniem molekularnym.

1.3. Podstawienie molekularne

Podstawienie molekularne (*molecular replacement* MR), to technika wykorzystująca model. Jeżeli badane białko jest mutantem białka, którego struktura jest już znana, można wykorzystać tę strukturę jako model. Dla białka dzikiego krystalizowanego po raz pierwszy jako model można wykorzystać struktury białek należących do tej samej rodziny, albo białek homologicznych pochodzących z innych organizmów (na przykład ludzkie i cielęce PNP są do siebie bardzo podobne).

Jeśli żadna z powyższych metod nie jest dostępna, można zbudować model wychodząc od sekwencji badanego białka. Istnieje wiele baz danych zawierających sekwencje białek (UniProt, PIR, etc) i wiele narzędzi biofizycznych, które pozwalają przeszukiwać takie bazy pod kątem podobieństwa sekwencji i fragmentów sekwencji (PHYRE2, MSD, Blocks). Gdy uda się znaleźć białka lub domeny białkowe podobne do badanego obiektu, należy przeszukać bazy struktur białkowych (PDB) w poszukiwaniu trójwymiarowej struktury modelu lub jego fragmentów. Wykorzystanie metody podstawienia molekularnego stało się możliwe i coraz częściej stosowane wraz z rosnącą liczbą struktur białkowych zdeponowanych w bazach takich jak PDB. Obecnie około 70% struktur jest rozwiązywanych metodą podstawienia molekularnego [3].

O ile odtworzenie struktury trójwymiarowej obiektu na podstawie jego obrazów dyfrakcyjnych jest problematyczne (problem fazowy!) to „w drugą stronę” jest o tyle prosto, że wszystkie dane są znane: współrzędne atomów i kwantowo-mechaniczny opis zjawiska dyfrakcji promieniowania. Zatem można policzyć jak wygląda obraz dyfrakcji dowolnej cząsteczki białka o znanej strukturze. Efekty dyfrakcji przedstawia się w postaci czynników struktury. Czynniki struktury reprezentują fale, które uległy rozproszeniu na badanym obiekcie, zatem tak samo jak fale są opisane przez amplitudę i fazę. Dla cząsteczki modelu należy obliczyć czynniki struktury.

Teraz cząsteczkę modelu należy „włożyć” do komórki elementarnej badanego kryształu, to oznacza, że należy wyznaczyć jej położenie i orientację. I tu znów nie ma metody analitycznej pozwalającej na wykonanie tej operacji. Należy przetestować wszelkie możliwe położenia i orientacje modelu w komórce elementarnej, żeby odtworzyć jak najwierniej zmierzony wzór dyfrakcyjny. Aby określić położenie modelu trzeba wyznaczyć sześć parametrów: trzy kąty związane z obrotem cząsteczki i trzy wektory translacji określające jej położenie, jeżeli w komórce elementarnej znajduje się N cząsteczek białka, oznacza to konieczność wyznaczenia $6N$ parametrów. Przeszukanie sześciowymiarowej przestrzeni zajęłoby bardzo dużo czasu. Zadanie to można znacząco uprościć, jeżeli znalezienie położenia (translacja) i orientacji (rotacja) modelu zostanie rozdzielone na dwa poszukiwania w trójmiarowych przestrzeniach. Najpierw próbkowana jest przestrzeń rotacji i wybierane jest najlepsze rozwiązanie (jedno lub kilka) i tylko dla tych wybranych rozwiązań próbkowana jest przestrzeń translacji. W ten sposób znajdowane jest najlepsze dopasowanie modelu do danych eksperymentalnych, czyli do zbioru amplitud czynników struktury. Fazy poszczególnych czynników struktury policzonych dla modelu są przypisywane odpowiednim czynnikom struktury zmierzonym dla badanego kryształu białka. Początkowa mapa gęstości elektronowej jest liczona na podstawie zmierzonych

amplitud i faz przeniesionych z modelu.

1.4. Mapy gęstości elektronowej

Gdyby atomy w kryształach były zupełnie nieruchome, cząsteczki w jednakowej konformacji a kryształy idealnie uporządkowane, wtedy wszystkie rozproszone refleksy byłyby w fazie i jedynym ograniczeniem rozdzielczości danych dyfrakcyjnych byłaby długość fali użytego promieniowania. Niestety białka są dość elastycznymi cząsteczkami, kryształy mają defekty sieci krystalicznej. Dodatkowo im mniejsze detale chcemy obserwować, pod tym większymi kątami obserwuje się refleksy i tym są one słabsze. W efekcie tych wszystkich utrudnień typowe rozdzielczości uzyskiwanych map elektronowych nie pozwalają widzieć pojedynczych atomów, raczej rurkę gęstości elektronowej gdzie atomy nie są wyraźnie oddzielone od siebie.

Kolejną zgorą krystalografii białek jest niezbyt wysoki stosunek sygnału do szumu. Co daje jeszcze jeden wkład do znacznego „rozmycia” mapy gęstości elektronowej.

1.5. Dopasowanie modelu i udokładnianie struktury

Ponieważ mapy gęstości elektronowej rzadko mają rozdzielczość atomową, dopasowywanie modelu jest do pewnego stopnia arbitralne! Z tego powodu również, rozwiązanie struktury białka tylko na podstawie mapy gęstości elektronowej jest zwykle niemożliwe. Dlatego w krystalografii, przy rozwiązywaniu struktur wykorzystuje się wszystkie informacje o badanym białku i białkach w ogólności jakie tylko są osiągalne.

Podstawową informacją o badanym białku jest jego sekwencja aminokwasowa. Tylko dzięki jej znajomości można przypisać kolejne aminokwasy określonym „bąblom” gęstości elektronowej.

Budując i udokładniając modele stale korzysta się z ograniczeń na długości i kąty wiązań, znane dzięki temu, że wiele tysięcy struktur białek zostało już rozwiązanych. Te ograniczenia pozwalają w racjonalny sposób budować główny łańcuch polipeptydowy. Modelując łańcuchy boczne aminokwasów korzysta się dodatkowo z baz danych danych konformacji reszt aminokwasowych i częstotliwości ich występowania w białkach.

Udokładnianie struktury jest rekurencyjnym procesem, poprawiania, obliczanie rozbieżności, wprowadzania dalszych poprawek, aż do osiągnięcia założonego poziomu zgodności. Udokładnianie jest procesem, który nie ma końca, rezygnacja z dalszego udokładniania jest zawsze decyzją arbitralną.

1.6. Kontrola jakości(Validation)

Ostatnim etapem rozwiązywania struktury jest sprawdzenie jakości uzyskanej struktury. Ponieważ udokładnianie bywa arbitralne, a w rozmyte bąble gęstości elektronowej czasami można wfitować „wszystko”(ang. *overfitting*) potrzebne są narzędzia sprawdzające i weryfikujące strukturę, które będą niezależne od eksperymentatora.

Po pierwsze sprawdzono musi zostać jakość dopasowania struktury do danych dyfrakcyjnych, w tym celu liczone są wskaźniki rozbieżności R definiowane jako średnia różnica między amplitudami struktury obliczonymi dla modelu F^{calc} a wyznaczonymi doświadczalnie na podstawie zindeksowanych refleksów F^{exp} :

$$R = \frac{\sum_{hkl} ||F_{hkl}^{exp}| + |F_{hkl}^{calc}||}{\sum_{hkl} |F_{hkl}^{exp}|}$$

Im mniejsza wartość R tym lepsze dopasowanie modelu do danych dyfrakcyjnych. Struktury o $R < 0,2$ uznaje się za dobrze udokładnione. Do oceny, czy udokładniane parametry mają pełne oparcie statystyczne w informacji zebranej podczas eksperymentu oblicza się wskaźnik rozbieżności liczony dla refleksów wykluczonych z udokładniania R_{free} . Jego wartość nie powinna znacząco przekraczać wartości R .

Po drugie sprawdzana jest geometryczna poprawność uzyskanej struktury. Długości i kąty wiązań powinny zawierać się w

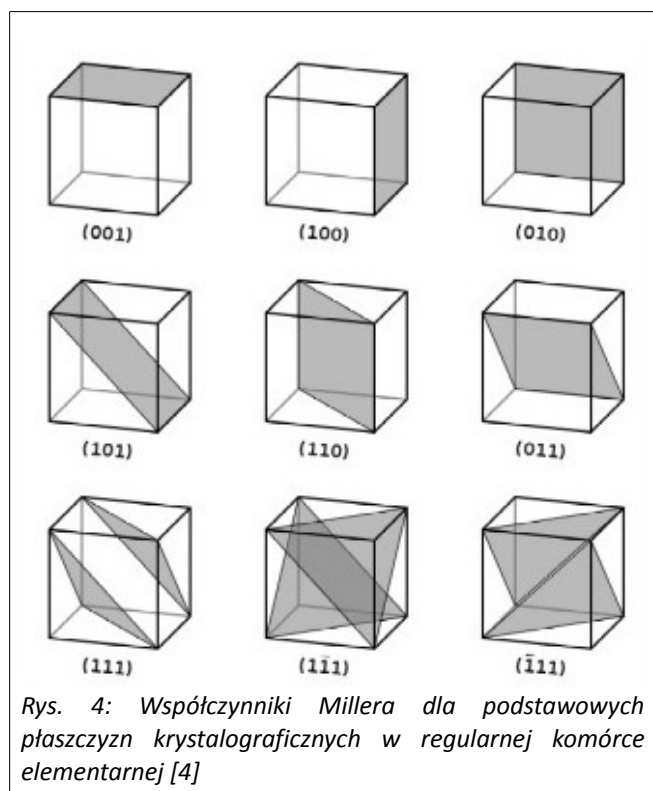
przedziałach typowych dla białek. Wiązania peptydowe muszą być płaskie, kąty ψ i θ zawierać się w określonych przedziałach, dlatego wiele programów wykreśla mapę Ramachandrana. Konformacje łańcuchów bocznych aminokwasów są porównywane z bazami danych. Sprawdzenie kontaktów van der Waalsa pozwala ocenić czy atomy nie są zbyt blisko siebie.

Kolejnym elementem kontroli może być sprawdzenie oddziaływań „środowiskowych”. Czy aminokwasy hydrofobowe oddziałują z innymi aminokwasami hydrofobowymi, a polarne z polarnymi lub wodą.

2. Budowa kryształów

W idealnym kryształcie cząsteczki ułożone są regularnie w przestrzeni tworząc trójwymiarową sieć. **Komórka elementarna** (*Unit cell*) to podstawowy element strukturalny powtarzający się w przestrzeni i tworzący sieć przestrzenną. Komórka elementarna jest równoległociąnem charakteryzowanym przez długości krawędzi: a, b, c i kąty między krawędziami: α, β, γ . Krawędzie komórek tworzą trójwymiarowe sieci zwane sieciami przestrzennymi. Długości i kierunki krawędzi komórki elementarnej definiują wektory $\vec{a}, \vec{b}, \vec{c}$, sieć rzeczywista jest rozpięta na tych trzech wektorach. Położenie dowolnego punktu w przestrzeni może być opisane wektorem translacji: $\vec{r}' = \vec{r} + n\vec{a} + m\vec{b} + l\vec{c}$, Punkty o współrzędnych (na, mb, lc) nazywa się węzłami sieci (gdzie n, m, l są dowolnymi liczbami całkowitymi).

Zespoły równoległych płaszczyzn Bragga opisujemy współczynnikami Millera: h, k, l . Wartości współczynników Millera wyznaczają punkty przecięcia z osiami układu współrzędnych tej z rodziny płaszczyzn, która przechodząc przez węzły sieci jest jednocześnie najbliższej środka układu odniesienia (rys. 4).



Sieć krystaliczna w przestrzeni odwrotnej jest również definiowana przez krawędzie komórki elementarnej: a^*, b^*, c^* i kąty między tymi krawędziami: $\alpha^*, \beta^*, \gamma^*$. Taktując a, b, c i a^*, b^*, c^* jako wektory, zależności między wektorami sieci rzeczywistej i odwrotnej są następujące:

$$\vec{a}^* \cdot \vec{a} = \vec{b}^* \cdot \vec{b} = \vec{c}^* \cdot \vec{c} = 1$$

$$\vec{a}^* \cdot \vec{b} = \vec{a}^* \cdot \vec{c} = \vec{b}^* \cdot \vec{c} = 0$$

Co oznacza, że a^* jest prostopadłe do b i do c a jego długość wynosi $1/a$.

Punkt, który w przestrzeni rzeczywistej opisany wektorem $\vec{r}' = n\vec{a} + m\vec{b} + l\vec{c}$, jest opisany wektorem

$\vec{H} = h\vec{a}^* + k\vec{b}^* + l\vec{c}^*$ w przestrzeni odwrotnej.

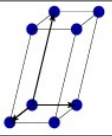
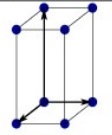
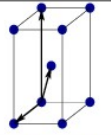
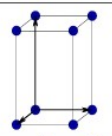
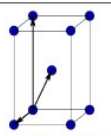
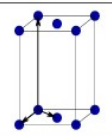
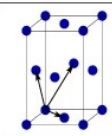
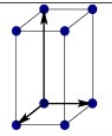
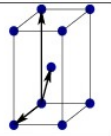
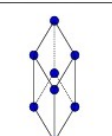
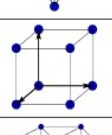
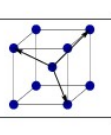
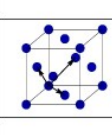
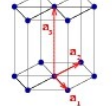
Symetria sieci rzeczywistej jest zachowana w sieci odwrotnej, zachowana jest nie tylko geometria, ale także intensywności rozproszonych refleksów.

Komórkę elementarną charakteryzuje sześć parametrów sieci: okresy identyczności sieci a, b, c (osie) oraz kąty α, β, γ . W zależności od kształtu komórki elementarnej wyróżniamy siedem układów krystalograficznych (Rys. 5)

- ≈ regularny (*cubic*)
- ≈ tetragonalny (*teragonal*)
- ≈ rombowy (*orthorombic*)
- ≈ heksagonalny (*hexagonal*)
- ≈ trygonalny (*trigonal*)
- ≈ jednoskośny (*monoclinic*)
- ≈ trójskośny (*triclinic*)

Sieci prymitywne, złożone z prymitywnych komórek (*primitive, P*) mają węzły tylko w wierzchołkach komórek. Sieci nieprymitywne mogą mieć węzły również na wszystkich ścianach (*face-centered, F*), dwóch przeciwległych ścianach (*side-centered, C*) lub pośrodku komórki (*body-centered, I*).

Siedem układów krystalograficznych wraz z czterema rodzajami komórek elementarnych tworzy czternaście sieci punktowych, sieci Bravais (Rys. 5)

Bravais lattice	Parameters	Simple (P)	Volume centered (I)	Base centered (C)	Face centered (F)
Triclinic	$a_1 \neq a_2 \neq a_3$ $\alpha_{12} \neq \alpha_{23} \neq \alpha_{31}$				
Monoclinic	$a_1 \neq a_2 \neq a_3$ $\alpha_{23} = \alpha_{31} = 90^\circ$ $\alpha_{12} \neq 90^\circ$				
Orthorhombic	$a_1 \neq a_2 \neq a_3$ $\alpha_{12} = \alpha_{23} = \alpha_{31} = 90^\circ$				
Tetragonal	$a_1 = a_2 \neq a_3$ $\alpha_{12} = \alpha_{23} = \alpha_{31} = 90^\circ$				
Trigonal	$a_1 = a_2 = a_3$ $\alpha_{12} = \alpha_{23} = \alpha_{31} < 120^\circ$				
Cubic	$a_1 = a_2 = a_3$ $\alpha_{12} = \alpha_{23} = \alpha_{31} = 90^\circ$				
Hexagonal	$a_1 = a_2 \neq a_3$ $\alpha_{12} = 120^\circ$ $\alpha_{23} = \alpha_{31} = 90^\circ$				

Rys. 5: Sieci Bravais

W kryształach białkowych dopuszczalne operacje symetrii to translacje, osie symetrii i osie śrubowe. Ponieważ w białkach występują tylko L-aminokwasy nie mogą pojawić się w kryształach białkowych takie operacje symetrii jak inwersja czy lustrzane odbicie.

W kryształach molekularnych komórki elementarne są zwykle dość duże i zawierają często więcej niż jedną cząsteczkę

białka, co oznacza, że w kryształach jest obiekt mniejszy od komórki elementarnej, który powielany zgodnie z operacjami symetrii tworzy kryształ. Najmniejszą częścią kryształu, która powtarza się w przestrzeni jest tzw. jednostka asymetrii (*asymmetric unit*, ASU).

Kryształy białek zawierają dużo wody, cząsteczki wody znajdują się zarówno wewnątrz cząsteczki białka jak i między nimi. Zawartość wody może wynosić nawet 50-60%.

W większości przypadków, kryształy białkowe nie są monokryształami, składają się z monokrystalicznych domen, które na ogół nie są ułożone idealnie, lecz lekko poprzesuwane i poprzekęcane względem siebie. Parametr, który określa stopień nieuporządkowania domen krystalicznych w białka to mozaikowość (*mosaicity*). Efektem mozaikowości są fizyczne rozmiary refleksów na obrazach dyfrakcyjnych i ich nieostre krawędzie. Gdyby kryształ był uporządkowany idealnie, na obrazie dyfrakcyjnym byłyby punkty.

3. Budowa dyfraktometru

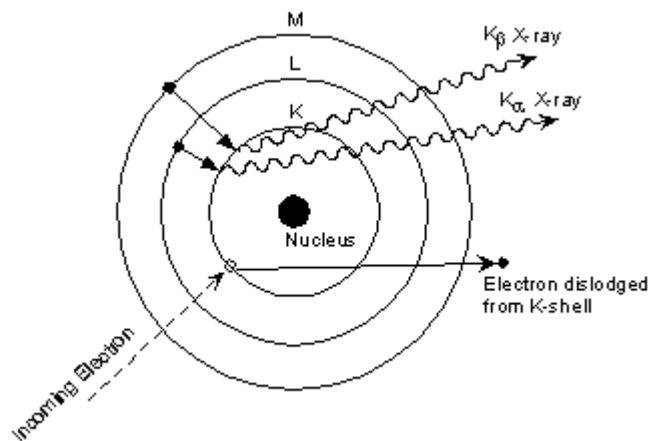
Do badań dyfrakcyjnych używa się różnego typu urządzeń, zaczynając od tzw. Kieszonkowych dyfraktometrów, a kończąc na wiązkach wytwarzanych w synchrotronach. Kluczowe elementy dyfraktometrów to źródło promieniowania rentgenowskiego, detektor przy pomocy którego rejestruje się obrazy dyfrakcyjne, goniometr, który pozwala na dowolne ustawienie kryształu względem wiązki promieniowania, oraz system kriogeniczny zapewniający utrzymanie kryształu we właściwej temperaturze. Każde z tych urządzeń wykonywane jest na wiele sposobów, rozwiązania zaprezentowane poniżej są wykorzystane są w dyfraktometrze SuperNova firmy Oxford Diffraction znajdującym się w Zakładzie Biofizyki.

3.1. Źródło

Źródłem promieniowania rentgenowskiego jest lampa z nieruchomą miedzianą anodą i mikroogniskowaniem wiązki.

Lampa rentgenowska, to próżniowa bańka w której umieszczone są katoda i anoda. Elektrony wyrwane są z żarzącej się katody i przyspieszane w polu elektrycznym. Napięcie przykładane między katodą i anodą jest rzędu kilkudziesięciu do ok 150kV. Wysokoenergetyczne elektrony uderzają w anodę, wybijając elektrony z wewnętrznych powłok. Elektrony wracając na swoje powłoki emitują kilka charakterystycznych długości fal, do badań białek wykorzystuje się przejście $K\alpha$, o długości fali 1,5418Å (Rys. 6).

Zaledwie ok 1% energii absorbowanej przez anodę jest oddawana jako promieniowanie, reszta jest oddawana na sposób ciepła, a schłodzenie anody w próżni nie jest sprawą trywialną. Stosowane są dwa rozwiązania: wirująca anoda lub mikroogniskowanie wiązki elektronów. W przypadku wirującej anody elektrony padają na coraz to inne miejsce anody, co ogranicza jej przegrzewanie się, w przypadku mikroogniskowania wiązkę elektronów starannie się kolimuje, dzięki czemu trafia ona w obszar anody o średnicy kilkudziesięciu mikrometrów, co pozwala obniżyć jej natężenie i w ten sposób ograniczyć przegrzewanie się anody. Dodatkową zaletą mikroogniskowania jest lepsza kolimacja wiązki promieniowania rentgenowskiego wychodzącego z anody, oraz obniżenie poboru mocy i wydłużenie czasu życia lampy rentgenowskiej. Promieniowanie następnie przechodzi przez układ optyczny, który dodatkowo kolimuje wiązkę, dając wiązkę o średnicy kilkudziesięciu – kilkuset mikrometrów.

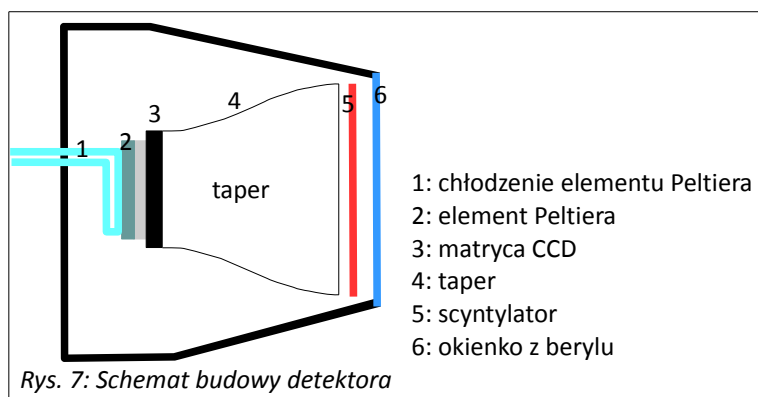


Rys. 6: [5]

3.2. Detektor

Do detekcji promieniowania wykorzystywany jest licznik scyntylacyjny w połączeniu z kamerą CCD.

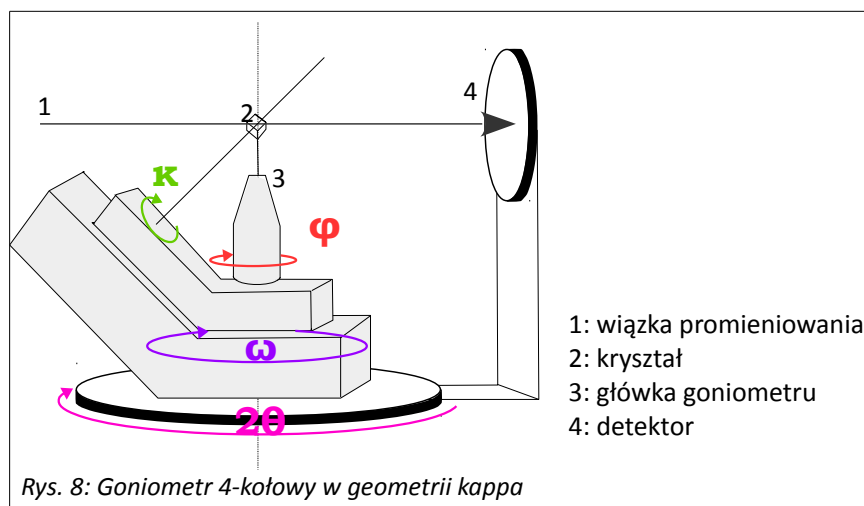
Scyntylator to substancja, która po zaabsorbowaniu kwantu promieniowania jonizującego emituje kwant promieniowania w zakresie widzialnym. Za scyntylatorem znajduje się taper, czyli element optyczny, przy pomocy którego przeskalowuje się obraz z detektora (którego średnica wynosi 135mm), na macierz kamery CCD, która ma mniejszy rozmiar. Aby ograniczyć szum matryca CCD schłodzona jest do -40°C za pomocą elementu Peltiera. Całość również umieszczona jest w próżni. Przed scyntylatorem znajduje się okienko z berylu, który jest dobrze przezroczysty dla promieniowania rentgenowskiego.



Rys. 7: Schemat budowy detektora

3.3. Goniometr

Kryształ białka montuje się na goniometrze. Goniometr zapewnia możliwość zebrania refleksów rozproszonych we wszystkich kierunkach, ponieważ łącznie z detektorem, ma możliwość obrotu wokół czterech osi, jest to tak zwany 4-kołowy goniometr w geometrii kappa (Rys. 8). Goniometr wyposażony jest w precyzyjne silniki krokowe, które pozwalają na precyzję ustawienia kryształu ok $10\mu\text{m}$.



Kryształ podczas pomiaru znajduje się w strumieniu gazowego azotu o temperaturze 100K.

4. Pomiar dyfrakcji kryształów białek

4.1. Przygotowanie kryształów do pomiaru

Wysoko energetyczne fotony promieniowania rentgenowskiego powodują degradację kryształów białka. Dyfraktometry „domowe” mają słabsze wiązki, więc pomiary trwają długo, kilka, kilkanaście minut na pojedyncze zdjęcie, na synchrotronach pomiary są krótsze, bo uzyskiwane wiązki mają większą intensywność, sumarycznie zniszczenia kryształu białka są podobne. Zwykle aby zebrać kompletne dane potrzebne było kilka kryształów białka. Do momentu gdy odkryto, że zeszkliwienie kryształu poprawia jego stabilność. Kryształ po wyłowieniu z kropli zamraża się w ciekłym azocie. Pomiary odbywają się w strudze gazowego azotu w temperaturze 100K.

Mrożenie ma również wady, ponieważ zwykle zwiększa mozaikowość kryształów, co wpływa na większe rozmycie plamek na obrazach dyfrakcyjnych.

4.2. Zebranie danych dyfrakcyjnych

Zebranie kompletnych danych dyfrakcyjnych może polegać na zebraniu refleksów z całej sfery, ale w większości przypadków jest to niepotrzebne. Ze względu na symetrię wewnętrzną kryształu obraz dyfrakcyjny będzie się powtarzał co 180, 90 lub 60°, zatem zbieranie danych z pełnej sfery jest niepotrzebny wysiłkiem.

Dla zaprojektowania rozsądnego eksperymentu dyfrakcyjnego konieczna jest znajomość symetrii wewnątrz kryształu. Eksperyment wstępny, polega na zebraniu kilka obrazów, najczęściej dwóch z kryształem obróconym o 90°, to pozwala wyznaczyć symetrię kryształu, wstępnie wyznaczyć grupę przestrzenną, oszacować rozmiary komórki elementarnej. Na podstawie tych informacji planuje się eksperyment w taki sposób aby zebrać „komplet” danych pomiarowych. Np dla kryształu o sześciokrotnej osi symetrii wystarczyłoby zebrać obrazy dyfrakcyjne dla sześćdziesięciostopniowego wycinka koła. Zwykle jednak zbiera się więcej ponieważ po pierwsze nie da się ustawić kryształu w wiązce tak by zacząć „od brzegu” komórki elementarnej, zatem musimy dodać „margines” lub „zapas” najlepiej +30° w obie strony, czyli już mamy 120°, po drugie dobrze jest zebrać pewien „nadmiar” danych bo to poprawia późniejszą statystykę wyników.

Zebranie dyfraktogramów jest ostatnim eksperymentalnym elementem rozwiązywania struktury białka, wszystkie kolejne etapy to obliczenia komputerowe.

5. Rozwiązanie struktury białka – część obliczeniowa

Analiza obrazów dyfrakcyjnych jest wieloetapowym procesem. Stworzonych zostało bardzo wiele programów i pakietów programów do analizy danych dyfrakcyjnych. Najczęściej używanymi są ccp4 i Phenix, są one dostępne za darmo!

W czasie ćwiczenia będą Państwo korzystać z pakietu ccp4.

5.1. Od obrazów dyfrakcyjnych – do czynników struktury (Data reduction)

5.1.1. Indeksowanie obrazów (Indexing)

Jeżeli pomiary były wykonane na wiązce synchrotronowej używa się programu *iMosflm* znajdującego się w pakiecie ccp4.

Jeżeli pomiary były wykonane na tzw. domowym dyfraktometrze dane analizuje się w oprogramowaniu dostarczonym przez producenta aparatury. Dane zostały zebrane przy pomocy dyfraktometru *SuperNova* firmy *Oxford Diffraction* i przeanalizowane przy pomocy oprogramowania *CrysAlisPro*.

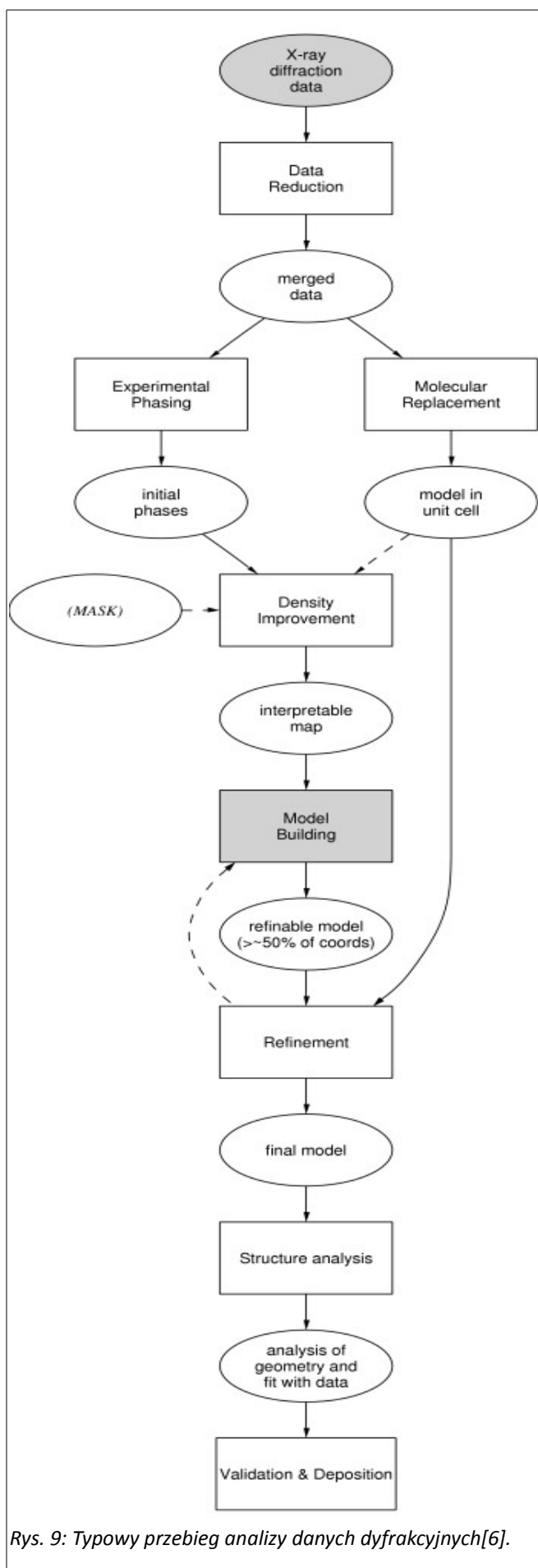
Indeksowanie obrazów dyfrakcyjnych to analiza, która ma na celu wykrycie wszystkich „mierzalnych” refleksów zarejestrowanych przez detektor, oraz na przypisaniu poszczególnym refleksom współczynników hkl.

Najpierw trzeba przeanalizować wszystkie obrazy dyfrakcyjne, określić na jakich obszarach obrazów znajdują się refleksy, a jakie obszary należy wykluczyć z analizy (cień beamstopera i blaszki na którym jest on umocowany), określić poziom szumów tła.

Przy pomocy odpowiednich programów analizuje się jasność poszczególnych pikseli i wyszukuje te, które są ciemniejsze niż tło. Uwzględnia się kształt plamki, odpowiedni stosunek sygnału do tła.

Efektom indeksowania jest jeden plik (o rozszerzeniu mtz), który zawiera tabelę wszystkich refleksów(hkl), ze wszystkich obrazów, każdemu przypisana jest intensywność piku wraz z odchyleniem standardowym (powtarzające się trójki hkl oznaczają, że dany refleks występuje na kilku kolejnych obrazach, każda z przypisanych intensywności obliczona na podstawie jednego obrazu)

Zbiór refleksów wykrytych na wszystkich obrazach, zapisanych w postaci zmierzonej intensywności(I) i szerokości



Rys. 9: Typowy przebieg analizy danych dyfrakcyjnych[6].

połówkowej sigl), wraz z przypisanymi im współczynnikami hkl zapisywany jest do pliku o rozszerzeniu mtz.

Plik zawiera

- pięć kolumn: h, k, l, l, sigl
- rozmiar komórki elementarnej
- możliwe grupy przestrzenne

Jeden wiersz w tej tabeli odpowiada jednej „plamce” na obrazie dyfrakcyjnym.

5.1.2. Skalowanie i scalanie (Scaling & Merging)

„Katalog refleksów” znajdujący się w omówionym powyżej pliku mtz trzeba przekształcić w dane pozwalające, na końcu, wykreślić mapę gęstości elektronowej jednostki asymetrycznej. Refleksy od tej samej rodziny płaszczyzn zarejestrowane na wielu obrazach muszą zostać „zebrane” w jedną intensywność. A następnie na podstawie scalonych intensywności należy obliczyć czynniki struktury.

Scalanie refleksów odbywa się przez całkowanie z wagami. Kryształ nie jest kulą, na poszczególnych obrazach piki mają różne intensywności wynikające także z tego, że wiązka przechodzi przez warstwy o różnej grubości. Wagi do refleksów z poszczególnych obrazów przypisywane są na podstawie analizy intensywności refleksów oraz poziomu tła na poszczególnych obrazach[6].

Podczas kolejnych etapów redukcji danych sprawdzana jest ich jakość, na podstawie spójności danych i porównania wyznaczonych natężeń refleksów I z ich odchyleniem standardowym obliczane są współczynniki korelacji oznaczane R:

$$R_{merge} = \frac{\sum_h \sum_L |I_{hl} - \langle I_h \rangle|}{\sum_h \sum_L \langle I_h \rangle} - \text{jest miarą rozbieżności między pomiarami, ma tendencję do rośnięcia wraz z liczbą}$$

zmierzonych refleksów

$$R_{meas} = \frac{\sum_h \left(\frac{n_h}{n_h - 1} \right) \sum_L |I_{hl} - \langle I_h \rangle|}{\sum_h \sum_L \langle I_h \rangle} - \text{ten czynnik nie zależy od liczby zmierzonych refleksów, jest poprawioną wersją}$$

R_{merge} [n_h jest liczbą zarejestrowanych refleksów h]

$$R_{p.i.m.} = \frac{\sum_h \left(\frac{1}{n_h - 1} \right) \sum_L |I_{hl} - \langle I_h \rangle|}{\sum_h \sum_L \langle I_h \rangle} - \text{jest miarą jakości danych po uśrednieniu refleksów.}$$

Te obliczenia wykonują Państwo wykorzystując trzy programy Pointless, Scala, cTruncate. Każdemu z tych programów, jako dane wejściowe trzeba podać plik mtz pochodzący z poprzedniego etapu obliczeń, plik ten zostanie zmodyfikowany i zapisany ponownie w formacie mtz.

5.1.2.1. Pointless

Pośród możliwych grup przestrzennych, wyliczonych na wcześniejszym etapie analizy, wyznacza najlepszą. Przygotowuje dane do kolejnych etapów skalowania i scalania.

5.1.2.2. Scala

łączy i skaluje w jedną intensywność piki od tej samej rodziny płaszczyzn znajdujące się na wielu obrazach.

Na tym etapie analizy danych wydzielamy 5% zmierzonych refleksów i traktujemy je od tej pory jako kontrolny zestaw danych. Zbiór refleksów zostanie losowo podzielony na 20 podzbiorów, numerowanych od 0 do 19, w taki sposób, żeby rozkład liczby refleksów względem rozdzielczości był podobny we wszystkich podzbiórach. W pliku mtz pojawi się dodatkowa kolumna opisana FreeR.

Dalsza analiza danych będzie niezależnie przeprowadzana na tych 5% i na pozostałych 95% (te 5% refleksów jest traconych z analizy badanego białka), ale zyskuje się kontrolę nad jakością kolejnych etapów analizy. Wyniki uzyskiwane dla tych 5% i reszty powinny być podobne, np wartości współczynników R.

5.1.2.3. *cTruncate*

Na podstawie intensywności oblicza amplitudy czynników strukturalnych.

Dla idealnych danych eksperymentalnych amplituda jest pierwiastkiem kwadratowym intensywności, ale dla danych obarczonych błędem eksperymentalnym lepszym oszacowaniem czynników strukturalnych, zwłaszcza dla refleksów o niewielkich intensywnościach, jest średnia rozkładu prawdopodobieństwa.

Program *cTruncate* wykorzystuje taką procedurę do policzenia czynników struktury, dodatkowo, na podstawie analizy statystycznej zbioru intensywności wykrywa twinning.

Do wcześniejszego pliku dopisuje kolumny zawierające amplitudy i ich dyspersję (F , σF)

Tak przygotowany plik mtz zawiera wszystkie dane możliwe do uzyskania z obrazów dyfrakcyjnych: czynniki struktury wraz z błędem.

Cynniki struktury reprezentują falę, która uległa rozproszeniu na badanym obiekcie, co oznacza, że tak samo jak fala muszą być opisane przez amplitudę i fazę.

5.2. Rozwiązanie problemu faz metodą podstawienia molekularnego

5.2.1. Tworzenie modelu (*Model Building*)

Pakiet CCP4 zawiera rozmaite narzędzia, które pozwalają edytować sekwencję, modyfikować strukturę 3D, pozwalając stworzyć model z rozmaitych „fragmentów” znalezionych w różnych bazach danych. (np Chainsaw, Modeller)

Jako model wykorzystamy strukturę z bazy PDB.

5.2.2. Przygotowanie danych do podstawienia molekularnego

5.2.2.1. *Cell Content Analysis (Matthews coefficient)*

Na początku musimy ustalić ile cząsteczek białka znajduje się w jednostce asymetrycznej.

Aby wyznaczyć liczbę cząsteczek białka w jednostce asymetrycznej obliczamy tzw. współczynnik Matthews. Określa on średnie odległości między atomami w kryształach białka [$\text{\AA}^3/\text{atom}$]. Dla białek, średnia wartość tego współczynnika wynosi $2,35 \text{\AA}^3/\text{atom}$. Współczynnik Matthews jest obliczany na podstawie rozmiarów komórki elementarnej i masy cząsteczkowej badanego białka.

Jako dane wejściowe musimy podać rozmiary komórki elementarnej i grupę przestrzenną (te informacje zawarte są w pliku mtz) oraz parametr, który pozwoli określić jaką część objętości komórki elementarnej zajmuje białko (np masę, sekwencję, liczbę aminokwasów).

5.2.2.2. *Analyse data for MR*

Ten program oblicza funkcję Pattersona i wyznacza jej maksima i wyznacza średnią wartość współczynnika B na podstawie wykresu Wilsona.

Obrazem funkcji Pattersona jest mapa Pattersona, jest ona podobna do mapy gęstości elektronowej, ale jej maksima nie znajdują w tych położeniach co atomy, lecz w położeniach odpowiadających względnym odległościom między atomami. Jeśli w komórce znajduje się atom w pozycji x_1 i drugi atom w pozycji x_2 to na mapie Pattersona będą znajdowały się dwa pików w pozycji x_1-x_2 i w pozycji x_2-x_1 . Wysokość pików na mapie Pattersona jest proporcjonalna do iloczynu wysokości pików na mapie elektronowej. Jeżeli w komórce elementarnej znajduje się N atomów, to powinno się w niej znajdować N pików gęstości elektronowej. Na mapie Pattersona każdy z tych N atomów jest połączony wektorem z każdym z N atomów, co daje N^2 wektorów. Przy czym N spośród tych wektorów będzie łączyło poszczególne atomy z nimi samymi, co wygeneruje duży pik w

pozycji 0. Pozostałe N2-N pików będzie rozłożonych wokół niego.

Mapa Pattersona jest w ogólności bardzo trudna w interpretacji, ale na potrzeby podstawiania molekularnego potrzebnych jest z niej tylko kilka informacji: liczba i wzajemne położenie cząsteczek w komórce elementarnej, oraz wykrycie ewentualnych pseudotranslacji.

Pierwszy pik na mapie Pattersona jest znormalizowany do 100, jeżeli wartości żadnego z pozostałych pików nie przekracza 20% wysokości pierwszego pików to oznacza, że w komórce elementarnej znajduje się tylko jedna cząsteczka badanego białka.

Współczynnik B jest miarą efektywnej średnicy gęstości elektronowej danego atomu. Wpływają na nie przede wszystkim termiczne drgania atomu.

Wielkość ta bywa użyteczna podczas podstawiania cząsteczki modelu i późniejszego udokładniania struktury (można wprowadzić dodatkowe więzy).

5.2.3. Podstawienie molekularne

W pakiecie ccp4 jest kilka programów, przy pomocy, których można dokonać podstawienia molekularnego: Phaser, MolRep, MrDUMP, Balbes, Amore. Skorzystamy z programu MolRep:

Operacje, wykonywane w programie MolRep:

- obrót i przesunięcie cząsteczki modelowej do pozycji, w której znajduje się badana cząsteczka w komórce elementarnej – pozycjonowanie.

- przypisanie faz z modelu do odpowiednich czynników struktury w naszych danych eksperymentalnych – to są początkowe fazy do udokładniania(refinement) struktury.

Jako dane wejściowe należy podać plik mtz po skalowaniu u scaleniu oraz plik pdb z cząsteczką modelową.

Jako dane wyjściowe program wygeneruje nowy plik pdb z przeorientowaną cząsteczką modelową.

5.3. Udokładnianie struktury

One should bear in mind that a macromolecular refinement against high resolution data is never finished, only abandoned. George Sheldrick (2008), Acta Crystallogr. D64112–122.

5.3.1. Refmac 5

Do udokładniania struktury użyjemy programu Refmac5.

Program wykorzystując metodę maksimum likelihood „wpasuje” cząsteczkę modelu w gęstość elektronową, a następnie policzy mapę gęstości elektronowej(F) i różnicową mapę gęstości elektronowej(ΔF), mapy te zdefiniowane są w następujący sposób:

$$F = 2 \cdot F_{obs} - F_{calc}$$

$$\Delta F = F_{obs} - F_{calc}$$

Algorytmy udokładniania zaimplementowane w programie Refmac5 są niewystarczające, wciąż jeszcze ostateczną instancją jest człowiek. Poprawiamy dopasowanie wykonane w Refmac'u oglądając uzyskaną strukturę i porównując z nią mapę gęstości elektronowej. Dlatego właśnie w Refmac'u liczona jest mapa różnicowa. Pozwala ona łatwo znaleźć obszary gdzie model nie pasuje do gęstości elektronowej, gdzie albo czegoś „brakuje” czyli jest gęstość elektronowa wewnątrz której nie ma struktury, lub czegoś jest „za dużo” czyli fragment struktury nie przechodzi przez obserwowaną gęstość elektronową.

Jako dane wejściowe należy podać plik mtz zawierający czynniki struktury, oraz plik pdb z wyzycjonowaną cząsteczką modelu.

Efektom obliczeń programu będzie nowy plik pdb ze współrzędnymi wstępnie udokładnionej cząsteczki, a do pliku mtz zostaną dopisane cztery kolumny pozwalające wykreślić mapę gęstości elektronowej (FWT – czynniki struktury i PHWT – fazy) oraz mapę różnicową (DELFWT i PHDELWT)

Refmac oblicza też szereg parametrów pozwalających ocenić jak dobre jest dopasowanie modelu do danych dyfrakcyjnych

5.3.2. Coot

Jest to program wykorzystywany do ręcznego udokładniania struktur. Przy pomocy *Coot'a* wykreśla się dwie mapy, mapę gęstości elektronowej i mapę różnicową. Na mapie różnicowej znajdują się obszary o ładunku dodatnim (zielone) gdzie model powinien mieć więcej elektronów i ujemnym (czerwone) gdzie model powinien mieć mniej elektronów.

Coot pozwala na ręczną edycję cząsteczki modelu. Posiada wiele narzędzi, które pozwalają modyfikować położenie aminokwasów i kształt łańcucha polipeptydowego.

Procedura udokładniania jest rekurencyjna. Po dokonaniu serii poprawek, należy uzyskaną strukturę poddać ponownej analizie w *refmac'u*, który policzy nowe mapy różnicowe gęstości elektronowej, niejako „akceptując” dokonane zmiany i „pokazując” obszary, które jeszcze wymagają udokładnienia. Postępy udokładniania można dodatkowo kontrolować przy pomocy programu *sfcheck*, o którym więcej, poniżej.

Ostatnim elementem udokładniania struktury jest dodanie cząsteczek wody. Wykonujemy tę operację w *Coot'ie* lub *Refmac'u*

5.4. Kontrola jakości dopasowania (Structure Validation)

Po podjęciu decyzji o zaprzestaniu dalszego udokładniania jakość uzyskanej struktury można ocenić przy pomocy programu *sfcheck*. Program ten przeanalizuje uzyskaną strukturę pod kątem jest zgodności z mapą gęstości elektronowej, licząc współczynniki R. Sprawdzi również czy długości i kąty wiązań zawierają się w przedziałach typowych dla znanych struktury białek, jednym z uzyskanych wykresów będzie mapa Ramachandrana pokazująca, które aminokwasy mają konformacje w niedozwolonych obszarach kątów ψ i θ .

5.5. Prezentacja wyników

Rozwiązane struktury zwykle deponuje się w PDB.

Natomiast w publikacjach omawiających rozwiązane struktury zwykle przedstawia się table podobne do przedstawionej poniżej:

Data collection	
Wavelength (Å)	0.975
Resolution range (Å)	28.6 - 1.9
Space group	P 61 2 2
Unit cell	120.8 120.8 239.1 90 90 120
Total reflections	704235
Unique reflections	80026
Multiplicity	8.8
Completeness (%)	99.4
Rmerge	0.25
Rmerge in top intensity bin	0.019
I/sigma(I)	21.6
Wilson B-factor	26.14
Refinement	
Rwork	0.163
Rfree	0.188
Number of atoms	6187
Protein residues	718
Water molecules	623
RMS (bonds)	0.008
RMS (angles)	1.10
Ramachandran favored (%)	96
Ramachandran outliers (%)	0.69
Average B-factor	31.10

6. Bibliografia

- [1] Bernhard Rupp, Crystallography 101, <http://www.ruppweb.org/Xray/101index.html>
- [2] University of Cambridge, Cambridge Institute for Medical Research (CIMR), Protein Crystallography Course, <http://www-structmed.cimr.cam.ac.uk/course.html>
- [3] P. Evans, A McCoy, An introduction to molecular replacement,
- [4] Yuan Ming Huang, Solid State Physics: Miller indices, <http://www.lcst-cn.org/Solid%20State%20Physics/Ch16.html>
- [4] Stephen A. Nelson, X-Ray Crystallography, <http://www.tulane.edu/~sanelson/eens211/x-ray.htm>
- [5] P. Evans, Scalling & Assessment of data quality,
- [6] CCP4 Software for Macromolecular X-Ray Crystallography <http://www.ccp4.ac.uk/>, <http://ccp4wiki.org/>

7. Literatura uzupełniająca

Bernhard Rupp *Crystallography 101* <http://www.ruppweb.org/Xray/101index.html>

Protein Crystallography Course (University of Cambridge, Cambridge Institute for Medical Research (CIMR)) <http://www-structmed.cimr.cam.ac.uk/course.html>

CCP4 Software for Macromolecular X-Ray Crystallography <http://www.ccp4.ac.uk/>, <http://ccp4wiki.org/>

Basic Maths for Protein Crystallographers: <http://www.ccp4.ac.uk/dist/html/pxmaths/index.html>

Zbigniew Dauter: Data collection strategies, *Acta Cryst. D* (1999) 55,1703-1717

Philip Evans: Scalling & Assessment of data quality, *Acta Cryst. D* (2006). 62,72–82

Philip Evans, Airlie McCoy *An introduction to Molecular Replacement* *Acta Cryst. D*(2008) 64, 1–10

Garib N. Murshudov et. al. *REFMAC5 for the refinement of macromolecular crystal structures* *Acta Cryst. D*(2011). 67, 355–367

8. Zagadnienia do kolokwium wstępnego

- Oddziaływanie światła z materią – dyfrakcja i interferencja
- Budowa kryształów, cechy specyficzne kryształów białek.
- Symetrie kryształów, grupy przestrzenne, cechy specyficzne kryształów białek.
- Stabilność kryształów białkowych w wiązce rentgenowskiej.
- Co widać na obrazach dyfrakcyjnych?
- Indeksowanie obrazów dyfrakcyjnych.
- Problem faz i metody rozwiązania problemu faz.
- Metoda podstawienia molekularnego
- Walidacja struktury.

9. Wykonanie ćwiczenia

Celem ćwiczenia jest zebranie danych dyfrakcyjnych i rozwiązanie struktury ludzkiej albuminy krwi (human serum albumin, HSA).

Kryształy HSA zostaną dla Państwa przygotowane wcześniej.

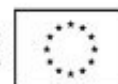
Na podstawie analizy wizualnej kropli na płytce wytypujemy kryształy do badań dyfrakcyjnych. Następnie wyłowimy je przy pomocy specjalnych pętelek i zamrozimy w ciekłym azocie. Tak przygotowane kryształy zamontujemy w dyfraktometrze i zbierzemy dane dyfrakcyjne. Analizując uzyskane dane przy pomocy pakietu CCP4 (<http://www.ccp4.ac.uk/>) rozwiążemy strukturę HSA metodą podstawienia molekularnego.



KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI



UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY



Projekt Fizyka wobec wyzwań XXI wieku współfinansowany ze środków Unii Europejskiej w ramach Europejskiego Funduszu Społecznego