

MARIUSZ JASKÓLSKI*

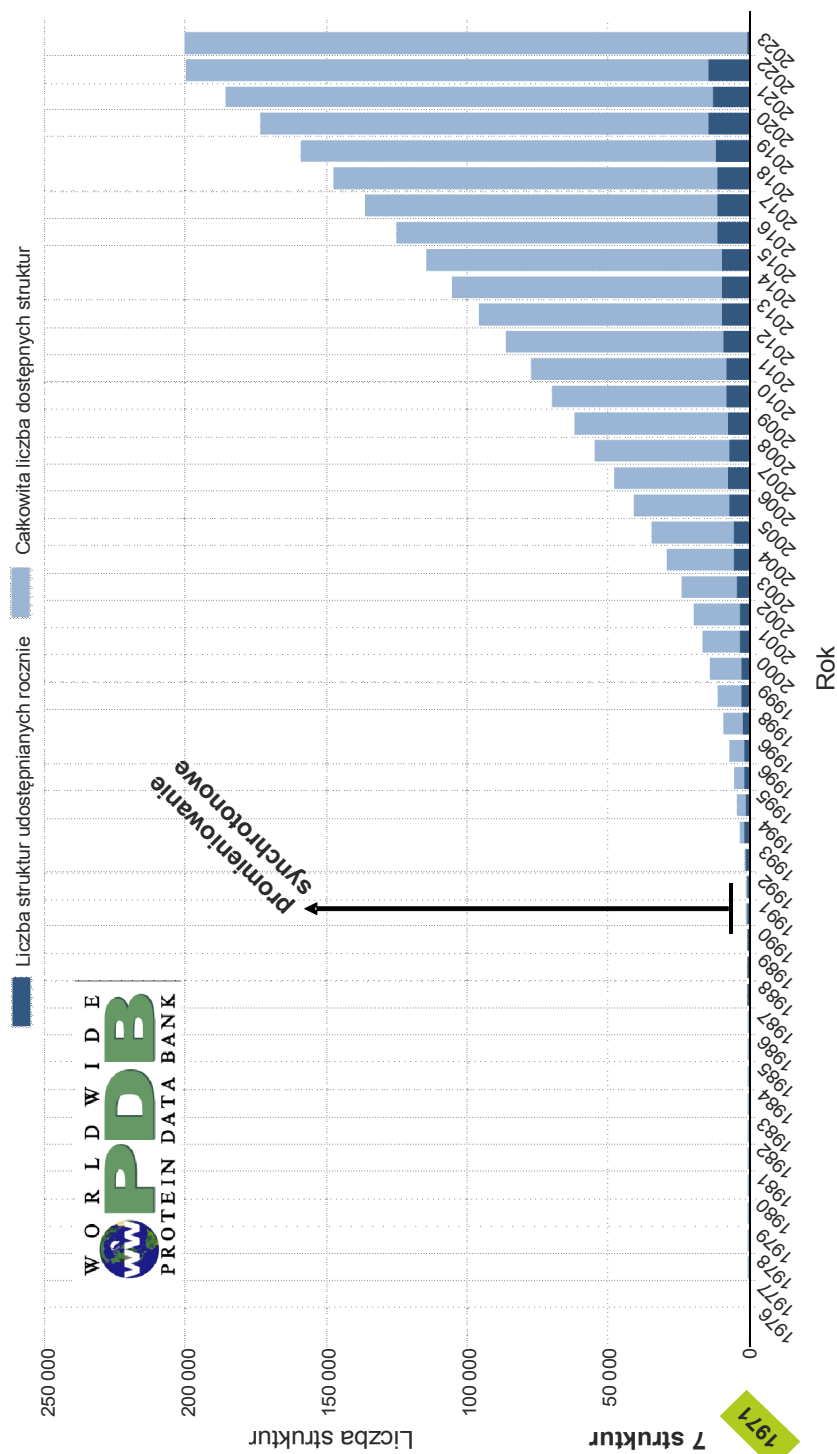
Biologia strukturalna: historia, teraźniejszość i perspektywy

Wstęp

Biologia strukturalna bada strukturę przestrzenną makromolekuł biologicznych, głównie białek i kwasów nukleinowych, w celu zrozumienia, a być może i regulacji, funkcjonowania molekuł życia. Ze względu na ogromne rozmiary i giętkość tych molekuł-olbrzymów, ich struktury 3D jeszcze do niedawna nie można było przewidzieć *ab initio*. Problem „ścieżki zwijania” białek pozostanie jeszcze długo nierozwiązany, chyba że pomogą go złamać komputery kwantowe.

Swoje początki w połowie XX. wieku wiąże biologia strukturalna z krystalografią rentgenowską, z którą przez wiele dekad była utożsamiana. Metoda krystalograficzna polega na badaniu struktury otrzymanego w postaci kryształu materiału biologicznego za pomocą dyfrakcji promieni X. W latach 80. ubiegłego wieku pewne wsparcie eksperymentalne przyszło ze strony jądrowego rezonansu magnetycznego (NMR) umożliwiającego badania struktury białek w roztworze, ale prawdziwy przełom nastąpił dopiero w ostatniej dekadzie wraz z rozwojem wysokorozdzielczej kriomikroskopii elektronowej (cryo-EM), pozwalającej na bezpośrednie obrazowanie makromolekuł z rozdzielczością niemal atomową. Protein Data Bank (PDB), który od 1971 roku gromadzi wyznaczone doświadczalnie struktury makromolekuł, początkowo odnotowywał kilka, najwyżej kilkadziesiąt nowych struktur rocznie (ryc. 1). W połowie lat 90. nastąpiła prawdziwa „eksplozja” zasobów PDB, związana z trzema czynnikami: (i) rozwojem biotechnologii, (ii) rozwojem technologii bardzo wydajnych komputerów osobistych oraz, przede wszystkim, (iii) z powszechnym zastosowaniem potężnych synchrotronowych źródeł promieniowania X. Postęp techniczny poszedł jednak dalej i dziś dysponujemy również laserami rentgenowskimi na swobodnych elektronach (XFEL), których promieniowanie ma wprost niewyobrażalną jasność. Umożliwia to skrócenie czasu naświetlania do kilkunastu femtosekund oraz badanie kryształów o wymiarach ~100 nanometrów, a nawet pojedynczych makromolekuł. Wszystko to sprawia, że w PDB jest zdeponowanych obecnie ponad 210 tys. wyznaczonych doświadczalnie struktur makromolekuł, a roczny przyrost sięga 15 tys.

* Prof. dr hab. Mariusz Jaskólski (mariuszj@amu.edu.pl), członek rzeczywisty PAN, Zakład Krystalografii, Wydział Chemii, Uniwersytet im. Adama Mickiewicza w Poznaniu oraz Instytut Chemii Bioorganicznej PAN w Poznaniu; ORCID: 0000-0003-1587-6489



Ryc. 1. Rozwój Banku Struktur Białkowych (Protein Data Bank, PDB) (źródło: wwPDB)

Wiele z tych struktur, jak np. białka wirusa HIV czy SARS-CoV-2, ma ogromne znaczenie medyczne, gdyż na ich podstawie skutecznie projektuje się leki nowej generacji.

Obok prawdziwej rewolucji na polu doświadczalnym, w ostatnim czasie nastąpił też przełom w metodach przewidywania trójwymiarowej struktury białek (3D) na podstawie ich sekwencji, tj. informacji jednowymiarowej (1D). Stało się to możliwe dzięki algorytmom głębokiego uczenia maszynowego, które wytrenowano pod kątem rozpoznawania w sekwencjach białek wzorców strukturalnych. Nie jest to co prawda jeszcze rozwiązanie problemu ścieżki zwijania, a dla zupełnie nowych sekwencji wynik nie jest w pełni wiarygodny; niemniej dla łatwiejszych problemów sztuczna inteligencja (AI) potrafi znaleźć poprawne rozwiązanie błyskawicznie. Dostępny powszechnie serwer AlphaFold zaproponował już strukturę dla setek milionów białek. Inny serwer obiecuje niebawem miliard struktur, choć wydaje się, że taka numerologiczna sztuka dla sztuki traci już sens.

Wymiary w świecie atomów, molekuł i makromolekuł

W świecie atomów i molekuł obowiązuje specjalna miara długości: 1 angstrom ($\text{\AA}$), czyli stumilionowa część centymetra, $1\text{\AA} = 10^{-8}\text{ cm}$. Choć angstrom nie należy do systemu SI, jest przez ten system akceptowany, gdyż znakomicie upraszcza poruszanie się w świecie atomów i molekuł. Odległości między atomami związanymi w molekułach, a więc długości wiązań chemicznych, są bowiem rzędu 1–2 $\text{\AA}$. Na przykład długość wiązania peptydowego OC–N, łączącego reszty aminokwasowe w łańcuchy białkowe, wynosi 1,33 $\text{\AA}$.

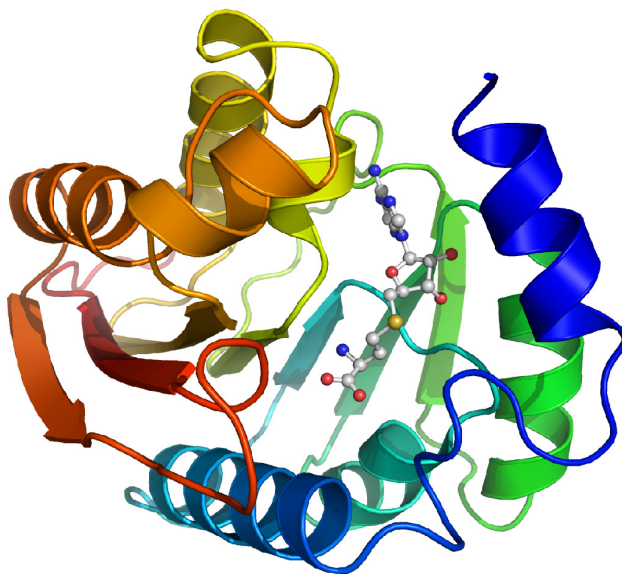
To, że wiązania chemiczne są tak krótkie, nie oznacza jednak, że molekuły życia, słusznie nazywane makromolekułami lub molekułami-olbrzymami, są „mikroskopijne”. Tak nie jest, gdyż makromolekuły biologiczne, głównie białka i kwasy nukleinowe, zbudowane są z wielu tysięcy, czasem nawet wielu milionów, atomów i mogą osiągać wymiary liniowe (długość) dostępne naszym zmysłom. Na przykład liniowy łańcuch białka złożonego z połączonych wiązaniami peptydowymi 2000 reszt aminokwasowych miałby po rozciągnięciu długość 0,7 mikrona – do zmierzenia w warsztacie mechanicznym. Jeszcze dłuższe są zbudowane z nukleotydów łańcuchy kwasów nukleinowych. Genom ludzki zawiera ponad 3,1 mld par zasad. Odległość par zasad w kanonicznej strukturze B-DNA wynosi 3,4 $\text{\AA}$. Tak więc długość podwójnej helisy DNA kodującej nasz genom to ponad 1 m. Ponieważ w każdej komórce somatycznej mamy dwa garnitury chromosomów, sumaryczna długość podwójnej helisy w komórce to ponad 2 m, a wszystkich nici DNA ponad 4 m. I to wszystko w maleńkim, kilkumikronowym jądrze komórkowym!

Nie pomijając kwasów nukleinowych, kodujących przecież informację genetyczną, w naszej opowieści skupimy się na białkach, gdyż w dzisiejszym świecie to one właśnie odpowiadają za większość procesów chemicznych ukrytych pod terminem „życie”.

Przede wszystkim to enzymy białkowe są katalizatorami większości reakcji biochemicznych; białka są materiałem strukturalnym komórek i tkanek, transportują inne biomolekuły, przenoszą sygnały, zarządzają oddychaniem, energią oraz ruchem; białka wreszcie kontrolują samą ekspresję informacji genetycznej. Jeden wszakże fundamentalny proces pozostał w gestii kwasów nukleinowych: odbywająca się w rybosomach, pod dyktando kodu genetycznego, biosynteza białek. Najkrócej rzecz ujmując: za biosyntezę białek odpowiadają kwasy nukleinowe, a za biosyntezę kwasów nukleinowych – białka.

Krystalografia białek – rys historyczny

Podstawowe motywy struktury białek, czyli tzw. struktury drugorzędowe, a więc uwiłta przestrzennie helisę α i złożone z rozciągniętych łańcuchów białkowych arkusze β , zaproponował w 1951 roku najgenialniejszy chemik wszechczasów, Linus Pauling (ryc. 2).



Ryc. 2. Przykład łańcucha białkowego zwinętego do poprawnej struktury 3D. Przedstawiony schematycznie w postaci wstęgi łańcuch białkowy w rzeczywistości składa się z połączonych liniowo „cegiełek” chemicznych, zwanych aminokwasami. Sekwencja aminokwasów (w tym przypadku 216) stanowi strukturę pierwszorzędową białka. Lokalna konformacja, widoczna w postaci helisy α (linia śrubowa) i rozciągniętych łańcuchów β – to struktura drugorzędowa. Struktura trzeciorzędowa to ostateczny zwój białkowy. Niektóre białka, które złożone są z kilku łańcuchów zwanych podjednostkami (tu nie występują), mają jeszcze strukturę czwartorzędową, określającą wzajemną aranżację podjednostek. W pokazanym na ilustracji białku związany jest małocząsteczkowy ligand, przedstawiony jako model kulkowy (atomowy). Na takim dopasowaniu ligandu do białka polega racjonalne projektowanie leków

Natomiast pierwsze eksperymentalne struktury białek ustalili pod koniec lat 50., wykorzystując dyfrakcję promieniowania rentgenowskiego na kryształach białek, krystalografowie: mioglobiny – John Kendrew i hemoglobiny – prawdziwy pionier w tej dziedzinie – Max Perutz. Tak narodziła się biologia strukturalna. Na jej narodziny złożyło się kilka wcześniejszych, czasem pozornie skromnych, lecz w istocie przełomowych, odkryć. Było wśród nich odnotowane w 1840 roku przez Friedricha Ludwiga Hünefelda pojawienie się kryształów hemoglobiny (nazwanych przez niego *Blutkristalle*) w kropli krwi dżdżownicy [1], a więc pierwsza naukowa obserwacja krystalizacji białka.

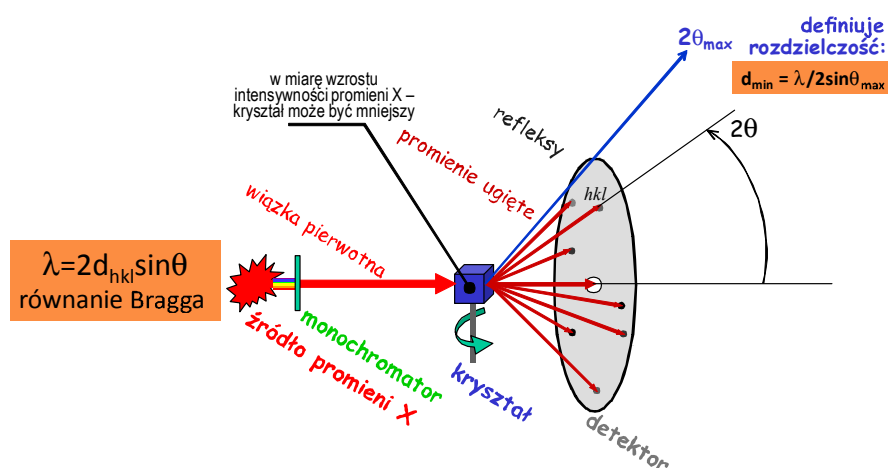
Kolejnym krokiem, który umożliwił narodziny biologii strukturalnej, było odkrycie promieni X, którego dokonał w 1895 roku Wilhelm Conrad Röntgen, pierwszy noblista z fizyki. Przełomowego odkrycia, ukazującego dyfrakcję promieni X na kryształach, dokonali w 1912 roku Walter Friedrich i Paul Knipping, wykonując doświadczenie według wskazówek Maxa Lauego [2]. Jako fizyk-teoretyk Max von Laue nie wykorzystał potencjału tego odkrycia na gruncie chemii. Zrobili to rok później w duecie syn i ojciec Braggowie w Anglii, rozwiązując na podstawie dyfrakcji rentgenowskiej struktury prostych kryształów, takich jak halit (NaCl) czy diament (kryształ węgla). Dzięki wizjonerom, takim jak Desmond Bernal, Dorothy Hodgkin i Max Perutz, już na początku lat 30. XX wieku pojawił się pomysł, że tą samą metodą będzie można ustalić trójwymiarowe struktury atomowe kryształów białek. Zrealizowali to po blisko trzydziestu latach John Kendrew i Max Perutz. Większość wspomnianych tu badaczy i odkryć uhonorowały Nagrody Nobla [3].

Przez kolejne trzy dekady (1960–1980) biologię strukturalną utożsamiano w zasadzie z krystalografią białek. Ustalenie struktury przestrzennej hemoglobiny zajęło Perutzowi 22 lata. W roku 1971, gdy znano zaledwie siedem struktur białek, utworzono Protein Data Bank (PDB). Dziś PDB zawiera ponad 210 tys. ustalonych eksperymentalnie struktur makromolekuł (ryc. 1).

Krystalografia białek w pigułce

Typowy eksperyment dyfrakcyjny ma dziś ten sam schemat, co w czasach Lauego i obejmuje trzy podstawowe elementy: źródło promieniowania rentgenowskiego, kryształ oraz detektor promieni ugiętych na kryształach, zwanych refleksami (ryc. 3). Zwykle (choć będą wyjątki) stosujemy promieniowanie monochromatyczne o jednej długości fali λ , gdyż pozwala to łatwo interpretować wzór Bragga $\lambda = 2d_{hkl} \times \sin\theta$; wystarczy zmierzyć kąt odbicia θ , by ustalić, od jakiej rodziny płaszczyzn hkl w sieci kryształu nastąpiło odbicie, oraz jaki jest odstęp międzypłaszczyznowy d_{hkl} (w angstromach!) w tej rodzinie. Na marginesie zauważymy, że przekształcony wzór Bragga dla maksymalnego kąta ugięcia θ_{max} pozwala wyliczyć minimalny odstęp międzypłaszczyznowy $d_{min} = \lambda/2\sin\theta_{max}$, czyli rozdzielczość, z jaką wyznaczamy strukturę kryształu. Rozdzielczość,

tak jak w mikroskopii, odpowiada na pytanie, jaka jest najmniejsza odległość dwóch punktów, które uda się nam rozdzielić w uzyskanym obrazie.



Ryc. 3. Rentgenowski eksperyment dyfrakcyjny. Niebieski promień (refleks) ma największy kąt ugięcia (θ_{max}) i on definiuje rozdzielczość d_{min} tego eksperymentu. W przypadku promieniowania monochromatycznego cały obraz dyfrakcyjny składamy z wielu zdjęć wykonanych np. przy różnych orientacjach kryształu podczas jego obrotu lub dla wielu przypadkowo zorientowanych kryształów. W metodzie polichromatycznej (bez monochromatora) wystarczy jedno zdjęcie, bowiem różne długości fali λ pozwolą odtworzyć cały obraz dyfrakcyjny przy jednej orientacji kryształu. Przez ok. pierwszych 100 lat promieniowanie X do badań strukturalnych pochodziło z lamp rentgenowskich. Dziś generowane jest w potężnych akceleratorach zwanych synchrotronami (patrz ryc. 5)

Dyfrakcja monochromatyczna ma jednak tę wadę, że na pojedynczym „zdjęciu” (tj. przy jednej ekspozycji) ukazuje tylko mały fragment obrazu dyfrakcyjnego; aby uzyskać pełen obraz, należy „obfotografować” kryształ z wielu stron, najlepiej podczas obrotu. Można też, zamiast wykonywać kolejne zdjęcia obracającego się kryształu, zarejestrować pojedyncze obrazy dla wielu przypadkowo zorientowanych kryształków (np. wstrzykiwanych do wiązki promieniowania X szeregowo), a następnie złożyć je w pełen obraz dyfrakcyjny. Można też użyć jednego kryształu i jednej ekspozycji, ale w promieniowaniu polichromatycznym (w szerokim zakresie λ), co również da cały obraz dyfrakcyjny (ale tym razem złożony z superpozycji obrazów uzyskanych przy różnych długościach fali).

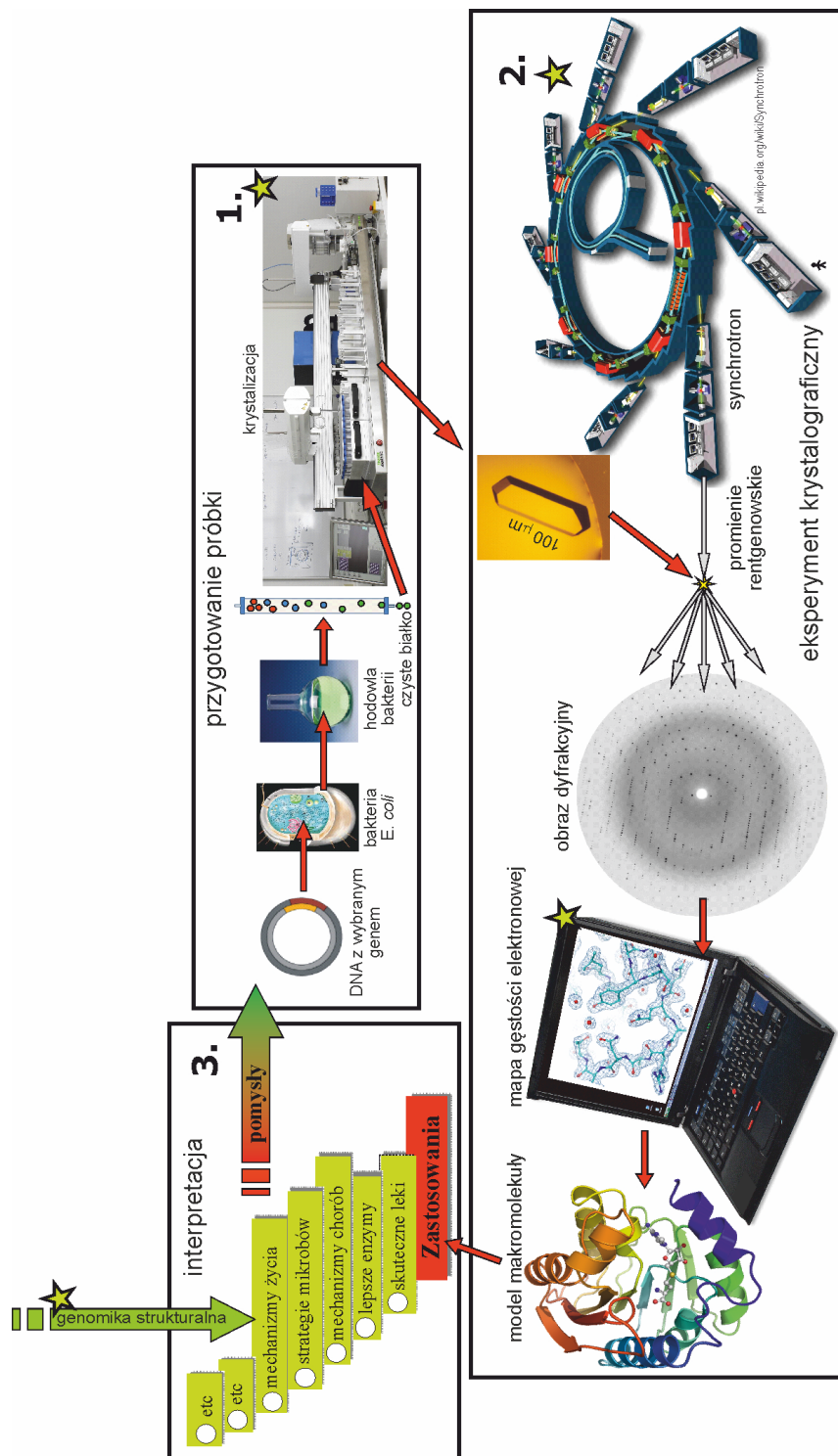
Jednak samo zarejestrowanie pozycji (kątów θ) refleksów (których mogą być setki tysięcy lub miliony) nie wystarczy do rozwiązania struktury kryształu. Znacznie ważniejsze są intensywności, a w zasadzie związane z czynnikami struktury F_{hkl} amplitudy poszczególnych refleksów. Wiąże się to z faktem, że obraz dyfrakcji promieni X, które rozpraszane są na chmurach elektronowych atomów, jest transformatą Fouriera rozkła-

du gęstości elektronowej $\rho(xyz)$ w kryształach. Na mocy teorii Fouriera prawdziwa jest też zależność odwrotna: rozkład gęstości elektronowej jest odwrotną transformatą funkcji $F_{hkl} = |F_{hkl}| \times \exp(i\varphi_{hkl})$. (Ten nieco skomplikowany wzór oznacza falę o amplitudzie $|F_{hkl}|$ i fazie φ_{hkl}). Z pomiarów bardzo łatwo wyznaczymy amplitudy $|F_{hkl}|$, lecz nie fazy φ_{hkl} refleksów. Jest to podstawowa trudność matematyczna wyznaczania struktury kryształów, zwana „problemem fazowym”.

Krystalografia białek dziś

Cykl badawczy w krystalografii białek rozpoczyna się na długo przed otrzymaniem kryształu (ryc. 4). Białka wybranego do badań nie izoluje się obecnie z materiału naturalnego. Byłoby to zbyt żmudne i czasochłonne, gdyż wymagałoby często „przerobienia” dziesiątków kilogramów jednorodnego materiału wyjściowego, a w przypadku białek ludzkich byłoby etycznie problematyczne. Zwykle prościej jest ustalić sekwencję DNA kodującego interesujące nas białko i otrzymać taki „gen” syntetycznie, a następnie wkłonić (tj. „wkleić”) go do odpowiedniego plazmidu, tj. kolistego DNA, który jest elementem genetycznym bakterii. (Na marginesie zauważymy, że takie postępowanie stało się możliwe dzięki spektakularnemu sukcesowi genomiki, która ustaliła sekwencje genomów tysięcy organizmów, z genomem ludzkim włącznie.) Po transformacji takim plazmidem, komórki bakteryjne (zwykle *Escherichia coli*) produkują w sposób kontrolowany pożądane białko (zwane rekombinowanym) w dużej ilości. Oczyszczanie białek rekombinowanych odbywa się metodą chromatografii i jest kapitalnie uproszczone przez fakt, że na etapie klonowania można dołączyć do sekwencji białka znacznik chemiczny umożliwiający wiązanie („powinowactwo”) na odpowiedniej kolumnie chromatograficznej. Na przykład wyposażenie białka w tzw. metkę histydynową 6xHis umożliwia wiązanie na kolumnie z jonami niklu.

Większość procesów, o których tu mowa, udało się częściowo lub w pełni zautomatyzować i zrobotyzować, tj. przystosować do trybu wysokoprzepustowego (*high throughput*) (ryc. 4). Wiele czynności związanych z ekspresją i oczyszczaniem białek przebiega automatycznie; przede wszystkim jednak kolejne etapy: krystalizacja i pomiary dyfrakcyjne odbywają się niemal wyłącznie pod kontrolą robotów. Obliczenia komputerowe z natury rzeczy są automatyczne. Jeśli idzie o krystalizację, to w niektórych przypadkach udaje się wyeliminować nawet roboty. Na przykład zauważono, że podczas wysokiej nadekspresji białek rekombinowanych w komórkach bakteryjnych, nadprodukowane białko ma tendencję do krystalizowania już w komórce bakteryjnej. Ponieważ do zaawansowanych pomiarów dyfrakcyjnych (np. z wykorzystaniem promieniowania lasera XFEL, *vide infra*) wystarczą supermałe, nanometrowe kryształy – taka krystalizacja *in cellulo* (w komórce) spełnia swoje zadanie w 100%: białko otrzymuje się od razu w postaci czystej i krystalicznej [4].





Ryc. 4. Proces badawczy biokrystalografii rozpoczyna się zwykle od (1) przygotowania na drodze inżynierii genetycznej bardzo czystego białka, które następnie poddaje się procesowi krystalizacji (obecnie z zastosowaniem robotów). Właściwe badania krystalograficzne (2) polegają na zarejestrowaniu dyfrakcji na kryształach przy użyciu superintensywnego promieniowania rentgenowskiego generowanego w cyklotronie zwanym synchrotronem. Obraz dyfrakcyjny po skomplikowanych obliczeniach zostaje przekształcony w mapę gęstości elektronowej kryształu, która ujawnia obraz atomowy badanej molekuly. Najciekawszy etap badań polega na interpretacji uzyskanych wyników (3) i wiąże się nie tylko z odkrywaniem tajemnic molekularnych życia, ale i z zastosowaniami praktycznymi, takimi jak projektowanie leków, lepszych enzymów czy rozwiązań dla produkcji żywności. W cyklu badawczym najczęściej występuje sprzężenie zwrotne, tj. uzyskane wyniki podpowiadają nowe koncepcje badawcze. Gwiazdką oznaczono zautomatyzowane etapy wysokoprzepustowe (*high throughput*). Wiele z innowacji w tym procesie badawczym wprowadzono dzięki ambitnym projektom genomiki strukturalnej. Genomika strukturalna stawia sobie za cel rozwiązanie struktury wszystkich białek kodowanych przez dany organizm, np. patogenną bakterię *Mycobacterium tuberculosis*, która powoduje gruźlicę

Inny pomysł polega na zmieszaniu roztworów białka i czynnika strącającego tuż przed wstrzyknięciem do wiązki promieniowania synchrotronowego lub laserowego. Ponieważ kryształy nie muszą być duże, wystarczy czas krystalizacji rzędu sekundy [5].

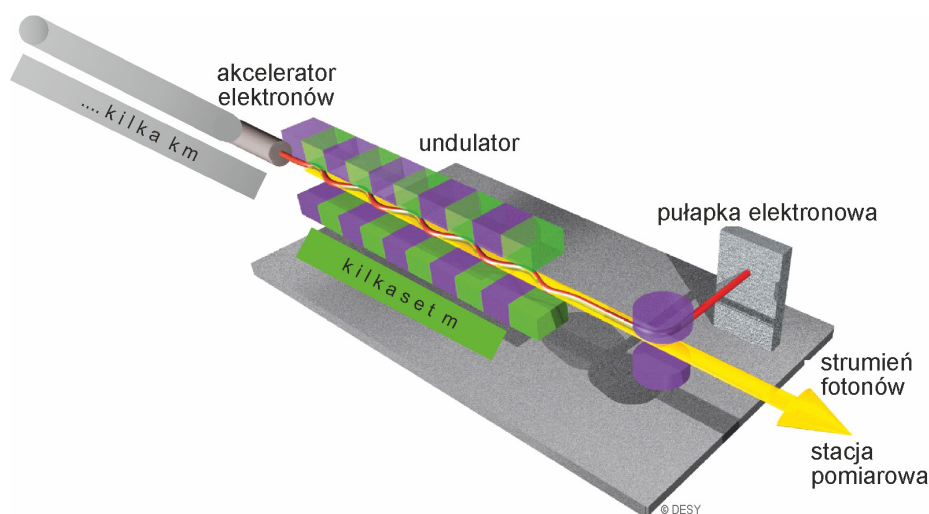
Promieniowania rentgenowskiego do badań dyfrakcyjnych kryształów makromolekularnych dostarczają dziś głównie synchrotrony nowej generacji [6]. Synchrotron to potężny cyklotron elektronowy, w którym przyspieszone do prędkości podświetlnych elektrony lub pozytony (elektrony antymaterii) krążą po torze zamkniętym, przelatując co rusz przez potężne pola magnetyczne generowane przez elektromagnesy nadprzewodzące (ryc. 5). Przelot relatywistycznych (tj. o podświetlnych prędkościach) elektronów przez pole magnetyczne wiąże się z dwoma zjawiskami: zakrzywieniem ich toru oraz emisją kwantów promieniowania elektromagnetycznego. Przy odpowiednim doborze parametrów pierścienia synchrotronowego będzie to promieniowanie rentgenowskie o niezwyklej intensywności, 10^{15} – 10^{17} fotonów na sekundę w standardowej geometrii (bardzo wąski kąt bryłowy) i zakresie energii emisji (bardzo wąski przedział energii, tj. pasmo przenoszenia). Intensywność promieniowania synchrotronowego można zwiększyć co najmniej o pięć rzędów wielkości (tj. sto tysięcy razy), umieszczając między magnesami uginającymi dodatkowe, śródwiązkowe urządzenia magnetyczne zwane wiglerami lub undulatorami. Są to multipole magnetyczne, zbudowane z wielu naprzemiennie ustawionych biegunów magnetycznych. Elektrony przechodzące przez zmienne pole magnetyczne w szczeliny undulatora poruszają się slalomem, generując jednocześnie dodatkowe fotony. Typowy synchrotron-fabryka fotonów ma kształt pierścienia, z którego wyprowadzone są stycznie do obwodu (w liczbie kilkudziesięciu) wiązki pomiarowe.



Ryc. 5. Synchrotron APS (*Advanced Photon Source*) w Argonne National Laboratory pod Chicago (miasto widoczne na horyzoncie) w USA (źródło: dr A. Joachimiak)

Ze względu na m.in. kolosalną energię, jaką musi „znieść” naświetlany kryształ, pomiary synchrotronowe prowadzi się zwykle w temperaturze 100 K (tj. -173°C), którą zapewnia skierowany na kryształ nadmuch oziębionych par azotu.

Fizyka wysokich energii poszła jednak dalej i dziś oferuje już lasery rentgenowskie na swobodnych elektronach (XFEL, *X-ray Free Electron Laser*). Laser rentgenowski to urządzenie otwarte, przypominające rozciąty i wyprostowany synchrotron (ryc. 6).



Ryc. 6. Schemat lasera rentgenowskiego na swobodnych elektronach XFEL (źródło: DESY)

Najpotężniejszy laser rentgenowski, European XFEL, działa w ośrodku DESY w Hamburgu. W istocie jest to urządzenie tak potężne (3,4 km długości), że swój początek (akcelerator) ma w Hamburgu, a koniec (stacje pomiarowe) w mieście Schenefeld w landzie Schleswig-Holstein. Akcelerator tego lasera to 1,7 km superprecyzyjnych wnęk rezonansowych wykonanych z rzadkiego metalu – niobu. Cały akcelerator umieszczony jest w tunelu o wysokiej próżni i zanurzony w ciekłym helu, by zapewnić temperaturę 2 K, w której niob staje się nadprzewodnikiem. Chodzi o to, by podtrzymanie przyspieszającego pola mikrofalowego we wnękach rezonansowych odbywało się jak najmniejszym nakładem energii. Po przejściu przez akcelerator, przyspieszone do prędkości bliskiej prędkości światła (a więc relatywistyczne) i „spakowane” w pakiety elektrony wlatują do długiego na kilkadziesiąt metrów undulatora, gdzie następuje ich dalsze porządkowanie w pakietach i pompowanie energią, aż do zajścia efektu SASE (*Self-Amplified Spontaneous Emission*), w którym cała zgromadzona energia wyemitowana jest w postaci bardzo krótkiego (10 femtosekund, tj. 10^{-14} sekundy) spójnego błysku o astronomicznej jasności 10^{25} fotonów. Po przeliczeniu na intensywność w maksimum emisji, otrzymujemy przyprawiającą o zawrót głowy wartość niemal 10^{34} fotonów na sekundę. Umoż-

liwia to pomiary kryształów o wymiarach ~ 100 nm, tj. tak małych, że złożonych zaledwie z około $10 \times 10 \times 10$ komórek elementarnych. Dla ilustracji: 100 nm to w przybliżeniu jedna tysięczna grubości włosa ludzkiego; a w czasie trwania błysku lasera XFEL, tj. 10 femtosekund, światło przebywa drogę zaledwie 3 mikronów, czyli jedną trzydziestą grubości włosa. Kryształów o wielkości 100 nm nie uda się zobaczyć nawet pod najlepszym mikroskopem optycznym, którego możliwości kończą się przy $1/2$ długości fali użytego światła. Dla najkrótszego, fioletowego, promieniowania widzialnego jest to granica ok. 200 nm.

Błysk lasera XFEL dostarcza do kryształu niewyobrażalną wprost energię 700 MGy, podczas gdy żadna materia nie jest w stanie „przeżyć” nawet 30 MGy (Gy, czyli grej, to jednostka pochłoniętej dawki promieniowania jonizującego, równa 1 dżul (J) na kg materii). Jak to więc możliwe, że rejestrujemy jakąkolwiek dyfrakcję, skoro błysk lasera niszczy kryształ, „odparowując” go do stanu plazmy? Otóż okazuje się, że przejście do stanu plazmy, a więc „odparowanie” elektronów (tj. całkowita jonizacja), a następnie eksplozja elektrostatyczna dodatnio naładowanych jąder atomowych, również wymaga czasu, w przybliżeniu 100 femtosekund. Jeśli więc błysk trwa ok. 10 femtosekund, to spokojnie zarejestrujemy wywołaną przezeń dyfrakcję; kryształ przestanie istnieć dopiero „po chwili”. Ten typ eksperymentu nazwano „Dyfrakcją przed Destrukcją”.

Pomiary dyfrakcyjne przy źródle XFEL prowadzi się, wstrzykując w bieg promienia laserowego cieniutką stróżkę zawiesziny nanokryształów. W tym samym czasie następują błyski laserowe, w Hamburgu z częstotliwością 20 000 Hz, tj. 20 tys. błysków na sekundę. Jeśli zawiesina jest dość gęsta, to większość błysków trafi w jakiś (przypadkowo zorientowany) kryształ, który właśnie znalazł się na drodze promienia. W czasie minuty można więc zarejestrować do 1 200 000 cząstkowych obrazów dyfrakcyjnych, z których bez trudu skompletuje się całość. Nawet gdyby tylko 10% (tj. 120 000) błysków było produktywnych, to i tak dane zostaną zarejestrowane z dużą redundancją. Na tym polega SFX, czyli szeregowo femtosekundowa nanokrystalografia rentgenowska [7].

Do niedawna woziliśmy do synchrotronów nasze drogie (bo w istocie droższe niż diamenty tej samej wielkości) kryształy białkowe o wymiarach ok. 0,1 mm. Obecnie mogą one mieć już rozmiar 100 nm, czyli $0,0001$ mm = 1000 Å. Jeśli udało się zejść o trzy rzędy wielkości, to czy nie będzie możliwy jeszcze mały kroczek i osiągnięcie wymiaru ~ 100 Å? Tak, taka możliwość jest już całkiem realna, lecz oznacza ona, że przez wiązkę lasera XFEL będą przelatywały nie kryształy, ale pojedyncze molekuly! Taki eksperyment, tj. rejestracja rozpraszania rentgenowskiego i wyznaczanie struktury pojedynczych makromolekuł biologicznych, jest już w zasięgu ręki. Będzie to osiągnięcie granicy eksperymentu w biologii strukturalnej: stosowanie metody opartej na krystalografii do obiektu (pojedynczej makromolekuły), który już nie jest kryształem!

Czasowo-rozdzielcza krystalografia białek

Aspekt czasowy pomiarów dyfrakcyjnych wiąże się z możliwością mapowania przemian chemicznych. Taką możliwość dają nawet pomiary synchrotronowe kryształów prowadzone metodą Lauego, a więc polichromatyczną. Metoda ta ma dwie zalety; ze względu na szerokie spektrum długości fali λ , cały obraz dyfrakcyjny można uchwycić na jednym zdjęciu; dodatkowo brak filtrowania fotonów przez monochromator sprawia, że dysponujemy potężną intensywnością promieniowania X, pozwalającą na skrócenie pomiaru do nanosekundy (10^{-9} sekundy). Przykładem procesu, który można badać w ten sposób, jest reakcja enzymatyczna, wiodąca od substratów do produktów, przez bogaty krajobraz energetyczny, przebiegający przez bardzo wysokie szczyty stanów przejściowych i głębokie doliny produktów pośrednich. Enzymy (tj. katalizatory biochemiczne) znacznie niwelują bariery energetyczne stanów przejściowych, umożliwiając w ogóle zajście reakcji, lecz ich nie likwidują zupełnie; ciągle przejście przez stan (*nomen omen*) przejściowy wymaga nakładu energii i odbywa się bardzo szybko, gdyż czas życia stanów termodynamicznych jest, zgodnie z prawem Arrheniusa, proporcjonalny do $e^{-E/kT}$, gdzie E oznacza energię, a T temperaturę bezwzględną. Produkty pośrednie, o stosunkowo niskiej energii, żyją nawet milisekundy, tj. wystarczająco długo, by wykonać w tym czasie setki, a nawet tysiące zdjęć dyfrakcji rentgenowskiej. Każdy z serii takich obrazów dyfrakcyjnych będzie odpowiadał konkretnemu stanowi na ścieżce reakcji. Montując wyznaczone z tych obrazów struktury jak kadry filmu, można uzyskać poklatkowy obraz przebiegu reakcji biochemicznej w kryształach.

Inny sposób badania przebiegu reakcji biochemicznych w 3D nosi nazwę MISC (*Mix & Inject Serial Crystallography*) [8]. Stosuje się w tym celu miniaturową dyszę (średnica całkowita 0,5 mm!), w której współosiowo podawane są dwa płyny: zawiesina kryształów enzymu oraz roztwór substratu, a w zewnętrznym płaszczu płynie ogniskujący strumień gazowego helu. W miejscu zwanym strefą ogniskowania dochodzi do zmieszania obu płynów i wówczas, na skutek dyfuzji substratu do kryształu z enzymem, rozpoczyna się reakcja enzymatyczna. Jej przebieg można śledzić przez dyfrakcję promieni X w różnych odstępach czasu, wyznaczonych przez prędkość przepływu i odległość punktu pomiarowego od strefy ogniskowania.

Biologia strukturalna w służbie medycyny

Nowoczesne leki powstają dziś głównie na drodze racjonalnego projektowania, mającego na celu stworzenie precyzyjnego „pocisku” molekularnego, który bezbłędnie i skutecznie trafi w makromolekularną „tarczę”. W podejściu tym najpierw prowadzi się rozpoznanie molekularne procesu patologicznego, a następnie identyfikuje odpowiedzialną za niego makromolekule-tarczę, zwykle enzym lub fragment kwasu nukleinowego. Najczęściej są to makromolekuły organizmów patogennych, wirusów, bakterii czy grzybów,

ale mogą one też być pochodzenia ludzkiego, np. w przypadku terapii genowych. Struktury makromolekuł stanowiące podstawę tych badań, jak również małych cząsteczek – potencjalnych kandydatów na leki, są zwykle ustalane metodami krystalograficznymi.

Stosuje się w tym celu dwa podejścia eksperymentalne, nazywane po angielsku *structure-based drug design* (SBDD, projektowanie leku wspomagane strukturą) oraz *fragment-based drug design* (FBDD, składanie leku z fragmentów). W metodzie SBDD wychodzi się od celu makromolekularnego (z angielska „targetu”), którym jest najczęściej enzym związany z procesem patologicznym (np. infekcją wirusową), dla którego ustala się (zwykle metodą krystalograficzną) strukturę 3D. Do struktury tej dopasowuje się następnie metodą projektowania molekularnego małą cząsteczkowy ligand, który jak *silver bullet* („cudowny pocisk”) będzie w stanie „unieszkodliwić” target (ryc. 2). W metodzie FBDD poszukuje się obiecujących fragmentów „pocisku”, nasączaając kryształy makromolekularnego targetu koktajlami przeróżnych związków chemicznych (w tym molekuł z uznanymi farmakoforami) i badając, które z nich wykazały powinowactwo do targetu, tj. utworzyły z nim kompleks. Tak zidentyfikowane „kawałki” łączy się następnie w większe cząsteczki, uzyskując obiecujące ligandy. Na ostatnim etapie ligandy te mogą być jeszcze optymalizowane metodą SBDD.

Biomolekularna spektroskopia NMR

W połowie lat osiemdziesiątych krystalografia, która dotąd była jedynym źródłem wiedzy o strukturze przestrzennej makromolekuł biologicznych, otrzymała wsparcie od spektroskopii jądrowego rezonansu magnetycznego, NMR. Metoda NMR pozwala wyznaczyć położenia atomów o niezerowym spinie jądrowym, takich jak ^1H , przez rejestrację absorpcji fal elektromagnetycznych o wysokiej częstotliwości (np. 800 MHz) przez próbkę umieszczoną w bardzo silnym polu magnetycznym. Ta możliwość wynika stąd, że absorpcja (a więc sygnał) konkretnego jądra zależy od jego otoczenia chemicznego. Dla cząsteczki etanolu, $\text{CH}_3\text{-CH}_2\text{-OH}$, posiadającej sześć protonów, rozwiązanie takiej zagadki jest dziecinnie łatwe. Jednak dla olbrzymich makromolekuł, zawierających tysiące lub dziesiątki tysięcy różnych chemicznie protonów, liniowe widmo ^1H NMR jest tak zatłoczone, że nie ma szans na jego rozwikłanie. Widmo ^1H NMR można „rozgęścić”, stosując techniki korelacyjne 2D, gdzie wybiórczo rozpręga się poszczególne sygnały H...H. W ten sposób powstają widma COSY (sprzężenia przez wiązania) i NOESY (*Nuclear Overhauser Spectra*, sprzężenia przez przestrzeń). Dla większych białek jednak i to nie wystarcza i konieczne są widma HSQC (*Heteronuclear Single-Quantum Coherence*) pokazujące sprzężenia protonów z innymi atomami. Niestety typowe izotopy ^{12}C , ^{14}N i ^{16}O budujące naturalne białka są nieaktywne w NMR. Dlatego należy badane białko wzbogacić w izotopy NMR aktywne, takie jak ^{13}C i ^{15}N . Wymaga to kosztownej produkcji białka rekombinowanego na pożywcze zawierającej takie właśnie izotopy

zamiast izotopów naturalnych. Wysoki koszt to jeden z minusów metody NMR; drugi to konieczność używania bardzo wysokich stężeń (niebezpieczeństwo agregacji białka), porównywalnych z warunkami krystalizacji, trzeci to ograniczenie do białek o niewielkiej masie, do 20 kDa. Do zalet metody NMR należy wizualizacja struktury w roztworze oraz jej dynamiki. Widma NMR można bowiem „rozwiązać” na kilka/kilkanaście równoważnych sposobów, ukazujących niektóre elementy struktury w różnych konformacjach, a więc pośrednio ich dynamikę. Pionierem zastosowań metody NMR do makromolekuł biologicznych jest noblista (2002) Kurt Wüthrich. Początkowe wielkie nadzieje związane ze spektroskopią NMR z czasem nieco przygasły i dziś odsetek struktur w PDB wyznaczonych tą metodą wynosi ~8%.

Rozpraszanie niskokątowe (bio-SAXS)

Jeśli mowa o badaniach struktury w roztworze, to warto wspomnieć metodę niskokątowego rozpraszania promieni X, SAXS (*Small Angle X-ray Scattering*). Podobnie jak w krystalografii, rozpraszaniu ulega tu promieniowanie synchrotronowe, z tym że próbka nie jest kryształem, lecz zawiera badane makromolekuły w roztworze. Nie otrzymamy w związku z tym silnie wzmocnionej informacji skoncentrowanej w pikach (refleksach) braggowskich, lecz ciągłe rozpraszanie, które dla małych kątów θ zależy od wielkości i kształtu obiektów rozpraszających, tj. makromolekuł obecnych w roztworze. Po transformacji krzywej rozpraszania otrzymujemy tzw. funkcję par, informującą o rozkładzie odległości międzyatomowych w cząsteczce. Można stąd wyznaczyć maksymalny rozmiar molekuly oraz najczęściej występujące odległości atom-atom. Stąd, przy pewnej dozie wyobraźni, można wydedukować kształt badanej makromolekuly. Można też, budując modele gruboziarniste, próbować zbudować układ najlepiej odtwarzający krzywą doświadczalną. Choć modele SAXS są bardzo zgrubne i prowizoryczne, jedna rzecz jest łatwa do określenia, mianowicie tworzenie kompleksów i oligomerów w roztworze.

Kriomikroskopia elektronowa (cryo-EM)

Metodą, która w ostatnim czasie spowodowała prawdziwą rewolucję w biologii strukturalnej, jest kriomikroskopia elektronowa (cryo-EM). Mikroskop elektronowy wynaleziony przez Ernsta Ruskę w 1931 roku (Nagroda Nobla w 1986 roku) przypomina mikroskop optyczny, z tym że źródło światła zastąpione jest działem elektronowym, a soczewki szklane soczewkami magnetycznymi. Przyspieszone ogromnym napięciem (w najnowszych mikroskopach 300 tys. elektronowoltów) elektrony poruszają się z prędkością bliską prędkości światła, a ich ruchowi towarzyszy fala materii de Broglie’a o długości λ ułamka angstroma. O ile więc światło widzialne (najkrótsza długość fali fioletowej to 3800 Å) nie nadaje się do obserwacji obiektów o wielkości angstromowej, o tyle elektronowa fala materii nadaje się jak najbardziej. Problemem przez wiele dekad

była rozdzielczość mikrografii elektronowych, wynosząca kilkadziesiąt, w najlepszym razie kilkanaście angstromów. Powodem była kiepska jakość źródeł elektronów, „optyki” magnetycznej i brak dobrych detektorów dla słabo przenikliwych elektronów. Przede wszystkim jednak nie umiano przygotować dla mikroskopu transmisyjnego (TEM) preparatu z materiałem biologicznym, który zawieszony jest w wodzie (słabo kontrastującej z zawierającym podobną gęstość elektronową białkiem) i nie może być zwyczajnie umieszczony w wysokiej próżni kolumny mikroskopu. Przełom przyszedł dzięki pracom trzech badaczy, uhonorowanych za to Nagrodą Nobla w 2017 roku. Najpierw Richard Henderson wpadł na pomysł, by mrozić preparat mikroskopowy w ciekłym azocie. Pomysł był dobry, ale wynik średni, gdyż mrożenie w ciekłym azocie powodowało krystalizację lodu i degradację preparatu białkowego. Jacques Dubochet zastąpił azot ciekłym etanem, który ma większą pojemność cieplną i witrifikuje (tj. przeprowadza w stan szklisty) uwodnione próbki błyskawicznie. Z kolei Joachim Frank opracował algorytmy komputerowe pozwalające z zaszumionych mikrografii elektronowych wydobyć użyteczną informację strukturalną, podobne obserwacje uśrednić i na tej podstawie zrekonstruować trójwymiarowy obraz badanej makromolekuły. To wszystko, w połączeniu z postępem technicznym w zakresie „optyki” mikroskopu elektronowego, oraz przede wszystkim wynalezienie bardzo czułych detektorów pojedynczych elektronów, sprawiło, że za pomocą nowoczesnego mikroskopu cryo-EM można uzyskać rekonstrukcję struktury przestrzennej pojedynczej makromolekuły z rozdzielczością lepszą niż 2 Å, a więc dorównującą przeciętnemu eksperymentowi krystalograficznemu [9]. Rozdzielczość obrazu cryo-EM nie jest jednak jednolita i może być znacznie gorsza z dala od dobrze określonego środka obrazu.

Sztuczna inteligencja

Równolegle do tych wszystkich wspaniałych osiągnięć na polu eksperymentalnym, w roku 2021 do biologii strukturalnej wkroczyła z impetem sztuczna inteligencja (AI). Wtedy właśnie stworzono oparte na sieciach neuronowych dwa algorytmy głębokiego uczenia maszynowego (Deep Learning), które nauczono – dzięki istniejącej informacji doświadczalnej – przewidywać strukturę trzeciorzędową (3D) łańcuchów białkowych na podstawie ich sekwencji, tj. struktury pierwszorzędowej (1D). Pierwszy z tych programów, o nazwie AlphaFold, opracowała firma DeepMind należąca do Alphabet/Google [10]. Drugi program, RoseTTAFold, jest dziełem bioinformatyków zespołu prof. Davida Bakera na Uniwersytecie Waszyngtona w Seattle [11]. Nie są to algorytmy zwijania białka, lecz rozpoznawania wzorców strukturalnych w sekwencji. Przewidują bardzo dobrze struktury białek, dla których istnieją homologi w strukturalnych bazach danych, lecz słabo radzą sobie z sekwencjami zupełnie nowymi (unikatowymi) oraz z białkami oligomerycznymi. W takich przypadkach, nawet gdyby predykcja była poprawna, rezultat jest

ciągle niepewny i wymaga weryfikacji doświadczalnej. Algorytmy te działają błyskawicznie, przewidując struktury białek w czasie od minut do ułamków sekund. W chwili opublikowania algorytmu AlphaFold autorzy ogłosili, że program przewidział już strukturę ponad 260 tys. białek. Wkrótce liczba ta przekroczyła 220 mln. Jakby tego wszystkiego było mało, w 2023 roku opublikowano trzeci program, ESMFold, oparty na modelach językowych, stworzony przez Meta/Facebook [12]. W chwili opublikowania tego algorytmu autorzy ogłosili, że program przewidział już strukturę ponad 600 mln metagenomowych białek i że 1 mld jest w bliskim zasięgu... Trudno powiedzieć, czy ta numeryczna sztuka dla sztuki ma jakiś sens. Możemy „rzucić” miliardem, 10 miliardami, czy miliardem miliardów, ale co z tego? Nikt nie jest w stanie przetrawić tej informacji, a nawet powiedzieć, czy jest ona cokolwiek warta. Sztuczna inteligencja ma niewątpliwe zalety, choćby szybkość znajdowania odpowiedzi. Jednak te odpowiedzi, np. przewidziane struktury białek, są tylko sprytnym powtórzeniem tego, co już widziano. W dodatku masowe zalewanie pola naukowego pomieszaną informacją prawdziwą i nieprawdziwą, a przynajmniej niezawierającą weryfikacji i niepewną, może być bardzo groźne dla przyszłości nauk doświadczalnych.

Podsumowanie i perspektywy

Biolog strukturalny jest dziś w zupełnie innym świecie niż jeszcze 25 lat temu. Ma w ręku narzędzia, takie jak laser XFEL czy mikroskop cryo-EM, pozwalające badać w bardzo krótkim czasie struktury nawet pojedynczych biomolekuł, a także śledzić ich przemiany w procesach biochemicznych. Na dodatek eksperyment wsparty jest potężnymi narzędziami obliczeniowymi, ze sztuczną inteligencją włącznie. Czy pojawienie się na scenie AI oznacza koniec eksperymentu? Z całą pewnością nie, przynajmniej w możliwej do przewidzenia przyszłości. Nawet jeśli AI łatwo i szybko wygeneruje interesujący nas model, to w przypadku zupełnie nowego obiektu biostrukturalnego będzie trzeba i tak potwierdzić go doświadczalnie. A czy może, jak to już wielokrotnie prognozowano, umrze krytalografia? Wydaje się, że w możliwej do przewidzenia przyszłości na pewno nie. Choć jest to metoda trudna, gdyż w przeciwieństwie do innych wymaga pokonania problemu krystalizacji, wyniki, jakich dostarcza, są nieprześcignione pod względem jakości, rozdzielczości i możliwości weryfikacji w oparciu o bogate dane doświadczalne. Poza tym innym metodom potrzebne są wiarygodne modele do uczenia maszynowego czy rozpoznawania obrazu. Wydaje się więc, że omówione metody będą współistniały na zasadzie komplementarności.

Podsumowując naszą opowieść, można powiedzieć, że biologia strukturalna przeszła z fazy „problem driven” do fazy „technology driven”. Dziś to nie pytania badawcze inspirowały inżynierów, lecz na odwrót: niesamowite wprost osiągnięcia techniki motywują badaczy do poszukiwania problemów, do których te nowoczesne technologie można by

zastosować. Jedno jest pewne: żyjemy w czasach gwałtownej rewolucji – metodologicznej i koncepcyjnej – w biologii strukturalnej. Trudno dobrze przewidzieć, co czeka nas w przyszłości. Przewidywania są zresztą zawsze niepewne, szczególnie gdy dotyczą przyszłości.

Bibliografia

- [1] F.L. Hünfeld (1840) *Die Chemismus in der tierescher Organisation*. p. 160, FA Brockhouse, Leipzig.
- [2] K. Dziubek, M. Jaskólski, A.B. Więckowski (2023). O mitach w nauce na kanwie „oryginalnej aparatury Lauego”. *Nauka* **1/2023**, 107–121.
- [3] M. Jaskolski, Z. Dauter, A. Włodawer (2014). A brief history of macromolecular crystallography, illustrated by a family tree and its Nobel fruits. *FEBS J.* **281**, 3985–4009.
- [4] R. Schönherr, J.M. Rudolph, L. Redecke (2018). Protein crystallization in living cells. *Biol. Chem.* **399**, 751–772.
- [5] A. Henkel et al. (2023). JINXED: just in time crystallization for easy structure determination of biological macromolecules. *IUCr* **10**, 253–260.
- [6] Z. Dauter, M. Jaskolski (2023). Zastosowanie promieniowania synchrotronowego w krystalografii białek. [W:] *Promieniowanie synchrotronowe w fizyce i chemii ciała stałego: wybrane zagadnienia* (red. B.J. Kowalski, W. Paszkowicz, E.A. Gorlich), s. 527–566. Wydawnictwo Naukowe UAM Poznań.
- [7] H.N. Chapman et al. (2011). Femtosecond X-ray protein nanocrystallography. *Nature* **470**, 73–33.
- [8] J.L. Olmos et al. (2018). Enzyme intermediates captured “on the fly” by mix-and-inject serial crystallography. *BMC Biology* **16**, 59.
- [9] T. Nakane et al. (2020). Single-particle cryo-EM at atomic resolution. *Nature* **587**, 152–156.
- [10] J. Jumper et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589.
- [11] M. Baek et al. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871–876.
- [12] Z. Lin et al. (2022). Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* **379**, 1123–1130.

Biologia strukturalna: historia, terażniejszość i perspektywy

Biologia strukturalna bada strukturę przestrzenną, tj. 3D, makromolekuł biologicznych, głównie białek i kwasów nukleinowych. Początkowo utożsamiana była z krystalografią białek, która dała jej początek. Metoda krystalograficzna polega na badaniu struktury materiału biologicznego za pomocą dyfrakcji promieni X. W latach 80. ubiegłego wieku wsparcie przyszło ze strony jądrowego rezonansu magnetycznego (NMR), ale prawdziwy przełom nastąpił w ostatniej dekadzie wraz z rozwojem wysokorozdzielczej kriomikroskopii elektronowej (cryo-EM).

W 1971 r. powołano Protein Data Bank (PDB) z misją gromadzenia wyznaczonych eksperymentalnie struktur makromolekuł. Po okresie powolnego wzrostu, w połowie lat 90. XX w. nastąpiła „eksplozja” PDB, związana z rozwojem biotechnologii oraz z powszechnym zastosowaniem synchrotronowych źródeł promieniowania X. W PDB jest obecnie ponad 210 tys. wyznaczonych doświadczalnie struktur makromolekuł, a roczny przyrost sięga 15 tys. Wiele z tych struktur, np. białek wirusa HIV czy SARS-CoV-2, ma znaczenie medyczne, gdyż na ich podstawie projektuje się leki nowej generacji. Postęp techniczny poszedł dalej i dziś dysponujemy laserami rentgenowskimi na swobodnych elektronach (XFEL) o niewyobrażalnej jasności promieniowania X. Umożliwia to skrócenie czasu naświetlania do kilkunastu femtosekund oraz badanie kryształów o wymiarach nanometrów, a nawet pojedynczych molekuł. Obok rewolucji na polu doświadczalnym, przełom nastąpił też w metodach przewidywania struktury 3D białek na podstawie ich sekwencji, tj. informacji jednowymiarowej. Stało się to możliwe dzięki algorytmom uczenia maszynowego, które wytrenowano pod kątem rozpoznawania w sekwencjach białek wzorców strukturalnych. Dla łatwiejszych problemów sztuczna inteligencja (AI) potrafi znaleźć poprawne rozwiązanie błyskawicznie. Dostępny powszechnie serwer AlphaFold zaproponował już strukturę dla ponad 200 mln białek. Inny serwer obiecuje miliard struktur, choć wydaje się, że taka numerologiczna sztuka dla sztuki traci już sens. Czy pojawienie się na arenie AI oznacza koniec eksperymentu? Z całą pewnością nie. Nawet jeśli AI wygeneruje interesujący nas model, to w przypadku zupełnie nowego białka trzeba będzie i tak potwierdzić go doświadczalnie. Poza tym innym metodom potrzebne są wiarygodne modele do uczenia maszynowego czy rozpoznawania obrazu. Wydaje się, że omówione metody będą współistniały na zasadzie synergii.

Słowa kluczowe: biologia strukturalna, krystalografia, dyfrakcja rentgenowska, makromolekuły biologiczne

Structural biology: its history, present status, and perspectives

Structural biology is concerned with the three-dimensional atomic structure of the molecules of life, proteins and nucleic acids. It was born in mid-1950s with a visionary application of X-ray diffraction to structure determination of protein crystals, and for several decades “structural biology” and “protein crystallography” were synonymous. In the 1980s structural biology received new experimental support from NMR spectroscopy, but a true breakthrough occurred only recently, with the development of atomic-resolution cryo-electron microscopy (cryo-EM), enabling direct visualization of macromolecular objects without the need of growing crystals. The Protein Data Bank (PDB) was created in 1971 with merely seven protein structures known. In mid-1990s the PDB entered an explosive growth phase, ignited by advances of biotechnological methods of protein production and, even more importantly, by widespread use of synchrotrons

as extremely powerful X-ray sources. The technological advances did not stop there, and today we have on offer ever more powerful X-ray Free Electron Lasers (XFELs), producing astronomically bright femtosecond X-ray pulses, which allow studying the structure of nanometer-sized crystals or even of single macromolecules. Thanks to all those methodological developments, the PDB holds today over 210,000 experimental macromolecular structures, many of which (such as those related to HIV or SARS-CoV-2) have fundamental importance for medicine as targets for rational drug design. In addition to innovative experimental methodology, structural biology has recently seen a huge progress of artificial intelligence (AI)-based methods of protein structure prediction, capable now of quite accurate divination of the three-dimensional structure for billions of protein sequences in very short time. However, those machine-learning algorithms, such as AlphaFold, recognize patterns that have been seen before, while for truly new sequences and for oligomeric proteins the prediction is still less than certain and needs experimental validation. It appears then that experimental structural biology is not quite dead yet and will remain the main source of reliable novel structural information for the foreseeable future.

Key words: structural biology, crystallography, X-ray diffraction, molecules of life