

Informatyka w służbie biologii

Gromadzone przez naukowców dane i dokonywane odkrycia oraz generowanie niesamowitych ilości informacji przysparzają problemów i fundują badaczom mnóstwo żmudnych analiz. Biologia wchodzi w erę cyfryzacji z podniesionym czołem, a terabajty danych przekazywane są specjalistom – bioinformatykom, którym udało się połączyć w swojej dziedzinie wszystkie najważniejsze umiejętności matematyczne, informatyczne i biologiczne.

Joanna Stojak

Na barkach bioinformatyków spoczywa ogromna odpowiedzialność. Do ich obowiązków należy bieżące katalogowanie mierzonych się informacji biologicznych w postaci baz danych z odpowiednimi adnotacjami (ramka 1).

Bioinformatyka zajmuje się także analizą sekwencji DNA (składanie sekwencji, adnotacje, poszukiwanie sekwencji regulatorowych, kodujących, markerów molekularnych) i genomów. Dzięki temu możliwe jest opisywanie wzajemnych relacji ewolucyjnych różnych organizmów, zbadanie ich filogenezy oraz genotypowanie, tak popularne w kryminalistyce i medycynie (np. w onkologii do poszukiwania genów odpowiedzialnych za zmiany nowotworowe). Narzędzia bioinformatyczne ułatwiają także analizę ekspresji genów (głównie mikromacierzowo) oraz analizy proteomiczne (porównywanie sekwencji białek, identyfikacja domen i motywów białkowych, przewidywanie struktury i lokalizacji w komórce). Co więcej, katalogowanie funkcji genów i białek ułatwia badanie szlaków metabolicznych i sygnalizacyjnych.

Ostatecznie bioinformatyka zajmuje się tworzeniem algorytmów i programów komputerowych,

które umożliwiają pełną analizę danych biologicznych, w tym właśnie cząsteczek takich jak DNA, RNA czy białka.

Co z tym białkiem?

Białkiem nazywamy cząsteczki zbudowane z reszt aminokwasów połączonych wiązaniami peptydowymi (-CONH-), których synteza przebiega w rybosomach. Większość białek to cząsteczki złożone, z przyłączonymi do reszt aminokwasowych m.in. cukrami lub jonami metali.

Z reguły białko musi przejść proces zwijania, najczęściej za pomocą białek pomocniczych, tzw. chaperonów, przyjmując strukturę przestrzenną zwaną konformacją białka (ramka 2). Te, którym się ta sztuka nie uda lub przyjmą nieprawidłową konformację, zostają zniszczone – nie mogą przecież pełnić wtedy swojej funkcji.

Strukturę białek można opisać pionowo. Najniższą i podstawową strukturą, zwaną strukturą pierwszorzędową, jest liniowa kolejność aminokwasów w łańcuchu polipeptydowym. Przestrzenne ułożenie fragmentów łańcuchów białkowych nazwano strukturą drugorzędową. Fragmenty te mogą przybierać kształt helisy α lub harmonijki β (ramka 3).

Wzajemne ułożenie tych elementów względem siebie stanowi strukturę trzeciorzędową, a naj-

Ramka 1. Popularne serwisy biologiczne

- NCBI (National Center for Biotechnology Information) – baza przechowująca sekwencje nukleotydowe (baza GeneBank), abstrakty i informacje o artykułach biomedycznych (baza PubMed) oraz inne wiadomości z zakresu biotechnologii.
- PDB (Protein Data Bank) – bank danych białkowych dotyczących struktury przestrzennej białek i kwasów nukleinowych. Można w nim znaleźć graficzne rekonstrukcje oraz dokładny opis cząsteczek: sekwencje aminokwasowe, współrzędne atomów i informacje o strukturze drugorzędowej.
- EBI (European Bioinformatics Institute) – centrum badań i usług bioinformatycznych, część Europejskiego Laboratorium Biologii Molekularnej (EMBL). Podstawowym celem było stworzenie centralnej bazy gromadzącej wszystkie poznane sekwencje DNA. Obecnie składa się z dwóch baz danych: sekwencji nukleotydowych (EMBL-Bank) i sekwencji białkowych (Swiss-Prot-TrEMBL, znanej jako UniProt).
- Pfam (Protein families) – baza rodzin białkowych wraz z adnotacjami i porównaniami wielu sekwencji (MSA, ang. *multiple sequence alignments*) wykonanymi z wykorzystaniem ukrytych modeli Markowa (ciąg, w którym prawdopodobieństwo każdego kolejnego kroku zależy tylko od wyniku kroku poprzedniego).

wyższą, czwartorzędową, tworzy wzajemne ułożenie łańcuchów białkowych i innych struktur niebiałkowych (zwanymi również grupami prostetycznymi, np. lipidy, cukry, barwniki).

Białko białku nierówne

Główną metodą uzyskiwania pełnej informacji o strukturze białek jest krystalografia rentgenowska lub NMR. Jest to jednak zadanie bardzo zmusne (ramka 4). Narzędzia bioinformatyczne pozwalają na przewidywanie struktury i sekwencji białek, wykorzystując metody porównawcze (ewolucja zachowuje funkcjonalne fragmenty białek jako konserwatywne domeny, przeszukiwanie baz danych w celu wykrycia podobieństwa do znanych rodzin białkowych) lub korzystające z termodynamiki (naprężenia w strukturze, grupowanie, powinowactwo między pewnymi grupami chemicznymi).

Modelowanie molekularne jest wykorzystywane np. w projektowaniu leków w oparciu o strukturę celu, poszukiwaniu miejsc funkcjonalnych oraz analizowaniu interakcji białko – białko. Dokładność symulacji zależy od mocy obliczeniowej komputera, stąd pomysł tworze-

Ramka 2. Chaperony i białka szoku cieplnego

Chaperony (ang. *chaperone* – opiekun) są białkami, które łączą się w sposób odwracalny z N-końcem nici polipeptydowej, zmierzającej do uzyskania zwiniętej konformacji. Chaperony nie mają wpływu na przebieg fałdowania białka, wynikający wyłącznie z sekwencji aminokwasowej. Dodatkową funkcją chaperonów jest stabilizacja struktury fałdującego się białka w trakcie przemieszczania go do określonych kompartmentów (np. z cytoplazmy do mitochondrium).

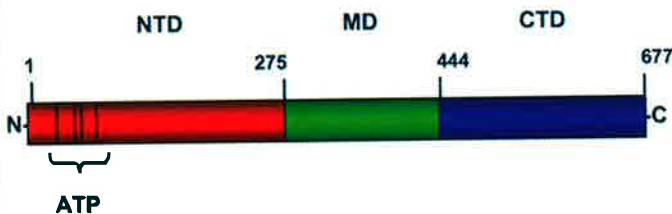
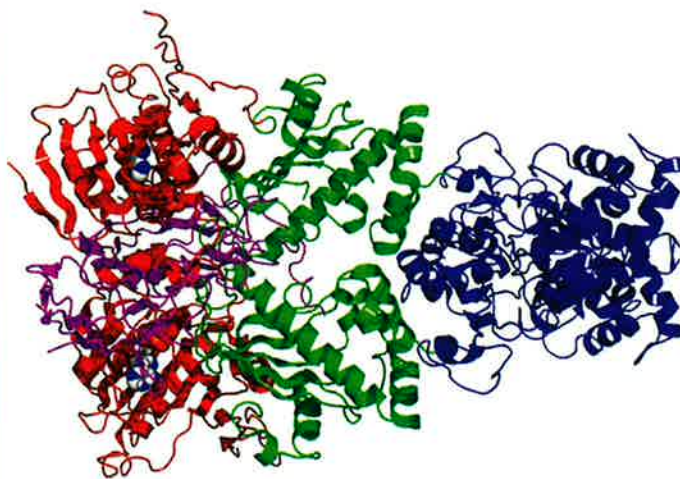
Chaperony są ATPazami. Początkowo istniejący kompleks ADP-chaperon wykazuje olbrzymie powinowactwo do niepołażowanych łańcuchów polipeptydowych, a po ich połączeniu ADP jest zastępowany przez ATP. Hydroliza ATP uwalnia połażowane białko, a wolne tempo tej reakcji sprawia, że łańcuch odłącza się stopniowo, ulegając tym samym poprawnemu fałdowaniu.

Większość chaperonów należy do rodziny wysoce konserwatywnych (zarówno w komórkach eukariotycznych, jak i prokariotycznych) białek Hsp (ang. *heat shock proteins*, białka szoku cieplnego).

Hsp70 – wiąże powstające łańcuchy już podczas translacji, uczestnicząc w ich transporcie do mitochondriów i retikulum endoplazmatycznego.

Hsp60 – ułatwia fałdowanie się łańcuchów polipeptydowych, a tym samym przyjęcie przez nie odpowiedniej konformacji. Hsp60 przypomina kształtem cylinder, do którego wnętrza transportowane są niepołażowane łańcuchy, które następnie stopniowo są z niego uwalniane i przybierają poprawną strukturę przestrzenną.

Hsp90 – ekspresja tych białek wzrasta wraz z nasileniem się czynników stresowych działających na komórkę. Białka te regulują aktywność ATPazową i umożliwiają połączenie odpowiedniego chaperonu z łańcuchem polipeptydowym.



Rys. 1. Budowa Hsp 90

Ramka 3. Struktura drugorzędowa białka

Helisa α – struktura kształtem przypominająca cylinder utworzony przez prawoskrętnie skręconą śrubę. Ściany cylindra tworzy łańcuch białkowy, a łańcuchy boczne wysunięte są na zewnątrz. Struktura stabilizowana jest przez wiązanie wodorowe między grupą karboksylową jednego a grupą aminową drugiego aminokwasu (co cztery aminokwasy w łańcuchu białkowym).

Harmonijka β – przestrzenne ułożenie aminokwasów w łańcuchu białkowym przypomina połażowaną w harmonijkę kartkę papieru. Kształt tej harmonijki jest stabilizowany przez wiązania wodorowe między sąsiednimi, oddzielnymi fragmentami łańcucha, tzw. nićmi beta. Można wyróżnić trzy rodzaje harmonijek: równoległe, antyrównoległe i mieszane.

nia projektów takich jak Rosetta@home, które angażują wolontariuszy z całego świata i udostępniają swoje serwery przez całą dobę.

Przewidywania *in silico* (łac. w krzemie, tu: w komputerze) są wprawdzie obciążone prawdopodobieństwem błędu, jednak są szybsze i tańsze niż metody tradycyjne. Pozwalają na przeanalizowanie setek, a nawet tysięcy cząsteczek w krótkim czasie, pomagając w formułowaniu hipotez, które następnie są analizowane doświadczalnie. Na czym to polega?

Założmy, że otrzymaliśmy jakąś sekwencję białkową. Jeśli jest ona identyczna z inną znaną sekwencją, wiadomo już, które to białko. Ciekawie robi się dopiero wtedy, gdy sekwencja okazuje się nieznaną. Może być ona podobna do innej – wiemy wtedy, że znaleźliśmy nowego członka jakiejś znanej rodziny. Jeśli znaleźliśmy w niej fragmenty podobne do znanych nam konserwatywnych motywów, regionów lub domen, możemy podejrzewać, jaką funkcję będzie pełniło badane białko. W przypadku całkowitego braku podobieństwa do jakiegokolwiek znanego nam białka jedynym wyjściem jest przeprowadzenie pełnego eksperymentu, czyli komputerowego poszukiwania jego struktury i porównania jej z rzeczywistym, skryształizowanym białkiem.

Należy jednak zadać sobie pytanie, czy aby na pewno podobieństwo krótkich fragmentów nie jest przypadkowe oraz czy białka o tej samej funkcji muszą należeć do jednej rodziny? I czy wszystkie spokrewnione ze sobą białka na pewno pełnią podobne funkcje?

Rozpatrzmy dwa istotne dla tej kwestii procesy ewolucyjne: dywergencję i konwergencję. Dywergencja (ewolucja rozbieżna) dotyczy różnokierunkowego kształtowania się cech homologicznych w wyniku oddziaływania różnych warunków środowiska (np. różne typy kończyn ssaków). Z kolei konwergencja, zwana także ewolucją zbieżną, polega na powstawaniu funkcjonalnie i morfologicznie podobnych cech ana-

Ramka 4. Jak poznać strukturę skryształizowanego białka?

- Krystalografia rentgenowska – najstarsza metoda krystalograficzna stosowana w celu ustalenia wielkości, kształtu i pełnej struktury cząsteczki białka. Rejestracja wielu obrazów dyfrakcji promieni rentgenowskich powstających w wyniku interakcji z powierzchnią kryształów białka pod różnymi kątami pozwala na odtworzenie pełnej jego struktury.
- Spektroskopia NMR (magnetycznego rezonansu jądrowego) – szybkie zmiany pola magnetycznego powodują wzbudzenie spinów jądrowych znajdujących się w zewnętrznym polu magnetycznym oraz emisji promieniowania elektromagnetycznego, które jest wynikiem relaksacji (powrót układu spinów jądrowych do stanu równowagi termodynamicznej). Rejestracja tego promieniowania pozwala poznać budowę białka.
- Mikroskopia elektronowa – mikroskop wykorzystuje do obrazowania wiązkę elektronów, co pozwala na zbadanie cząsteczki atom po atomie, umieszczonej w próżni.

logicznych u organizmów odlegle spokrewnionych w odpowiedzi na podobne warunki środowiskowe (np. posiadanie skrzydeł przez ptaki i owady).

Sekwencje homologiczne łączą pochodzenie od wspólnego przodka – jeżeli rozdzielenie ich nastąpiło w wyniku duplikacji, sekwencje te nazywane są paralogami, natomiast gdy zaszła specjacja – ortologami. Doskonałym przykładem są trzy homologiczne białka: hemoglobina, leghemoglobina i mioglobina. Wszystkie trzy przenoszą tlen, różnią się jednak miejscem występowania – hemoglobina znajduje się w erytrocytach, leghemoglobina w brodawkach korzeniowych roślin motylkowych, a mioglobina w mięśniach poprzecznie prążkowanych. Natomiast sekwencje analogiczne pełnią tę samą funkcję, w wyniku czego są do siebie podobne, nie łączy ich jednak wspólne pochodzenie ewolucyjne.

Z powyższych rozważań wypływa zatem jasny wniosek – sekwencje homologiczne są podobne, ale nie wszystkie sekwencje podobne muszą być sekwencjami homologicznymi.

Misja na lata

Bioinformatyka jako nauka stosunkowo młoda ma przed

sobą świetlaną przyszłość i mnóstwo problemów do rozwiązania. Kolejne białka o nieznannej strukturze krystalizowane w laboratoriach chemicznych coraz częściej potwierdzają, że stworzone w komputerze modele są poprawne, i tym samym podnoszą prestiż programów opracowanych do przewidywania struktury białek.

Na swoje modele czeka jednak nie tylko olbrzymia ilość białek, ale i mnóstwo cząsteczek kwasów nukleinowych. Czy bioinformatyczne metody modelowania, tak skuteczne w przypadku białek, sprawdzą się także podczas poszukiwań innych struktur, czy też bioinformatyków czekają kolejne innowacje i nowe programy? W końcu jednak *nic w biologii nie ma sensu, jeśli rozpatruje się to w oderwaniu od ewolucji* (T. Dobzhansky). A czy ewolucja RNA i białek na pewno przebiegała zupełnie inaczej?

I tu okazuje się, że bioinformatycy po raz kolejny zwyciężają – zastosowanie w modelowaniu struktur RNA metod stosowanych przy modelowaniu struktur białkowych przynosi efekty i sprawdza się bardzo dobrze.

mgr Joanna Stojak

Institut Biologii Ssaków PAN w Białowieży