

Nagroda Nobla z chemii za AlphaFold, czyli przewidywanie kształtu białek

© Margit Kossobudzka

Gazeta Wyborcza, 09.10.2024

https://wyborcza.pl/7,75400,31370782,nagroda-nobla-z-chemii-za-alphafold-czyli-przewidywanie-ksztaltu.html#s=S.embed_article-K.C-B.1-L.1.zw

Tegoroczną [w 2024] Nagrodę Nobla z dziedziny chemii otrzymują David Baker, John M. Jumper i Demis Hassabisa za program AlphaFold.

Już w lipcu tego roku pisaliśmy, że sztuczna inteligencja jest blisko pierwszej Nagrody Nobla. Spełniły się przewidywania naukowców, którzy prognozowali właśnie taką nagrodę!

Twórcy sztucznej inteligencji AlphaFold zostali w zeszłym roku laureatami Breakthrough Prize w naukach o życiu, dostali też medyczną nagrodę Laskera, uznawaną za przedsiónek do Nagrody Nobla. Niedawno ukazała się wersja programu, która może zrewolucjonizować badania biologiczne i medyczne.

Nagroda Nobla z chemii za AlphaFold, czyli przewidywanie kształtu białek

Niektórzy twierdzą, że ta rewolucja już się dokonuje. Zaczęła się w pandemii, kiedy firma Deep Mind (należy obecnie do Google'a) pokazała wersję AlphaFold, która nauczyła się przewidywać kształty, w jakie zwijają się nici białek, złożone z długich łańcuchów aminokwasów.

Białka to podstawowy budulec żywych organizmów. A o tym, jak działają i co robią w naszych komórkach, w dużym stopniu decyduje trójwymiarowa struktura, w jaką się zwijają.

Dekady zajęło naukowcom opisanie tylko niewielkiej liczby białek, tymczasem program AlphaFold błyskawicznie ustalił trójwymiarowe struktury prawie wszystkich 20 tys. białek spotykanych w ludzkich komórkach i tkankach.

Dwa lata później skompletował przewidywane kształty ponad 200 mln białek, czyli wszystkich, jakie znamy dziś na Ziemi.

Połowa Nagrody Nobla z dziedziny chemii trafia do Demisa Hassabisa i Johna Jumpera, którzy wykorzystali sztuczną inteligencję, aby pomyślnie rozwiązać problem, z którym chemicy zmagali się przez ponad 50 lat: przewidzieć trójwymiarową strukturę białka na podstawie sekwencji aminokwasów.

Drugą połowę nagrody otrzymuje David Baker, który opracował skomputeryzowane metody osiągnięcia tego, co wiele osób uważało za niemożliwe: stworzenia białek, które wcześniej nie istniały i które w wielu przypadkach mają zupełnie nowe funkcje.

Pierwszy, najważniejszy, czyli białko

Białka to podstawa życia i funkcjonowania naszego organizmu. Jak są ważne, dowodzi fakt, że czasami używana jest ich inna nazwa - proteiny (wywodzi się ona z języka greckiego od słowa protos – „pierwszy, najważniejszy”).

Wydawałoby się, że już wszystko o nich wiemy. Podstawowym elementem budującym białka są aminokwasy (jest ich 20), które łącząc się ze sobą, tworzą białka właśnie. Jednak w przyrodzie nic nie jest takie proste.

Białka to jednak wielocząsteczkowe związki chemiczne o skomplikowanej budowie. Różnią się własnościami w zależności od liczby aminokwasów i ich wzajemnego położenia w cząsteczce.

Dodatkowo, ze względu na budowę chemiczną białka dzieli się na proste i złożone. Białka proste zbudowane są tylko z aminokwasów, natomiast białka złożone oprócz aminokwasów mają także związki niebiałkowe, tzw. grupy prostetyczne (reszta kwasu fosforowego, kwasy nukleinowe, hem, atom metalu ciężkiego, węglowodany, lipidy). Należą do nich fosfoproteidy, nukleoproteiny, chromoproteiny, metaloproteiny, glikoproteiny i lipoproteiny.

Białka pełnią też liczne funkcje: strukturalne (kolagen, elastyna, keratyny, glikoproteiny), enzymatyczne (enzymy), ochronne (immunoglobuliny, fibrynogen, interferon), transportowe (hemoglobina, albuminy osocza, lipoproteiny, transferyna), białka biorące udział w skurczu (aktyna, miozyna), hormony (insulina, glukagon, parathormon, kalcytonina), białka błon komórkowych...

Bogactwo białek w przyrodzie wręcz przytłacza. Jak się odnaleźć w tym gąszczu?

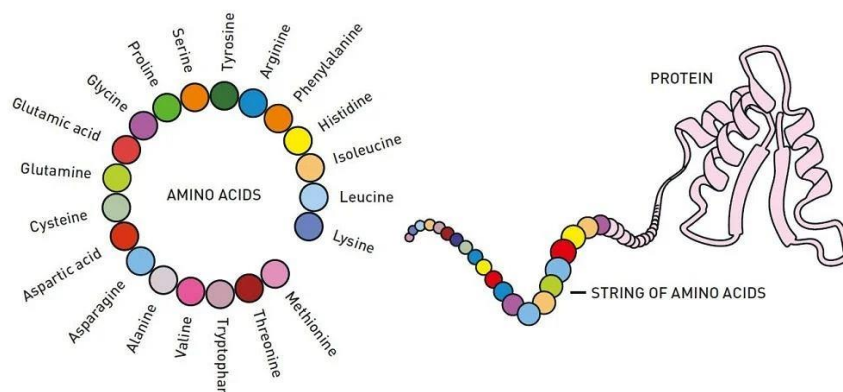
Pokaż mi swój kształt 3D a powiem ci, co robisz w komórce

Badanie ich kształtu 3D (tzw. trzeciorzędowej struktury) było przez lata jednym z najpilniejszych i najtrudniejszych zadań biochemii.

Kształt białka jest bardzo ważny dla jego funkcji.

Żeby zrozumieć, w jaki sposób białko uzyskuje swój ostateczny kształt lub konformację, musimy zrozumieć cztery poziomy struktury białka.

Najprostszy poziom struktury białka, struktura pierwszorzędowa, jest po prostu sekwencją aminokwasów w łańcuchu polipeptydowym - czyli prostym łańcuszkiem. Przepis na budowę takiego łańcuszka (kolejność nanizanych w nim aminokwasów) znajduje się w genach zawartych w DNA.



RYS. Białko może składać się z dziesiątek aminokwasów do kilku tysięcy. Łańcuch aminokwasów składa się w trójwymiarową strukturę, która jest decydująca dla funkcji białka. Johan Jarnestad/Królewska Szwedzka Akademia Nauk

Następny poziom struktury białka (struktura drugorzędowa), powstaje od zagięć w strukturze, które tworzą się w polipeptydzie z powodu interakcji pomiędzy atomami w szkieletcie. Najbardziej popularnymi rodzajami struktur drugorzędowych są helisa α oraz harmonijka β . Obydwie są stabilizowane przez wiązania wodorowe. Nasz łańcuszek nadal jest dość prosty, ale skręcony lub pozaginany na całej swojej długości.

Ogólna trójwymiarowa struktura polipeptydu jest nazywana jego strukturą trzeciorzędową. Wynika ona przede wszystkim z interakcji pomiędzy grupami aminokwasów tworzących białko. Nasz łańcuszek już nie jest prosty, ale zaczyna przypominać węzeł.

Wreszcie, istnieje specjalny rodzaj wiązania kowalencyjnego, który może wpływać na strukturę trzeciorzędową: wiązania disiarczkowe. Są one o wiele silniejsze od pozostałych rodzajów wiązań tworzących strukturę trzeciorzędową.

Działają one jak molekularne "agrafki", utrzymując części polipeptydu mocno połączone ze sobą.

Białko przyjmuje swój ostateczny i unikalny kształt. To on może powiedzieć uczonym, m.in. co dane białko w komórce robi.

Białka lubią się z Nagrodą Nobla

Droga do dzisiejszej Nagrody Nobla rozpoczęła się w połowie XX wieku, gdy naukowcy z Cambridge, John Kendrew i Max Perutz, dokonali przełomowego odkrycia. Z powodzeniem zastosowali metodę zwaną krystalografią rentgenowską, aby przedstawić pierwsze trójwymiarowe modele białek.

Przyznano im Nagrodę Nobla w dziedzinie chemii w 1962 r.

Prace nad opisywaniem struktury białek popchnął następnie Christian Anfinsen, amerykański naukowiec, który stosując różne sztuczki chemiczne rozłożył, a następnie ponownie złożył istniejące białko.

Niezwykłe było to, że białko za każdym razem przybierało dokładnie taki sam kształt.

W 1961 r. uczony doszedł do wniosku, że trójwymiarowa struktura białka jest w całości regulowana przez sekwencję aminokwasów. Doprowadziło to do przyznania mu Nagrody Nobla w dziedzinie chemii w 1972 r.

Jednak logika Anfinsena zawiera paradoks, na który zwrócił uwagę inny Amerykanin, Cyrus Levinthal, w 1969 r. Obliczył on, że nawet jeśli białko składa się tylko ze 100 aminokwasów, w teorii może przyjąć co najmniej 1047 różnych trójwymiarowych struktur.

Gdyby łańcuch aminokwasów składał się losowo, znalezienie prawidłowej struktury białka zajęłoby więcej czasu niż wiek wszechświata. W komórce zajmuje to zaledwie kilka milisekund. Jak więc właściwie składa się ciąg aminokwasów?

Odkrycie Anfinsena i paradoks Levinthala sugerowały, że fałdowanie jest z góry ustalonym procesem. I - co ważne - wszystkie informacje o tym, jak białko się fałduje, muszą być zapisane w sekwencji aminokwasów.

Problem przewidywania

Aby zachęcić do szybszego rozwoju w dziedzinie przewidywania struktury białek, w 1994 r. naukowcy rozpoczęli projekt o nazwie Critical Assessment of Protein Structure Prediction (CASP), który ewoluował potem w konkurs.

Co dwa lata uczeni z całego świata otrzymywali dostęp do sekwencji aminokwasów w białkach, których struktury zostały właśnie określone, ale były one utrzymywane w tajemnicy przed uczestnikami.

Wyzwaniem było przewidzenie struktur białek na podstawie znanych sekwencji aminokwasów.

Okazało się to niezwykle trudne. Zgodność między przewidywaniami, które naukowcy podali w konkursie, a rzeczywistymi strukturami była niska. Przełom nastąpił dopiero w 2018 r., kiedy wkroczył ekspert w dziedzinie neuronauki i pionier sztucznej inteligencji - Demis Hassabis.

W 2010 r. współzакładał firmę DeepMind, która opracowała mistrzowskie modele sztucznej inteligencji (AI) dla popularnych gier planszowych. Google kupiło tę firmę w 2014 r., a dwa lata później DeepMind zyskało globalną uwagę, gdy firma osiągnęła to, co wielu uważało wówczas za święty Graal AI: pokonała mistrza jednej z najstarszych gier planszowych na świecie, Go.

Jednak dla Hassabisa mistrzostwo w Go nie było celem, lecz środkiem do opracowywania lepszych modeli AI. Po tym zwycięstwie jego zespół był gotowy zająć się problemami o większym znaczeniu dla ludzkości, więc w 2018 r. zarejestrował się na trzynasty konkurs CASP.

W poprzednich latach struktury białek, które badacze przewidywali dla CASP, osiągały dokładność co najwyżej 40 proc. Dzięki swojemu modelowi AI, AlphaFold, zespół Hassabisa osiągnął prawie 60 procent.

Wygrali, ale aby odnieść sukces, przewidywanie musiało mieć dokładność 90 proc. w porównaniu do docelowej struktury.

Jednak mimo starań i dalszego rozwoju AlphaFold - nie było postępu, prace utknęły w martwym punkcie. Zespół był zmęczony i zniechęcony.

Jumper wskoczył do gry

John M. Jumper miał już duże doświadczenie, kiedy dołączył do zespołu Deep Mind. I miał kreatywne pomysły na ulepszenie AlphaFold, które fundamentalnie zreformowały model AI.

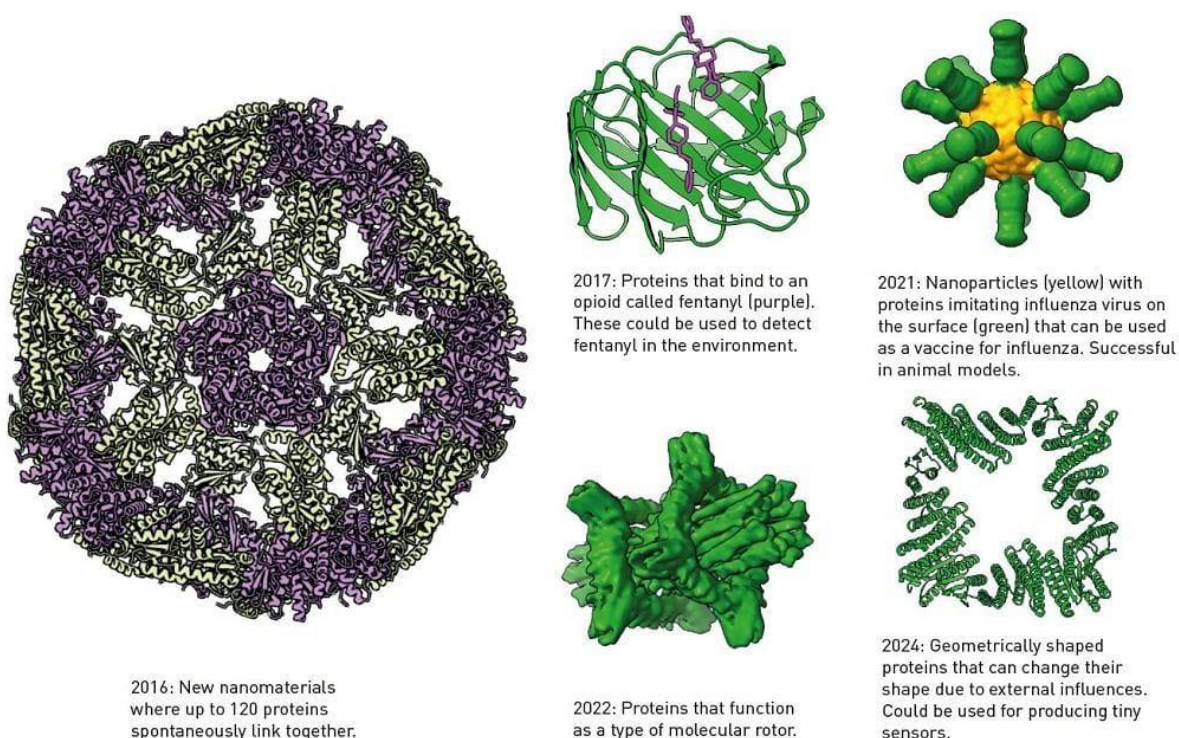
Nowa wersja - AlphaFold2 - została ubarwiona wiedzą Jumpera na temat białek. Zespół zaczął również wykorzystywać innowację stojącą za ostatnim ogromnym przełomem w dziedzinie AI: sieciami neuronowymi zwanymi transformatorami.

Potrąfią one znajdować wzorce w ogromnych ilościach danych w sposób bardziej elastyczny niż wcześniej i skutecznie określać, na czym należy się skupić, aby osiągnąć konkretny cel.

Zespół wyszkolił AlphaFold2 na ogromnych informacjach w bazach danych wszystkich znanych struktur białek i sekwencji aminokwasów, a nowa architektura sztucznej inteligencji zaczęła przynosić dobre wyniki na czas czternastego konkursu CASP.

Tak jak kamień z Rosetty

David Baker też za swój naukowy cel obrał sobie białka. Kiedy w 1993 r. rozpoczął pracę jako lider grupy na University of Washington w Seattle, wykorzystując sprytnie eksperymenty, zaczął badać, w jaki sposób składają się białka. Na podstawie tych doświadczeń zaczął rozwijać oprogramowanie komputerowe, które mogło przewidywać struktury białek: Rosettę.



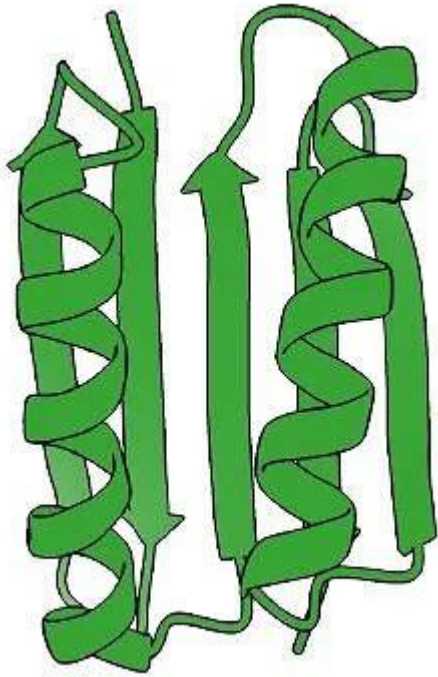
Białka stworzone przez program Rosetta Terezia Kovalova/The Royal Swedish Academy of Sciences

Oczywiście, jego także pociągał konkurs CASP. Zadebiutował w nim już w 1998 r., używając właśnie Rosetty. W porównaniu z innymi uczestnikami, wypadł naprawdę dobrze.

Ale Baker poszedł inną drogą - pod prąd. Jego zespół postanowił używać oprogramowania odwrotnie. Zamiast wprowadzać sekwencje aminokwasów w Rosettie i odczytywać struktury białek, uczeni wprowadzali pożądaną strukturę białka i otrzymywali sugestie dotyczące jego sekwencji aminokwasów.

Takie podejście pozwoliłoby tworzyć zupełnie nowe białka.

Zespół narysował białko o zupełnie nowej strukturze, a następnie zlecił programowi Rosetta obliczenie, jaki typ sekwencji aminokwasów może skutkować pożądanym białkiem. Czy jednak wynik pracy programu - mówiąc kolokwialnie - dobrze zagrał?



Białko TOP7 Terezia Kovalova/The Royal Swedish Academy of Sciences

Aby to zbadać, uczeni stworzyli gen dla proponowanej sekwencji aminokwasów i umieścili go w bakteriach, które produkowały pożądane białko. Następnie określili strukturę białka za pomocą krystalografii rentgenowskiej.

Okazało się, że Rosetta naprawdę może konstruować białka! To, które opracowali naukowcy, o nazwie Top7, miało niemal dokładnie taką strukturę, jaką zaprojektowali.

Top7 był niczym grom z jasnego nieba dla badaczy pracujących nad projektowaniem białek.

Jego unikalna struktura nie istniała w naturze. Ponadto, z 93 aminokwasami, białko to było większe niż cokolwiek wcześniej wyprodukowanego przy użyciu projektowania de novo.

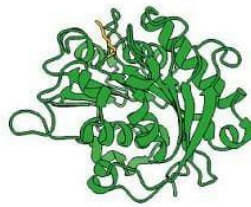
Po konkursie CASP 2020, kiedy David Baker dostrzegł potencjał modeli AI opartych na transformatorach, dodał jeden z nich do Rosetty, co ułatwiło również projektowanie białek de novo. W ostatnich latach w laboratorium Bakera pojawiało się jedno nowe białko za drugim.

To nie koniec, program się rozwija

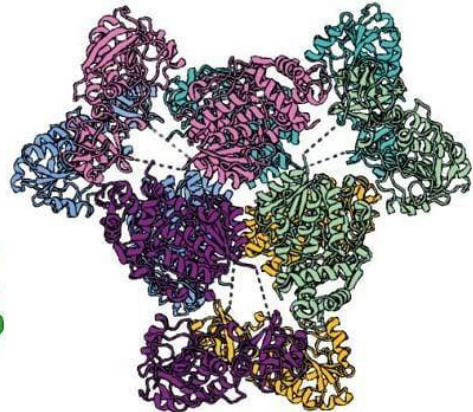
Na początku maja tego roku DeepMind i jej spółka Isomorphic Labs (ma komercjalizować umiejętności AlphaFold), przedstawiły w "Nature" najnowszą wersję programu, AlphaFold 3, która potęguje jego możliwości. Już nie tylko rozszyfrowuje strukturę białek, ale także wszelkich innych biomolekuł, które odgrywają ważną rolę w komórkach, a więc np. także kwasów nukleinowych DNA i RNA.



2022: Part of a huge molecular structure in the human body. More than a thousand proteins form a pore through the membrane surrounding the cell nucleus.



2022: Natural enzymes that can decompose plastic. The aim is to design proteins that can be used to recycle plastic.



2023: A bacterial enzyme that causes antibiotic resistance. The structure is important for discovering ways of preventing antibiotic resistance.

Struktury białek rozszyfrowane przez AlphaFold Terezia Kovalova/The Royal Swedish Academy of Sciences

Sztuczna inteligencja Deep Mind jest obecnie w stanie przewidzieć strukturę wielocząsteczkowych kompleksów, jakie białka tworzą z innymi obecnymi w komórkach molekułami.

Dzięki temu naukowcy mogą mieć wgląd we wszystkie złożone interakcje, które mają znaczenie w biologii czy medycynie.

Po co to wszystko?

Pozwala nam to lepiej zrozumieć, jak funkcjonuje życie, w tym dlaczego rozwijają się niektóre choroby, jak powstaje oporność na antybiotyki lub dlaczego niektóre mikroby mogą rozkładać plastik.

Zaś możliwość tworzenia nowych białek może prowadzić m.in do stworzenia unikalnych nanomateriałów, ukierunkowanych leków, szybszego rozwoju szczepionek i bardziej ekologicznego przemysłu chemicznego.

Red. Aleksander Gurgul