De har avslöjat proteinernas hemligheter med hjälp av datorer och artificiell intelligens

Att till fullo förstå och behärska livets kemiska verktyg – proteiner – har länge varit kemisters dröm. Nu är denna dröm inom räckhåll. ***Demis Hassabis*** *och* ***John Jumper*** *har med hjälp av artificiell intelligens lyckats förutspå formen hos nästan alla kända proteiner.* ***David Baker*** *har lärt sig behärska livets byggsats och skapat proteiner som aldrig tidigare har existerat. Potentialen i deras upptäckter är enorm.*

Hur är livets sprudlande kemi ens möjlig? Svaret på den frågan är att det finns proteiner. De kan beskrivas som en sorts geniala kemiska konstruktioner. De byggs av 20 olika typer av aminosyror som kan kombineras i det oändliga. Med den genetiska koden som mall, länkas de samman i våra celler till långa pärlband av aminosyror.

Sedan händer det som ger proteiner deras storhet: pärlbandet vrider och veckar sig till en särpräglad – ibland helt unik – tredimensionell struktur (figur 1). Genom denna struktur får proteiner sin funktion. En del stöps till kemiska byggstenar som kan bilda muskler, horn eller fjädrar, andra blir till hormoner eller antikroppar. Ett stort antal bildar olika enzymer, som med en makalös precision driver livets kemiska reaktioner. Viktiga är också alla proteiner som sitter på cellernas ytor och fungerar som kommunikationskanaler mellan celler och deras omgivning.

Det går knappt att överskatta den potential som ryms i livets kemiska byggsats, de 20 aminosyrorna. 2024 års Nobelpris i kemi handlar om att förstå och behärska denna byggsats på en helt ny nivå. Ena halvan av priset går till Demis Hassabis och John Jumper. Med hjälp av artificiell intelligens har de lyckats lösa ett problem som kemister har brottats med i över 50 år: att utifrån en sekvens av aminosyror kunna förutsäga ett proteins tredimensionella struktur. Genom detta har de kunnat förutspå formen för nästan alla 200 miljoner kända proteiner. Den andra halvan av priset går till David Baker. Han har

Nobelpriset® är av Nobelstiftelsen registrerat varumärke.

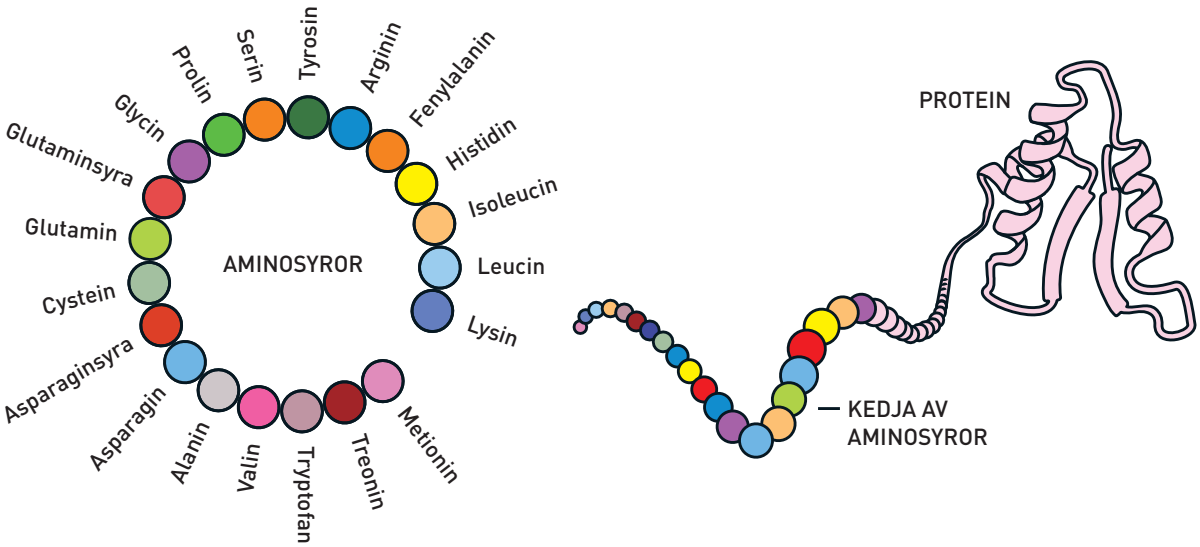
utvecklat datorbaserade metoder för att kunna uppnå det många trodde var omöjligt: att skapa proteiner som aldrig tidigare har funnits och som i många fall har helt nya funktioner.

2024 års Nobelpris i kemi belönar alltså två olika upptäckter, men som du kommer att märka är de starkt sammanlänkade. För att förstå vilka utmaningar som årets Nobelpristagare har övervunnit, behöver vi förflytta oss till gryningen av biokemins storhetstid.

De första gryniga bilderna av proteiner

Att proteiner är viktiga för livets processer har kemister vetat sedan 1800-talet, men först på 1950-talet blev de kemiska verktygen tillräckligt vassa för att forskare skulle kunna börja kartlägga proteiner mer i detalj. En banbrytande upptäckt stod Cambridgeforskarna John Kendrew och Max Perutz för. Med hjälp av metoden röntgenkristallografi lyckades de mot slutet av 1950-talet presentera de första tredimensionella modellerna av hur proteiner ser ut. För det belönades de med Nobelpriset i kemi 1962.

Efter detta har forskare med hjälp av främst röntgenkristallografi – och ofta stor möda – lyckats framställa bilder av cirka 200 000 olika proteiner, vilket har banat vägen för 2024 års Nobelpris i kemi.



Figur 1. Proteiner byggs av 20 olika typer av aminosyror. De kan kombineras i det oändliga. Vissa proteiner består av några tiotals aminosyror, andra tusentals. Kedjan av aminosyror veckar ihop sig till en tredimensionell struktur, som är avgörande för proteinets funktion.

En gåta: hur hittar ett protein sin unika form?

En annan tidig upptäckt stod amerikanen Christian Anfinsen för. Med hjälp av olika kemiska knep lyckades han få ett existerande protein att veckla ut sig och sedan veckla ihop sig igen. Det intressanta var att proteinet varje gång antog exakt samma form. 1961 kunde han därför dra slutsatsen att ett proteins tredimensionella struktur helt och hållet styrs av sekvensen av aminosyror i proteinet. Upptäckten belönades med Nobelpriset i kemi 1972.

I Anfinsens logik rymdes dock en paradox vilket en annan amerikan, Cyrus Levinthal, påpekade 1969. Han räknade ut att även om ett protein bara består av 100 aminosyror, så kan proteinet rent teoretiskt anta minst 10^{47} olika tredimensionella strukturer. Om aminosyrakedjan skulle vecka sig slumpmässigt, skulle det ta längre tid än universums ålder att hitta proteinets rätta struktur. I en cell tar det dock bara någon millisekund. Så hur fungerar aminosyrakedjans veckning?

Den oundvikliga slutsatsen av Anfinsens upptäckt och Levinthals paradox var att veckningen är en styrd process. Dessutom – och detta är viktigt – all information om hur veckningen ska gå till måste finnas inbakad i sekvensen av aminosyror.

Handsken kastas i biokemins stora utmaning

Insikterna ovan ledde till ytterligare en avgörande insikt: om kemister kände till aminosyrasekvensen hos ett protein, borde de kunna förutspå proteinets tredimensionella struktur. Det var en kittlande tanke. Om de lyckades med detta skulle de slippa hålla på med krånglig röntgenkristallografi och spara ofantligt med tid. Dessutom skulle de kunna få fram strukturer för alla proteiner där röntgenkristallografi inte gick att använda.

Genom dessa logiska slutsatser kastades handsken till det som har kommit att bli biokemins stora utmaning: prediktionsproblemet. För att sporra en snabbare utveckling inom området, drog forskare 1994 i gång projektet *Critical Assessment of Protein Structure Prediction* (CASP), som har utvecklats till en sorts tävling. Vartannat år har forskare från hela världen fått tillgång till aminosyrasekvenserna för olika proteiner vars strukturer just har kartlagts. Strukturerna har dock hemlighållits för deltagarna. Utmaningen har varit att utifrån de kända aminosyrasekvenserna förutsäga proteinernas strukturer.

CASP har engagerat många forskare, men att lösa prediktionsproblemet har visat sig vara oerhört svårt. Likheten mellan de förutsägelser som forskarna har tagit fram i tävlingen och de faktiska strukturerna har knappt förbättrats alls. Genombrottet kom först 2018 när en schackmästare, expert på neurovetenskap och pionjär inom artificiell intelligens gav sig in i matchen.

Mästaren på brädspel anmäler sig till protein–OS

Låt oss snabbspola Demis Hassabis bakgrund: han började spela schack som fyraåring och nådde mästarnivå vid 13 års ålder. Som tonåring påbörjade han en karriär som programmerare och framgångsrik spelutvecklare. Han började utforska artificiell intelligens och gav sig in i neurovetenskapen, där han gjorde flera banbrytande upptäcker. De insikter han fick om hjärnan använde han för att utveckla mer välfungerande neuronnätverk för AI. År 2010 bidrog han till att grunda företaget DeepMind, som utvecklade mästerliga AI-modeller för populära brädspel. Företaget såldes 2014 till Google. Två år senare fick DeepMind global uppmärksamhet när företaget uppnådde det som många då ansåg vara den heliga gralen för AI: att slå mästaren i ett av världens äldsta brädspel, Go.

För Demis Hassabis var dock inte Go målet, utan ett medel för att utveckla bättre AI-modeller. Efter segern var hans team redo att ta sig an viktigare problem för mänskligheten. År 2018 anmälde han sig därför till den trettonde CASP-tävlingen.

En oväntad vinst för Demis Hassabis AI-modell

Tidigare år hade de proteinstrukturer som forskare tagit fram inom ramen för CASP som bäst nått 40 procents träffsäkerhet. Med sin AI-modell AlphaFold nådde Demis Hassabis team upp till nästan 60 procent. De vann och det fina resultatet överrumplade många. Det var ett oväntat steg framåt, men lösningen var fortfarande inte bra nog. För det krävdes att förutsägelsen stämde till 90 procent med förlagan.

Efter framgången fortsatte Demis Hassabis och hans team att utveckla AlphaFold. Men hur de än försökte nådde algoritmen aldrig hela vägen fram. Den tuffa sanningen var att de hade hamnat i en återvändsgränd. Teamet var slutkört, men en relativt nyanställd person hade avgörande idéer om hur AI-modellen kunde förbättras: John Jumper.

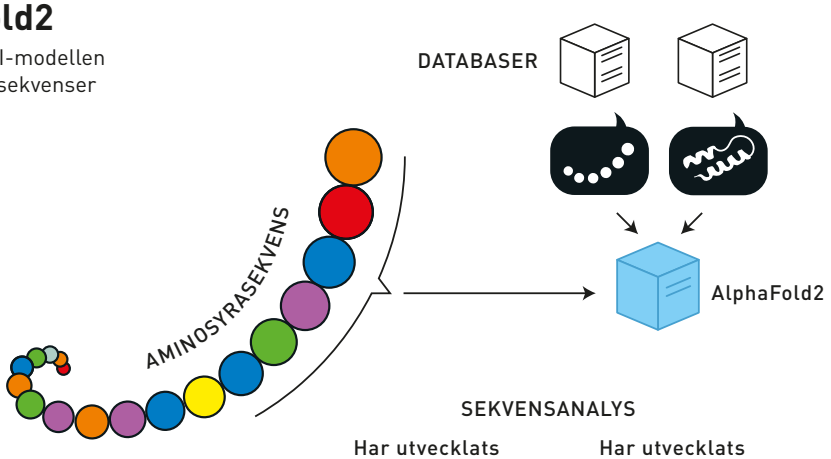
Figur 2.

Så fungerar AlphaFold2

I utvecklingen av AlphaFold2 har AI-modellen fått träna på alla kända aminosyrasekvenser och kartlagda proteinstrukturer.

1. INMATNING OCH DATABASSÖKNING

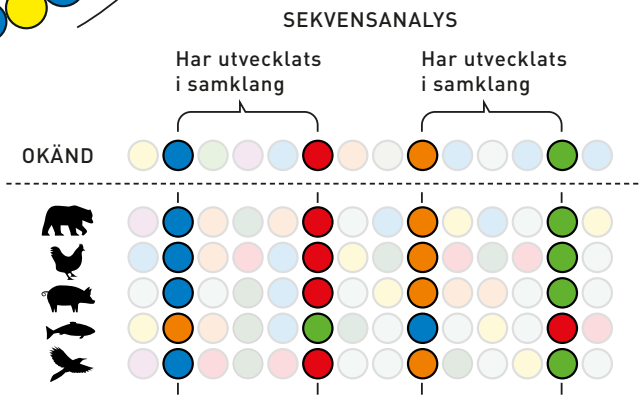
En aminosyrasekvens med okänd struktur matas in i AlphaFold2, som söker efter liknande aminosyrasekvenser och proteinstrukturer i olika databaser.

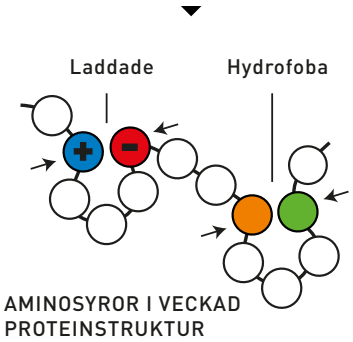


2. SEKVENSANALYS

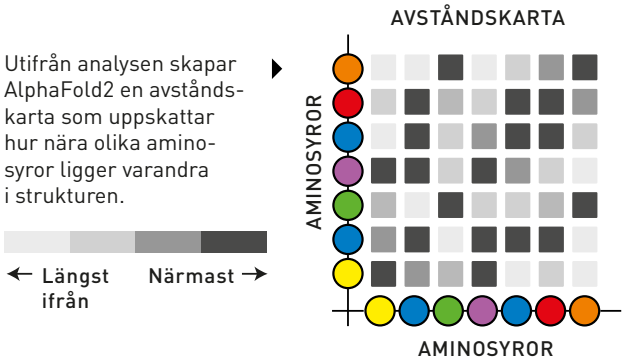
AI-modellen linjerar upp alla liknande aminosyrasekvenser – som ofta är från olika arter – och undersöker vilka delar som har bevarats under evolutionen.

I nästa steg utforskar AlphaFold2 vilka aminosyror som skulle kunna ha kontakt i den tredimensionella proteinstrukturen. Aminosyror som interagerar förändras i samklang med varandra under evolutionen. Är den ena laddad har den andra motsatt laddning, så att de dras till varandra. Byts en ut till en vattenavvisande (hydrofob) aminosyra, blir den andra också hydrofob.





Utifrån analysen skapar AlphaFold2 en avståndskarta som uppskattar hur nära olika aminosyror ligger varandra i strukturen.

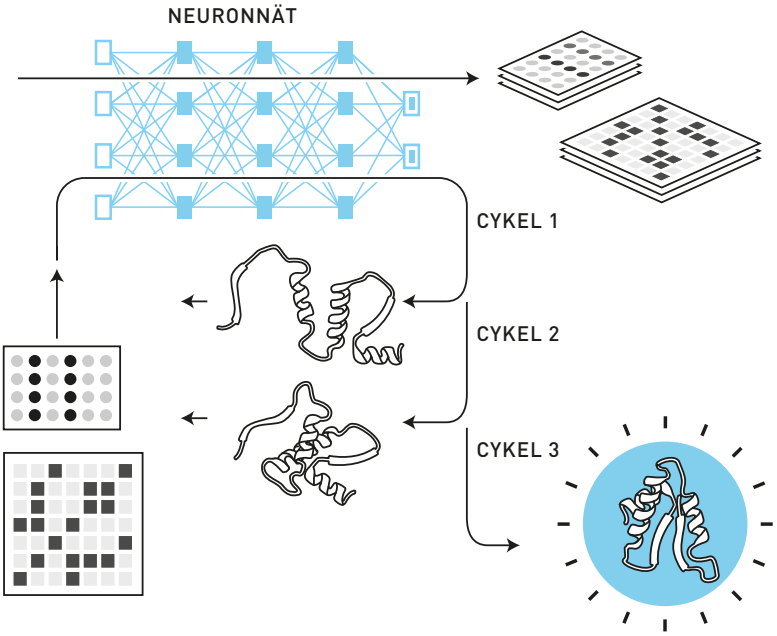


3. AI-ANALYS

I en sökande process förfinar AlphaFold2 sekvensanalysen och avståndskartan. AI-modellen utnyttjar neuronnät kallade transformers. De identifierar vad i all data som är viktigt att fokusera på. Information från andra proteinstrukturer – om sådan hittats i steg 1 – utnyttjas också.

4. HYPOTETISK STRUKTUR

AlphaFold2 lägger ett pussel med alla aminosyrorasekvenser och testar sig fram till en hypotetisk struktur. Denna körs in i steg 3 igen. Efter tre cykler har AlphaFold2 letat sig fram till en bestämd struktur. AI-modellen uppskattar hur stor sannolikheten är att olika delar av strukturen stämmer med verkligheten.



John Jumper plockar upp handsken till biokemins stora utmaning

Det var egentligen universums tjuskraft som fick John Jumper att plugga fysik och matematik. När han tog anställning 2008 vid ett företag som med hjälp av superdatorer simulerade proteiner och deras rörelser, insåg han dock att kunskaper i fysik kan bidra till att lösa medicinska problem.

Det nyvunna intresset för proteiner följde med när John Jumper 2011 påbörjade en forskarutbildning i teoretisk fysik. För att spara datorkraft – något det var ont om på universitetet – började han hitta enklare och smartare metoder för att simulera proteiners rörelser. Ganska snart plockade John Jumper också upp handsken till biokemins stora utmaning. När han som nydisputerad 2017 hörde rykten om att Google DeepMind i största hemlighet hade börjat spå i proteinstrukturer, sökte han sig dit. Med sin erfarenhet av att simulera proteiner hade han kreativa idéer om hur AlphaFold kunde förbättras. Så efter att teamet hade kört fast, ledde han arbetet tillsammans med Hassabis då AI-modellen reformerades i grunden.

Häpnadsväckande resultat med reformerad AI-modell

Den nya versionen – AlphaFold2 – färgades av John Jumpers kunskaper om proteiner. Dessutom började teamet utnyttja innovationen bakom de senaste årens enorma genombrott inom AI: neuronnätverk som kallas *transformers*. Dessa kan på ett mer flexibelt vis än tidigare hitta sammanhang i enorma mängder data och effektivt avgöra vad som är viktigt att fokusera på för att nå ett uppsatt mål.

Teamet lät bland annat AlphaFold2 träna på den enorma information som finns i databaserna för alla kända proteinstrukturer och aminosyrasekvenser (figur 2). Lagom till den fjortonde CASP-tävlingen började den nya AI–arkitekturen att leverera fina resultat.

När ledningen för CASP år 2020 utvärderade resultaten insåg de att biokemins 50-åriga utmaning var över. AlphaFold2 presterade i de flesta fall nästan lika bra som röntgenkristallografi, vilket var häpnadsväckande. När en av grundarna av CASP, John Moult, den 4 december 2020 avrundade tävlingen ställde han frågan: Vad händer nu?

Vi återkommer till det. Nu ska vi bege oss bakåt i tiden och rikta strålkastaren mot en annan deltagare i CASP. Det är dags att presentera den andra halvan av 2024 års Nobelpris i kemi, som handlar om konststycket att skapa nya proteiner från grunden.

En lärobok om cellen får David Baker att byta bana

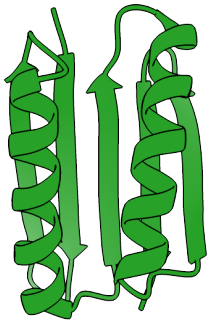
När David Baker började studera vid Harvarduniversitetet valde han filosofi och samhällsvetenskap. Under en kurs i utvecklingsbiologi stötte han dock på den första upplagan av den numera klassiska läroboken *Molecular Biology of the Cell*. Det fick honom att ändra bana i livet. Han började utforska cellens biologi och kom så småningom att fascineras av proteinernas struktur. När han 1993 blev gruppledare vid University of Washington i Seattle, tog han sig an biokemins stora utmaning. Med smart utformade experiment började han kartlägga vad som styr proteiners veckning. Det gav insikter som han tog med sig när han mot slutet av 1990-talet började utveckla ett datorprogram för att kunna förutspå proteinstrukturer: Rosetta.

Med detta program debuterade David Baker 1998 i CASP och i jämförelse med övriga deltagare presterade Rosetta riktigt bra. Framgången ledde till en ny idé: att David Bakers team skulle kunna använda programmet omvänt. I stället för att mata in aminosyrasekvenser i Rosetta och få ut proteinstrukturer, borde de kunna mata in önskade strukturer på proteiner och få ut förslag på vilka aminosyrasekvenser proteinerna bör ha. På så vis skulle de kunna skapa helt nya proteiner.

David Baker blir proteinkonstruktör

Området proteindesign – där forskare skräddarsyr nya proteiner med nya funktioner – började vid den här tidpunkten att ta fart. Många utgick från redan existerande proteiner och försökte skruva om dem, så att de till exempel kunde bryta ner miljöfarliga ämnen eller fungera som verktyg i den kemiska tillverkningsindustrin.

Det naturliga urvalet av proteiner är dock begränsat. För att öka möjligheterna att få fram proteiner med helt nya funktioner, ville Bakers forskargrupp i stället skapa dem från grunden. Som David Baker har uttryckt det: ”Om du vill bygga ett flygplan, börjar du inte med att förändra en fågel; i stället förstår du aerodynamikens grundläggande principer och bygger en flygande maskin baserat på dem.”



Figur 3. Top7 – det första proteinet som var helt olikt alla kända existerande proteiner.

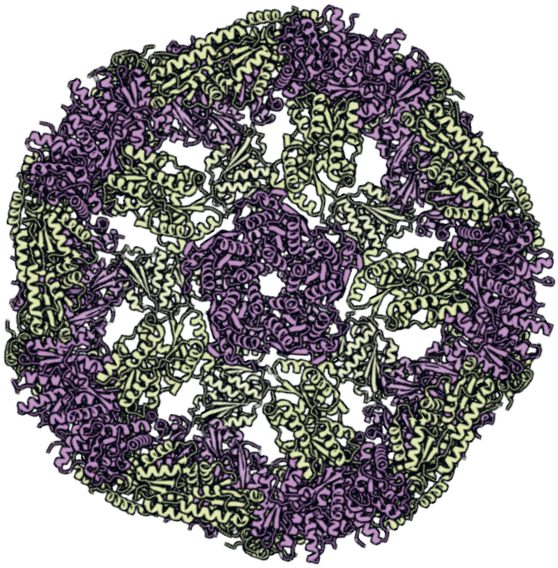
Ett unikt protein ser dagens ljus

Området där man från grunden bygger helt nya proteiner kallas *de novo*-design. Forskargruppen ritade upp ett protein med en aldrig tidigare skådad struktur, sedan lät de Rosetta ta reda på vilken aminosyrasekvens som skulle kunna ge det önskade proteinet. För att lyckas med detta anropade Rosetta en databas med alla kända proteinstrukturer. Bland dessa letade programmet fram korta fragment av proteiner som hade likheter med den önskade strukturen. Med hjälp av grundläggande kunskaper om proteiners energilandskap, optimerade sedan Rosetta dessa fragment och föreslog en aminosyrasekvens.

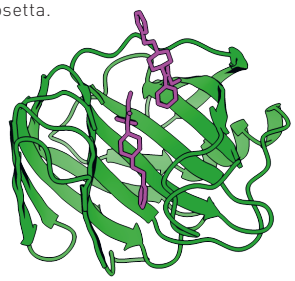
För att undersöka hur väl programmet lyckades, satte David Bakers forskargrupp in genen för den föreslagna aminosyrasekvensen i bakterier som framställde det önskade proteinet. Sedan kartlade de proteinets struktur med hjälp av röntgenkristallografi.

Det visade sig att Rosetta verkligen kunde konstruera proteiner. Proteinet som forskarna fick fram, *Top7*, hade i princip exakt den struktur som de hade önskat sig (figur 3).

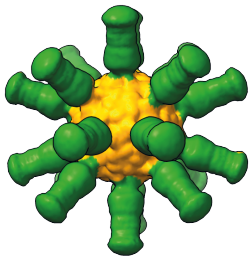
Figur 4. Proteiner framtagna med hjälp av Bakers program Rosetta.



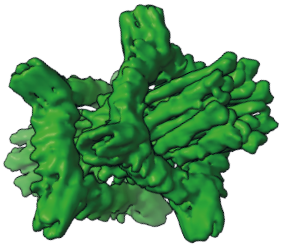
2016: Nya nanomaterial där upp till 120 proteiner spontant kopplar till varandra.



2017: Protein som binder till opioiden fentanyl (lila). Kan användas för detektion av fentanyl i miljön.



2020: Nanopartikel (gul) med protein som härmar influensavirus på ytan (grönt). Har fungerat som vaccin i djurförsök.



2022: Proteiner som fungerar som en minimal rotor.

2024: Proteiner med olika geometriska former som vid yttre påverkan kan ändra form. Skulle kunna användas för att få fram minimala sensorer.

Spektakulära skapelser från David Bakers laboratorium

Top7 slog ner som en bomb bland forskare som höll på med proteindesign. De som tidigare hade skapat proteiner *de novo* hade bara kunnat härma befintliga strukturer. Top7 hade en unik struktur som inte fanns i naturen. Med sina 93 aminosyror var proteinet dessutom större än allt man tidigare hade tagit fram med *de novo*-design.

David Baker publicerade upptäckten år 2003 och det var första steget på något som bara kan beskrivas som en enastående utveckling. Några av alla de spektakulära proteiner som har skapats vid Bakers laboratorium hittar du i figur 4. David Baker släppte också koden för Rosetta fri, så en global sammanslutning av forskare har vidareutvecklat programmet och gett det nya tillämpningsområden.

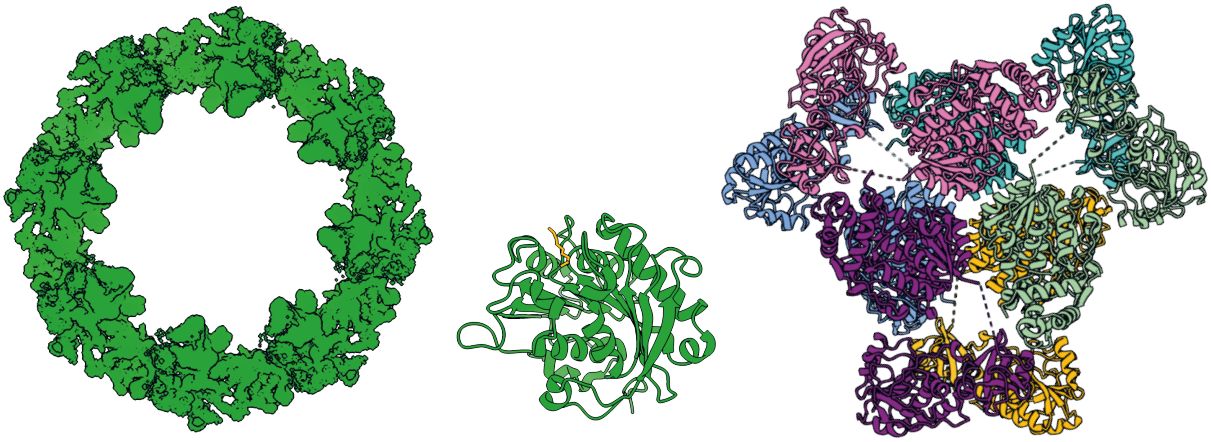
Det dags att knyta ihop säcken för 2024 års Nobelpris i kemi, för vad händer nu?

Arbete som tidigare tog åratal, tar nu några minuter

När Demis Hassabis och John Jumper hade fått bekräftat att AlphaFold2 verkligen fungerar, lät de räkna ut strukturen för alla mänskliga proteiner. Sedan förutspådde de strukturen för i princip alla de 200 miljoner proteiner som forskare hittills har hittat när de kartlagt jordens organismer.

Google DeepMind har också tillgängliggjort koden för AlphaFold2 och vem som helst kan dra nytta av AI-modellen. Den har blivit en guldgruva för forskare. I oktober 2024 hade AlphaFold2 använts av mer än två miljoner personer från 190 länder. Tidigare tog det ofta många år att få fram en proteinstruktur, om det ens gick. Nu går det på några få minuter. AI-modellen är inte felfri, men den gör en uppskattning av hur tillförlitlig struktur den får fram. Så forskare vet om strukturen riskerar att ha brister. I figur 5 ser du några av väldigt många exempel på hur AlphaFold2 hjälper forskare.

Figur 5. Proteinstrukturer framtagna med hjälp av AlphaFold2.



2022: En del av en enorm molekylär struktur i människokroppen. Mer än tusen proteiner formar en por genom membranet som omger cellens kärna.

2022: Naturliga enzymer som kan bryta ner plast. Målet är att designa proteiner som kan användas för att återvinna plast.

2023: Ett bakterieenzym som orsakar antibiotikaresistens. Strukturen är viktig för att hitta vägar att förhindra antibiotikaresistens.

När David Baker efter CASP 2020 insåg vilken potential det finns i transformerbaserade AI-modeller lade han in sådana även i Rosetta, vilket i sin tur underlättar *de novo*-design av proteiner. De senaste åren har den ena mer fantasifulla proteinkreationen efter den andra flödat ut från Bakers laboratorium (figur 4).

En svindlande utveckling till nytta för mänskligheten

Livets enorma mångfald vittnar om vilka enastående kemiska verktyg proteiner är. Det är svindlande att vi nu så lätt kan se hur dessa små molekylära maskiner ser ut. Därmed kan vi bättre förstå hur livet fungerar, exempelvis varför vissa sjukdomar utvecklas, hur antibiotikaresistens uppstår eller vad som gör att vissa mikrober kan bryta ner plast.

Lika stort är det att vi har förmågan att skapa proteiner som är laddade med nya funktioner. Det kan ge nya nanomaterial, träffsäkra läkemedel, snabbare vaccinutveckling, minimala sensorer och en miljövänligare kemisk industri. För att nämna några saker som gör mänskligheten den största nytta.

LÄS MER

Mer information om årets priser, bland annat en vetenskaplig bakgrundsartikel på engelska, finns på Kungl. Vetenskapsakademiens webbplats, *www.kva.se*, och på *www.nobelprize.org*. Där kan man också titta på presskonferenser, Nobelföreläsningar och annat videomaterial. Mer information om utställningar och aktiviteter kring Nobelprisen och Ekonomipriset finns på *www.nobelprizemuseum.se*.

Kungl. Vetenskapsakademien har beslutat utdela Nobelpriset i kemi 2024 med

ena hälften till

DAVID BAKER

Född 1962 (62 år) i Seattle, WA, USA. Fil.dr 1989 vid University of California, Berkeley, CA, USA. Professor vid University of Washington, Seattle, WA, USA och Investigator, Howard Hughes Medical Institute, USA.

”för datorbaserad proteindesign”

och med andra hälften gemensamt till

DEMIS HASSABIS

Född 1976 (48 år) i London, Storbritannien. Fil.dr 2009 vid University College London, Storbritannien. VD för Google DeepMind, London, Storbritannien.

JOHN JUMPER

Född 1985 (39 år) i Little Rock, AR, USA. Fil.dr 2017 vid University of Chicago, IL, USA. Senior Research Scientist på Google DeepMind, London, Storbritannien.

”för proteinstrukturprediktion”

Vetenskapsredaktörer: Peter Brzezinski, Johan Åqvist och Heiner Linke, Nobelkommittén för kemi
Text: Ann Fernholm
Illustrationer: Johan Jarnestad och Terezia Kovalova
Redaktör: Vincent von Sydow
©Kungl. Vetenskapsakademien