

Kan algoritmer finde bedre lægemidler? og fjerne den plet fra din trøje, du ikke kan få væk?

Mads Herbert Kern

January 2026

I mit ph.d.-projekt har jeg udviklet en algoritme, der kan forudsige, hvordan ændringer i et protein påvirker dets funktion. Ved at kombinere kunstig intelligens med eksperimentel data gør metoden det muligt hurtigere og mere præcist at designe proteiner, der virker bedre i praksis – som medicin, som enzymer i vaskemidler eller til nedbrydning af plastik.

Proteiner – et af verdens centrale værktøjer

Proteiner udfører meget af det arbejde, der får levende organismer – og det moderne samfund – til at fungere. De bruges i dag i medicin mod alvorlige sygdomme som diabetes og kræft, og som enzymer i alt fra fødevareproduktion til vaskemidler. Naturen har gennem evolutionen udviklet proteiner, der er specialister i utallige opgaver, hvilket medicinal- og biotekindustrien udnytter til at løse nogle af vores største udfordringer.

Problemet er, at disse proteiner sjældent er effektive nok til netop den opgave, vi ønsker, de skal løse. De virker ofte kun under snævre betingelser, for eksempel ved bestemte temperaturer eller pH-værdier. For at udnytte deres fulde potentiale er vi derfor nødt til måltrettet at ændre dem – hvilket er lidt af en gordisk knude.

Udfordringen er, at der findes næsten uendeligt mange måder at ændre et protein på. Et protein består af op mod flere hundrede byggesten, og hver af disse kan ændres på mange forskellige måder. Selv små ændringer kan have stor betydning for proteinets funktion. At teste alle muligheder i laboratoriet er praktisk talt umuligt, og man er derfor nødt til at vælge meget selektivt, hvilke varianter der skal undersøges.

Kunstig intelligens får os kun noget af vejen

Her kommer kunstig intelligens ind i billedet. I de senere år har store deep learning-modeller lært at “læse” proteiner på samme måde, som sprogmodeller,

som for eksempel ChatGPT, læser tekst. Ved at træne på millioner af naturlige proteiner har disse modeller opbygget en generel forståelse af, hvordan proteiner ser ud og fungerer.

Modellerne har dog en væsentlig begrænsning. De har primært lært af naturens egne proteiner og ved derfor meget om, hvad der findes, men langt mindre om, hvad der sker, når vi bevidst ændrer noget. Når vi introducerer mutationer for at forbedre et protein, bevæger vi os uden for det område, hvor modellerne har mest erfaring. I netop dette regime – hvor data er sparsomt, og ændringerne er målrettede – kommer de store deep learning-modeller ofte til kort.

Løsning: at kombinere generel og specifik viden

I mit ph.d.-projekt har jeg, i samarbejde med andre forskere, udviklet en model, der kombinerer viden fra store neurale netværk, kaldet deep learning, for proteiner med konkrete eksperimentelle data fra et specifikt protein. Modellen kan trænes på et lille datasæt af muteret varianter af et protein og lærer ud fra disse hvordan andre mutationer vil påvirke proteinet.

Modellen udnytter indlejrede, abstrakte repræsentationer af proteiner og sandsynlighedsbaserede forudsigelser fra deep learning-modeller, der er trænet på store mængder general protein data. Disse bruges til at konstruere en specialdesignet *Gaussian process*-model, som kan forudsige effekten af nye mutationer – også selvom kun få varianter er blevet testet eksperimentelt.

Med andre ord kombinerer modellen den brede, generelle viden om naturlige proteiner med få, men vigtige eksperimenter, der giver specifik viden om hvordan mutationer påvirker et givet protein. Det gør det muligt at forudsige effekten af mutationer, som aldrig før er blevet undersøgt i laboratoriet.

Vi har desuden udviklet en procedure til hvordan modellen kan guide forskere til at teste de mest lovende proteinvarianter frem for at teste i blinde. Dermed kan de hurtigere og mere effektivt udvikle optimerede proteiner, hvad enten det er som lægemidler til behandling af sygdomme eller som enzymer i vaskemidler, der fjerner pletter ved lavere temperaturer og sparer både vand og energi.

Læring fra få eksperimenter

I proteinoptimering tager man typisk udgangspunkt i et naturligt protein, en såkaldt vildtype, og konstruerer varianter ved at ændre enkelte byggesten. Selv hvis man kun ændrer en håndfuld byggesten, opstår der et enormt antal mulige varianter. Det er derfor hverken praktisk eller økonomisk muligt at teste dem alle, og man står ofte i en situation, hvor man kun har data fra nogle få hundrede eksperimenter.

Den centrale udfordring er derfor, hvordan man kan opnå pålidelige forudsigelser af, hvilke proteinvarianter der er bedst egnede til en given opgave, når data-grundlaget er stærkt begrænset.

Neurale netværk er velegnede til at modellere komplekse og højdimensionelle data som proteiner, men de kræver typisk meget store datasæt for at fungere godt. Gaussian process-modeller har derimod den fordel, at de kan lære effektive sammenhænge ud fra små datasæt, hvis de er konstrueret korrekt. Vores model, *Kermut*, kombinerer disse to tilgange ved at bruge neurale netværk til at repræsentere proteiner på en informativ måde så en ophygget konstrueret Gaussian process-model kan lære sammenhæng mellem mutationer og funktion fra bare få eksperimenter.

Denne kombination viste sig at være særligt effektiv og *Kermut* overgik eksisterende state-of-the-art metoder til forudsigelse af mutations-effekter. Desuden tager det kun sekunder at lave de nødvendige beregninger, mens den tidligere bedste model brugte dage. I dag er vores model gjort frit tilgængelig, så den kan anvendes af alle forskere og virksomheder inden for medicinal- og biotekindustrien.

Kunstig intelligens til gavn for samfundet

Ved at gøre proteinoptimering hurtigere, billigere og mere målrettet bidrager min forskning til udviklingen af nye biologiske løsninger på nogle af samfundets største udfordringer. Metoden kan anvendes til at optimere et hvilket som helst protein til ethvert anvendelsesområde samt reducerer behovet for omfattende og ressourcekrævende laboratorieforsøg. Således bliver det også lettere at udvikle ny medicin og bioteknologiske løsninger for mindre spillere, som ikke har de samme ressourcer som store virksomheder.

Jeg er i dag ansat som postdoc, hvor jeg arbejder videre med at udvikle algoritmer, der kan modellere endnu mere komplekse sammenhænge i proteiner, herunder hvordan mutationer påvirker hinanden gensidigt. Målet er at give forskere endnu bedre værktøjer til at designe proteiner, der kan helbrede sygdomme, beskytte miljøet – og måske endda gøre det lidt lettere at få den sidste genstridige plet ud af tøjet.