

# Chemie tussen mens en machine

*De nieuwe labrobot Eve roept de vraag op in hoeverre je moleculaire processen kunt automatiseren. 'De vragen die je met een lerend algoritme kunt beantwoorden zijn heel divers.'*

Eind vorig jaar verscheen een publicatie over *robot scientist* Eve, een autonoom laboratoriumsysteem dat experimenten doet, gegevens interpreteert, verbanden zoekt in resultaten en vervolgens nieuwe hypothesen formuleert. Op basis van die stappen start Eve een cyclus van nieuwe experimenten. Al doende vond de robot een stof die mogelijk een startpunt vormt voor een nieuw antimalariamiddel.

Eve is vooral een demonstratieproject van computerwetenschappers van de University of Manchester, dat illustreert hoe integratie van robotisering en lerende algoritmes wetenschappelijke processen kan automatiseren. Soms lijkt het alsof computertechnologie inmiddels tot alles in staat is, maar vooralsnog is dat fictie.

## Minder beperkingen

Mogelijk gaan algoritmes in de toekomst wel helpen om nieuwe moleculen te verzinnen; geavanceerder en creatiever dan wat robot scientist Eve nu laat zien.

Gerard van Westen, computational chemisch bioloog bij het Leiden Academic Centre for Drug Research: 'Het zou kunnen dat het die kant uitgaat. Vaak heeft menselijke creativiteit te maken met nieuwe verbanden leggen, op basis van eerdere ervaringen. Als een computeralgoritme met iets op de proppen komt waaraan mensen nog nooit dachten, dan kun je dat denk ik ook creativiteit noemen.'

Van Westen wil in zijn werk patronen ontdekken in bestaande kennis en laat computers een deel van dat werk uitvoeren. 'Ik bedenken experimenten op dezelfde manier als een biochemicus, alleen voer ik ze uit

op de computer. Daarmee heb je veel minder beperkingen. Terwijl in de echte wereld de kosten van chemicaliën, syntheses en antilichamen de mogelijkheden bepalen, kun je op de computer alles uitvoeren wat je maar verzint.'

Veel van wat Van Westen doet, draait om data en *machine learning*: computeralgoritmes trainen met gegevens, waarna het algoritme in staat is om nieuwe verbanden te zien of voorspellingen te doen. De laatste jaren is machine learning bijvoorbeeld ingezet om genoomkaarten accurater te annoteren op de aanwezigheid van genen. Zo'n lerend algoritme train je eerst om gen-kenmerken te herkennen en laat je vervolgens op eigen houtje een compleet genoom te lijf gaan.

Van Westen: 'De vragen die je met een lerend algoritme kunt beantwoorden zijn heel divers. Is een wetenschappelijke publicatie relevant voor een onderzoek? Gaat er resistentie ontstaan tegen een geneesmiddel of bestrijdingsmiddel? Gaat een medicijn binden aan een eiwit?'

## Trillende spaghetti

Eiwitten zijn het specialisme van Peter Bolhuis, hoogleraar simulaties van biomoleculaire systemen aan de UvA. Volgens

hem is niet alles zo eenvoudig in silico te vangen. 'Ik krijg weleens de vraag waarom het eigenlijk zolang duurt voordat we snel en nauwkeurig simulaties kunnen doen van bijvoorbeeld eiwitvouwing. Dit vak is immers al enkele decennia geleden uitgevonden en sindsdien zijn computers veel krachtiger geworden.'

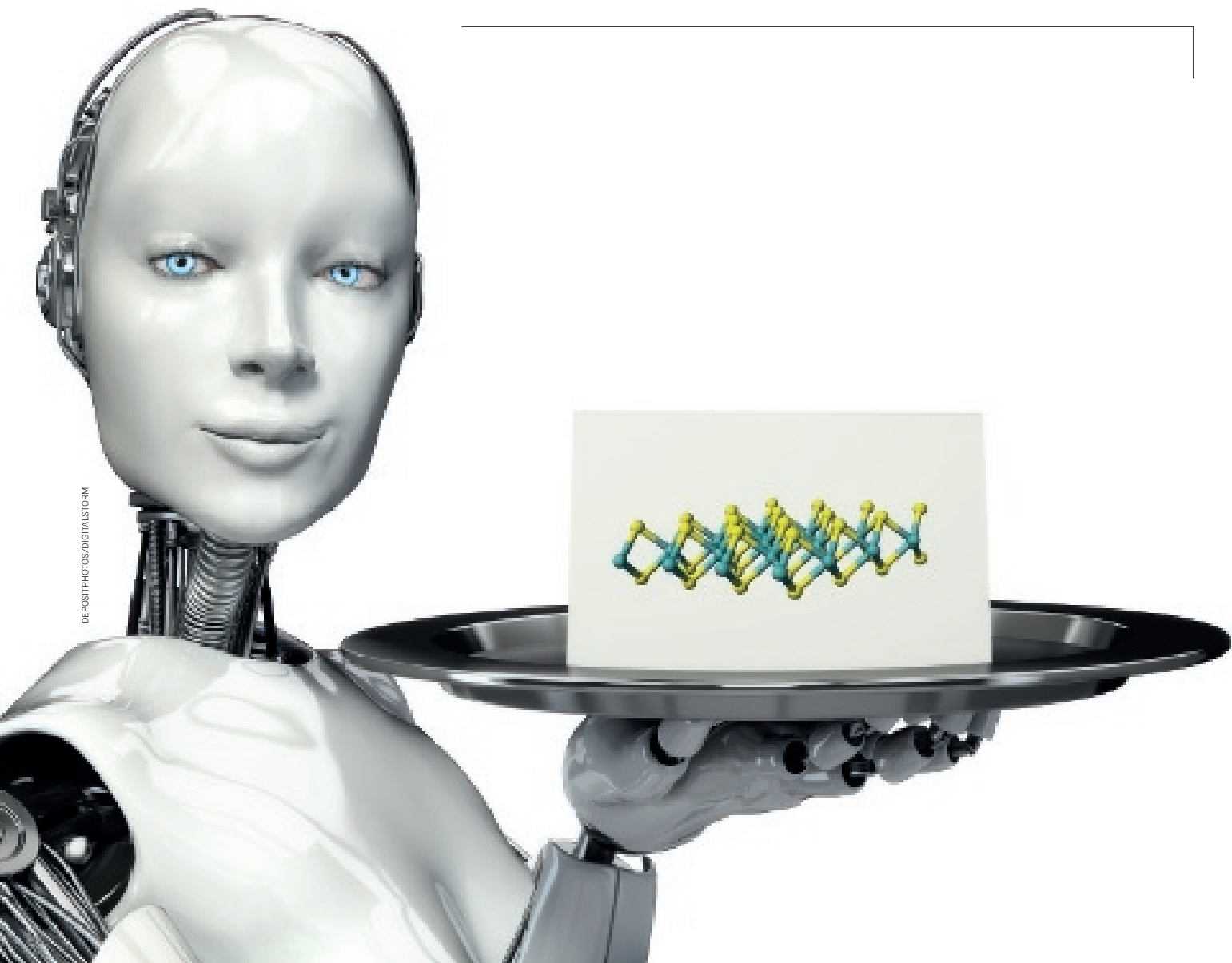
Ondanks de exponentieel toegenomen rekenkracht blijft het ingewikkeld en tijdrovend om de moleculaire dynamiek van een eiwit in een simulatie te vangen. Eiwitten zijn namelijk flexibel en metastabiel, aldus Bolhuis. Ze lijken welbeschouwd meer op een rommelig trillende spaghetti, dan op een starre paperclip. Wie bijvoorbeeld precies wil begrijpen hoe een stofje bindt aan een enzym, moet modelleren hoe in zo'n wiebelende peptidekluven de actieve site doorlopend van vorm verandert.

Met *biomolecular dynamics simulation* kun je zo'n moleculair filmpje maken, dat rekent aan de positie en beweging van elk individueel atoom, plus alle krachten die honderdduizenden atomen onderling op elkaar uitoefenen. De gedetailleerde simulatie van een paar microseconden uit het leven van een eiwit kost daardoor al snel enkele dagen aaneengesloten rekentijd op een zeer krachtige computer.

Bolhuis: 'Eiwitten krijgen hun 3D-vorm door de subtiele balans van vaak tegenstelde krachten en interacties. Daaraan nauwkeurig rekenen is ingewikkelder dan de bewegingen beschrijven van tienduizenden planeten en sterren.'

Volgens Bolhuis kan de wetenschap het precieze mechanisme van eiwitvouwing nog nauwelijks voorspellen. Bij veel ziektes, zoals Alzheimer, speelt ophoping van

***'Brute rekenkracht zal niet de oplossing bieden'***



DEPOSITPHOTOS/DIGITALSTORM

verkeerd gevouwen eiwitten juist een hoofdrol. Bovendien dirigeren chaperonneiwitten veel grote eiwitten in de cel richting de juiste vorm. Zulke complexe interacties probeert Bolhuis ook beter te begrijpen door simulaties.

‘Brute rekenkracht zal uiteindelijk niet de oplossing gaan bieden’, zegt Bolhuis. Zelfs simpelweg opknippen en verdelen van berekeningen over honderdduizenden computerprocessoren is niet snel genoeg.

### **Biologische activiteit**

De crux zit in slimmere onderzoeksmethodes ontwikkelen en efficiënte vereenvoudigingen. In plaats van te proberen de complete eiwitstructuur over lange tijd te simuleren, focussen onderzoekers op een of enkele toestanden. Bolhuis: ‘Je zoekt dan naar energetische barrières, naar conformaties die de stabiliteit van een eiwit of de binding van liganden bepalen. Door die te simuleren, kun je in veel kortere tijd inzichten krijgen.’

In vergelijking met eiwitvouwing is analyse van de relatie tussen een moleculaire structuur en biologische activiteit minder complex. Een *quantitative structure-activity relationship* (QSAR) is een model van het verband tussen structurele eigenschappen van een groep gelijksoortige moleculen en hun biologische effecten in bijvoorbeeld een toxicologische test.

Met QSAR kun je vervolgens extrapoleren: van een verwante stof die nog niet is getest, kun je iets zeggen over de toxicologische eigenschappen of biologische activiteit. ‘In de farmaceutische industrie is van oudsher veel onderzoek aan QSAR gedaan, maar daarbuiten, bijvoorbeeld in de chemische industrie was men afwachtend’, zegt Sabine Van Miert, als onderzoeker en docent werkzaam bij RADIUS-onderzoeksgroep van de Thomas More Hogeschool in Geel. ‘Met de komst van REACH en de wens om dierproeven terug te dringen is dat snel aan het veranderen. De EU zegt: als er helemaal geen data beschikbaar zijn

## **‘Met QSAR kun je zo’n snelle screening doen’**

van een stof, probeer dan eerst die kennis aan te vullen met modellen in plaats van dierproeven of andere experimenten. Met QSAR kun je zo’n snelle screening doen.’ Van Miert werkte samen met haar collega Geert Verheyen aan het Vlaamse project ASOPUS, een gelijknamige webtool die verschillende QSAR-instrumenten met elkaar vergelijkt. Een zekere basiskennis is wel vereist om deze tools te gebruiken, aldus moleculair bioloog Verheyen. Voeling met toxicologie is ook handig. ‘QSAR’s geven een inschatting van activiteit en toxiciteit op de mate waarop jouw molecuul lijkt op de moleculen in een bepaald domein. Daarover is soms discussie mogelijk ►

► en dan kan het nodig zijn om aanvullende informatie uit de literatuur of extra experimenten te verzamelen'

De kracht van QSAR wordt bepaald door de dataset waarop die is gebaseerd.

Verheyen: 'Voor veel eindpunten zijn de datasets niet toereikend om modellen te maken, zoals nanomaterialen. Daarvoor moeten we gewoon nog goede datasets

### ► Onbevangen blik

Het belangrijkste voordeel van het gebruik van machine learning is dat de computer onbevangen naar bestaande kennis kijkt, meent Van Westen. 'Een onderzoeker heeft tijdens zijn promotie en postdoc een bepaalde manier van denken aangeleerd op een specifiek onderzoeksveld. Een computer is onbevangen en ziet daardoor verbanden die misschien niet zo voor de hand liggen.' Dat kan zelfs op kleinere schaal. 'Je kunt bijvoorbeeld kijken naar alle stoffen die de afgelopen vijftien jaar in mijn lab zijn gesynthetiseerd, omdat je zoekt naar een nieuwe verbinding. Dan ben je ook *data-driven* bezig. Want dat is de kern van de zaak, in plaats van dat je alleen op je *gut feeling* vertrouwt, ga je daarnaast een objectieve *data-driven* beslissing nemen.

Gewoon door te kijken naar wat je in het verleden aan kennis hebt verzameld. En omdat er labs zijn die misschien in hetzelfde veld werken, kun je hun publicaties erbij nemen. Dat wordt alsmaar makkelijker, omdat data en publicaties via internet steeds toegankelijker worden.'

Het proces van wetenschap verandert, aldus Van Westen. 'Het nieuwe lab van wiskunde en natuurwetenschappen heeft meer kantoorruimte, minder labruimte en geen papieren bibliotheek. De rol van de computer wordt belangrijker, wetenschap wordt meer schermwerk. Waar het precies heengaat weet ik niet. Als we sommige arbeidsintensieve taken overdragen aan computers, kunnen wetenschappers andere dingen doen. Mijn hoop is dat we daardoor slimmere experimenten gaan ontwerpen.'



genereren.' Van Miert voegt toe: 'Er is ook veel meer kennis beschikbaar uit onderzoek op organische moleculen dan voor metalen en zouten.'

Hoewel modellen als QSAR zeker een meerwaarde hebben, zijn ze niet allesbepalend, meent Van Miert. Zo'n analyse kan een van de onderdelen van een REACH-dossier zijn. 'Bovendien gebruik je altijd minstens twee verschillende modellen om een inschatting te maken.'

### Trefwoorden selecteren

Van Westen houdt zich ook bezig met toxicologie. Zo stelde hij met een algoritme een database samen met meer dan tweeduizend publicaties over onderzoek naar de werking van agrochemicaliën, bijvoorbeeld over de interactie van bestrijdingsmiddelen met ionkanalen van plaaginsecten of de werking van fungiciden tegen schimmels. Een bibliotheek met honderdduizenden publicaties vergaren, bekijken en beoordelen is voor een mens een onmogelijke opgave. Ook daar biedt automatisering volgens Van Westen een oplossing. 'Je kunt een computeralgoritme eenvoudig trainen

om de juiste papers op basis van trefwoorden in abstracts te selecteren uit een onlinebibliotheek.'

Met de dataset agrochemische gegevens kunnen onderzoekers speuren naar nieuwe verbindingen, met een specifiekere werking, zoals stoffen die een luis doden, maar een bij ongemoeid laten. Van Westen: 'Agrochemisch onderzoek kun je goed vergelijken met bijwerkingen van medicijnen in een patiënt minimaliseren. Ook daar wil je de target – een kankercel, een virus – uitschakelen, maar de andere processen ongemoeid laten. Het draait bij landbouw en geneeskunde telkens om dezelfde mechanismen en vragen. We laten zien dat we met de database een model kunnen maken dat bijvoorbeeld het ontstaan van resistentie tegen bestrijdingsmiddelen kan voorspellen.'

Het kwam er niet van om die voorspellingen in de praktijk te toetsen met experimenten; het onderzoek was meer een zijlijn tussen zijn medicinale onderzoek. 'Het lab blijft onmisbaar', benadrukt Van Westen. 'Computeralgoritmes trainen met publieke data is relatief eenvoudig geworden, maar pas als je in staat bent om nieuwe moleculen te maken en te testen, kun je het onderscheid maken. Een computer doet een voorspelling en die dingen maken fouten.' ●

***'Een computer is onbevangen'***