



Universiteit
Leiden
The Netherlands

De computer als copiloot in de zoektocht naar nieuwe medicijnen

Westen, G.J.P. van

Citation

Westen, G. J. P. van. (2023). *De computer als copiloot in de zoektocht naar nieuwe medicijnen*. Leiden. Retrieved from <https://hdl.handle.net/1887/3618424>

Version: Publisher's Version

License: [Leiden University Non-exclusive license](#)

Downloaded from: <https://hdl.handle.net/1887/3618424>

Note: To cite this publication please use the final published version (if applicable).

Prof. Dr. Gerard J.P. van Westen

De computer als copiloot in de zoektocht naar nieuwe medicijnen



Universiteit
Leiden

Bij ons leer je de wereld kennen

De computer als copiloot in de zoektocht naar nieuwe medicijnen

Oratie uitgesproken door

Prof. Dr. Gerard J.P. van Westen

bij de aanvaarding van het ambt van hoogleraar

Kunstmatige Intelligentie & FarmacoChemie

aan de Universiteit van Leiden

op 26 mei 2023



Universiteit
Leiden

Mevrouw de rector magnificus, geacht faculteitsbestuur, geachte wetenschappelijk directeur, geacht bestuur der L.P.S.V. Aesculapius, lieve familie, vrienden, collega's, zeer gewaardeerde toehoorders,

'Op het gebied van de observatie begunstigt toeval alleen de voorbereide geest'

Deze uitspraak werd op 7 december 1854 gedaan door Louis Pasteur bij zijn benoeming tot decaan van de faculteit der natuurwetenschappen aan de universiteit van Lille.¹ De uitspraak is direct van toepassing op veel wetenschappelijke doorbraken, maar ook op mijn eigen carrière. Het is mij dan ook een eer en een genoegen hier vandaag mijn oratie uit te mogen spreken bij het aanvaarden van mijn leerstoel 'Kunstmatige intelligentie en farmacochemie'.

Zoals U uit de naam al af kunt leiden bestaat mijn leeropdracht uit het combineren van het beste van twee werelden en is daarmee met recht interdisciplinair. Voor mij persoonlijk combineert dit ook een tweetal hobby's, onderzoeken en computeren....

Het Geneesmiddelenonderzoek

Waarom zou men de computer betrekken bij het geneesmiddelenonderzoek?

Waarom zou een computer een copiloot kunnen zijn?

Het geneesmiddelenonderzoek is duur, complex, en tijdrovend. Het aantal theoretisch mogelijke kandidaat-medicijnen (drug like moleculen) wordt geschat op meer dan 10 tot de macht 33. Dat is meer dan er sterren in ons universum zijn...

Om vanuit die *chemische ruimte* een geneesmiddel naar de markt te krijgen is er volgens recente schattingen gemiddeld meer dan 10 jaar en 1.5 miljard dollar nodig.² De meeste van de kandidaat geneesmiddelen die dit proces instromen vallen af. Een van de redenen is dat de eisen die aan een geneesmid-

del gesteld worden bijzonder complex zijn. Om een geneesmiddel te worden moet een kandidaat geneesmiddel uiteraard voldoende activiteit hebben in de patiënt en dient het veilig te zijn. Maar het kandidaat middel moet ook de gewenste plek in het lichaam bereiken waar het nodig is met een concentratie die voldoende is voor werking. En het moet het liefst ook in pilvorm verstrekt kunnen worden.

U begrijpt dat dit een groot aantal, vaak tegengestelde, eisen aan het kandidaat middel stelt. De truc is om een optimaal compromis te vinden. In mijn onderzoek wil ik door het gebruik van kunstmatige intelligentie dit proces verbeteren. Maar voordat ik U daar het een en ander over kan vertellen, wil ik U vandaag eerst meenemen in een beknopt overzicht van de achtergrond de vakgebieden waarop mijn leerstoel steunt.

In deze hink-stap-sprong door de geschiedenis hoop ik U duidelijk te maken hoe de twee velden elkaar in het moderne geneesmiddelenonderzoek aanvullen en versterken. En daaruit hoe mijn onderzoek daarin een rol speelt op weg naar de copiloot.

Van kleinschalig natuurproduct naar industriële productie

Sinds de vroege historie wordt gebruik gemaakt van geneeskrachtige natuurproducten maar synthetische geneesmiddelen zijn een stuk recenter. Voor de farmacochemie wil ik graag illustreren hoe de industriële productie een steeds sterkere rol is gaan spelen de laatste 2 eeuwen. Deze transitie heeft het toepassen van kunstmatige intelligentie in het onderzoeksproces mogelijk gemaakt.

Ik neem U graag mee terug naar de 19^e eeuw, het jaar 1805 in het Pruisische Duitsland. Friedrich Sertürner schreef een brief aan de hoofdredacteur van "Trommsdorffs Journal der Pharmacie" waarin hij de extractie en isolatie van een natuurproduct rapporteerde. Hij testte de werkzaamheid van de geïsoleerde stof op ratten en honden en beschreef een slaap inducerend molecuul. Zijn ontdekking werd ook na een 2^e

brief in 1806 niet opgemerkt door de wereld. Ondertussen ging het leven verder en opende hij een apotheek. Op een avond werd hij midden in de nacht met kiespijn wakker. Hij nam een kleine dosis van de geïsoleerde stof en werd uren later wakker. Hij realiseerde zich dat de stof veilig *leek* voor mensen. Hierop volgde een mini klinische test op 3 jonge vrijwilligers en hemzelf. De optimale dosis werd 15 mg, 30 mg bezorgde een vrolijk, licht gevoel in het hoofd. Hij noemde de stof naar de Griekse god van de slaap 'Morphium'. Het artikel uit 1817 getiteld: "Ueber das Morphinum als Hauptbestandteil des Opiums" veranderde het farmaceutisch onderzoek van alchemie in een geaccepteerde tak van de wetenschappen.³

Zoals U ziet speelde het toeval, maar ook een voorbereide geest een grote rol in dit proces. Sinds dat artikel ruim 200 jaar geleden heeft het geneesmiddelenonderzoek ongelooflijke snelle ontwikkeling doorgemaakt, maar ik vind het fascinerend dat dit verhaal in grote lijnen identiek is aan hoe medicijnen vandaag de dag getest worden. Een werkzame stof wordt geïdentificeerd en deze maakt verscheidene stappen door tot een goedgekeurd medicijn.

Tien jaar later in 1828 werd een gelige substantie door Johann Buchner geïsoleerd uit de schors van de wilgenboom die hij Salicine noemde. Een extract van wilgenboomschors was een bekend middel bij de behandeling van pijn en koorts. De isolatie werd door Raffaele Piria in 1838 verbeterd door verwijdering van de suikergroep, wat salicylzuur opleverde. De Fransman Charles Gerhardt slaagde er vervolgens in salicylzuur te acetyleren en verminderde daarmee de bijwerkingen. Maar hij kon de stof niet zuiveren en daardoor bleef de stof instabiel en onbekend. Hierdoor bleef de belangrijkste bron van pijnstillers natuurlijke extracten.

Dat veranderde in 1897. Arthur Eichengrün, een joodse wetenschapper bij het chemische bedrijf Bayer en zijn assistent Felix Hoffmann, vonden een methode om acetylsalicylzuur op grote schaal te synthetiseren en te kristalliseren. Daarmee was de

stof geschikt voor therapeutische toepassingen. De stof werd bekend onder de naam Aspirine. Dit markeert de verandering van het farmaceutische onderzoek van kleinschalig, lokaal, en voornamelijk gebaseerd op natuurlijke producten naar een industrieel grootschalig en gecontroleerd proces aan het einde van de 19^e eeuw.

Data verzamelen, visualiseren, en analyseren in de 19^e eeuw

U verwacht het misschien niet, maar ook op het gebied van de kunstmatige intelligentie gebeurde er al veel in de 19^e eeuw. Ik wil graag 4 voorbeelden noemen.

In 1849 ontwikkelde John Snow in London statistische kaarten en grafieken om epidemiologische data van de cholera-epidemie te analyseren. Op basis van de symptomen concludeerde hij dat de infectiebron een besmette waterbron was en niet zoals tot op dan gedacht 'slechte lucht'. Hij publiceerde dit in een artikel "On the Mode of Communication of Cholera",⁴ een van de eerste artikelen met 'data-driven' conclusies op basis van statistiek.

Net zo analyseerde de Engelse zuster Florence Nightingale de informatie van een grote groep patiënten. Ze vertrok naar het slagveld van de Krimoorlog in 1854 geïnspireerd door een artikel over de slechte omstandigheden waarin de Engelse soldaten verbleven. Ze concludeerde dat er meer soldaten door de sanitaire omstandigheden en slechte hygiëne in de ziekenhuizen overleden dan door oorlogshandelingen. Door haar toedoen daalde het sterftepercentage spectaculair van 42% naar 2%... Bijvoorbeeld door de introductie van handen wassen! Terug in London probeerde zij ook daar hervormingen door te voeren met gematigd succes. Uniek aan haar werk was dat zij een voor de tijd revolutionaire collectie aan data visualisatie methoden ontwikkelde die in een oogopslag duidelijk maakten dat het merendeel van de slachtoffers te voorkomen was. Ook vandaag is heldere visualisatie essentieel voor de boodschap.

Ten derde mag het werk van Crum Brown en Fraser niet ontbreken als link naar de farmacochemie. In 1865 publiceerden

zij een artikel: 'On the connection between chemical constitution and physiological action'.⁵ Hoewel ze de link legden was hun werk nog niet heel concreet. Het artikel sluit af met de volgende zin (vrij vertaald):

'Dit onderzoek heeft weinig meer gedaan dan ons kennis laten maken met een enorm onderzoeksgebied, maar het heeft ons gerechtigd te verwachten dat belangrijke vruchten kunnen worden verkregen in de toekomst.'

Een vooruitziende blik als men naar het veld nu kijkt.

Tot slot volgde aan het einde van de 19^e eeuw een ontdekking van Herman Hollerith. In 1889 publiceerde Hollerith 'An Electric Tabulating System', dit artikel beschreef een systeem wat informatie kon bevatten door de aanwezigheid of afwezigheid van een gat in een kaart.⁶ Deze 'punch cards' werden gebruikt om data van de volkstelling tussen 1880 en 1890 te verwerken en zouden tot in de jaren 70 de dominante manier van automatische data verwerking blijven. Ze vormden de basis voor het bedrijf van Hollerith de "Tabulating Machine Company", dit werd na een aantal fusies hernoemd naar "International Business Machines Corporation" of IBM.

Samenvattend zijn er grote stappen gezet in de 19^e eeuw. Van simpelweg beginnen met het analyseren van relatief grote hoeveelheden data, tot het visualiseren van deze informatie en het automatische kunnen invoeren van deze informatie. In de 20^e eeuw zullen deze basis ontwikkelingen de verdere ontwikkelingen katalyseren.

De introductie van chemotherapie in de vroege 20^e eeuw

Eerst hink-stap-springen wij terug naar de farmacochemie in de 20^e eeuw. Paul Ehrlich, die zich richtte zich op het vinden van een 'Zauberkugel' (magische kogel) tegen infecties, mag niet ontbreken. In 1909 werd door hem Salvarseen ofwel 'reddend arseen' ontdekt. Het is een synthetisch molecuul voor de behandeling van Syfilis. Dit werd op de markt gebracht door

het chemische bedrijf Hoeschst. Het middel staat bekend als het eerste chemotherapeuticum en belichaamt industriële productie.

Salvarseen werd in 1928 vervangen door een middel ontdekt door Alexander Flemming. Flemming had zijn spullen niet netjes opgeruimd toen hij op vakantie ging. Hij had een aantal platen met bacteriën laten groeien en een daarvan was beschimmeld. Hij zag dat naast deze schimmels geen bacteriën groeiden en gebruikte de meest voorkomende uitspraak bij een grote wetenschappelijke ontdekking: "hé dat is gek". Hij isoleerde de stof die daarvoor verantwoordelijk was, penicilline. Minder bekend is dat Flemming al lang innoveerde met antibiotica. Flemming werkte al vroeg met Salvarseen, waarvan hij intraveneuze toediening pionierde, en ontdekte in 1921 lysozym uit oogvocht (met een bacteriedodende werking).

Men kan dus stellen dat er een grote mate van toeval was, maar zeker ook sprake van een voorbereide geest.

5

Naar schatting heeft penicilline zo'n 200.000 levens gered tijdens de 2^e wereldoorlog. Het was vanaf 1944 op grote schaal beschikbaar voor de geallieerden door een samenwerking van het leger met farmaceutische bedrijven. Bij de asmogendheden was het beperkt beschikbaar voor de troepen. Daarom wordt er gespeculeerd dat de industriële productie van penicilline een directe bijdrage aan de uitkomst van de oorlog geleverd heeft.

De periode tijdens en direct na de oorlog wordt gekenmerkt door een snelle groei van de farmacochemie en vele ontdekkingen. Zo was streptomycine in 1943, het eerste effectieve geneesmiddel tegen tuberculose. In 1955 volgde chloordiazepoxide, het eerste benzodiazepine kalmeringsmiddel, op de markt wordt gebracht door Hoffmann-La Roche. En in 1957 werd het eerste orale anticonceptiemiddel Enovid door het bedrijf Syntex, geïntroduceerd met grote maatschappelijke gevolgen.

Een vaderlands tintje, het bedrijf Organon uit Oss had Enovid nauwlettend gevolgd. Onder leiding van Stefan Szpilfogel werd het middel Lynestrol ontwikkeld, een synthetische stof met eenzelfde anticonceptie werking. Hierop werd octrooi werd aangevraagd en het werd in 1962 op de markt gebracht.

De groeiende rol van industriële productie, het betere begrip van de chemie, en de ruimte voor ongebonden onderzoek karakteriseren deze gouden jaren van het medicijnonderzoek die onze maatschappij blijvend veranderd hebben.

De eerste structuur-activiteitsrelaties begin 20^e eeuw

We springen terug naar de kunstmatige intelligentie in de eerste helft van de 20^e eeuw. Toen werd de basis gelegd voor de huidige ontwikkelingen. Ik wil dit graag met drie voorbeelden illustreren.

6 Allereerst het werk van Meyer en Overton gepubliceerd in 1899 en 1901,^{7,8} die verder gaan na Crum Brown en Fraser. Ze bevestigden onafhankelijk van elkaar dat de mate van toxiciteit van verdovende middelen op levende wezens gerelateerd kan worden aan de olie-water partitie coëfficiënt (de mate van vetoplosbaarheid). Meyer publiceerde een studie op kikkers waar hij een relatie vond tussen de vetoplosbaarheid en concentratie van een stof. Dit werd na twee jaar bevestigd door Overton. De publiceerden de Meyer-Overton correlatie. In 1935 werd een volgende stap gezet door Hammet, die elektro-nische substitutie constante introduceerde en deze kon correleren met reactie en equilibrium constanten.

Het werk van Overton en Meyer enerzijds en van Hammet anderzijds vormden in 1962 de bouwstenen voor het werk van Corwin Hansch. Hij liet met een aantal coauteurs zien dat de biologische activiteit van een serie fenoxiazijnzuren in haver verklaard kon worden op basis van de fysicochemische eigenschappen van de moleculen. Hansch wordt gezien als de vader van de kwantitatieve structuur-activiteitsrelaties.⁹ Op basis van

de vindingen van anderen zette zijn 'prepared mind' een fundamentele volgende stap.

Bij het eerdergenoemde IBM vonden interessante ontwikkelingen plaats. In 1955 werd daar een onderzoeksgroep opgezet door Nathaniel Rochester. Het doel was patroonherkenning en informatietheorie onderzoeken. Een van de projecten draaide om gesimuleerde neurale netwerken op een IBM 704 computer. In 1955 werd Rochester benaderd door John McCarthy, een jonge wiskundige die bij IBM werkte. Met de steun van hem en twee anderen schreven ze een subsidieverstrekker aan: *"Wij stellen voor om in de zomer van 1956 een seminar van twee maanden te houden over kunstmatige intelligentie met tien deelnemers op Dartmouth College."*¹⁰

Deze bijeenkomst, bekend als de Dartmouth Conference, wordt gezien als de geboorteplaats van de kunstmatige intelligentie. De theorie van de huidige deep-learning modellen van nu stamt uit de jaren 50.

Deze innovaties zijn de basis onder de methoden die later achtereenvolgens onder de namen 'expertsystemen', 'machine learning', en 'kunstmatige intelligentie' toegepast worden in het geneesmiddelenonderzoek. In mijn onderzoek worden nog steeds moleculaire eigenschappen (fysicochemisch, vingerafdrukken, of geleerde beschrijvingen) gekoppeld aan een biologische activiteit via een wiskundige vergelijking. We bouwen op de basis die in de 19^e en 20^e eeuw gelegd werd.

Doorbraken aan het einde van de 20^e eeuw

Terug naar de farmacochemie in de tweede helft van de 20^e eeuw. Het kloneren werd geïntroduceerd. Hiermee kan de rol van genen en hun rol in het ziekteproces bestudeerd worden. De farmaceutische industrie maakte een transitie door. Van het observeren van effecten van kandidaat geneesmiddel in diermodellen naar het identificeren van genen die een cruciale rol spelen in het ziekteproces (ontdekking van aangrijpingspun-

ten). Hierdoor werd de ontwikkeling van een nieuwe generatie industriële processen zoals 'high-throughput screening', 'virtual screening', en 'rational design' mogelijk.

Een van de succesverhalen waar deze technieken hun waarde konden bewijzen is de AIDS-epidemie in de jaren 80. Voor deze nieuwe ziekte was geen oorzaak, geen behandeling, en er was geen toekomst voor patiënten. Na de isolatie van HIV in 1983 werd in 1987 de eerste behandeling geïntroduceerd, Zidovudine een nucleotide analoog. Maar al snel ontstond resistentie en terugval. Hoewel meerdere analogen werden geïntroduceerd kon geen het virus permanent onder controle houden. Dit veranderde in 1996 met de publicatie van twee onafhankelijke artikelen die de behandeling van patiënten met twee nucleotide analogen en de protease inhibitor Indinavir (goedgekeurd in maart 1996) introduceerde. De combinatie zorgde voor een reductie van de sterfte door AIDS van maar liefst 60% binnen 2 jaar en vormde de basis voor een groot aantal combinatie therapieën. Bijvoorbeeld via middelen met minder bijwerkingen zoals Darunavir van het bedrijf Tibotec uit 2006. Het is bijzonder hoe de farmaceutische industrie in een tijdspanne van ongeveer 15 jaar de patiënten een toekomst heeft kunnen bieden die zij in de jaren 80 niet hadden.

Mijn eigen eerste contact met de computationeel geneesmiddelenonderzoek bij het farmaceutische bedrijf Tibotec was ook op het gebied van anti-HIV middelen, namelijk de non-nucleoside reverse transcriptase inhibitoren.

Farmacochemie en kunstmatige intelligentie in de 21^e eeuw

Eind 20^e en begin 21^e eeuw kwamen verdere doorbraken waaronder het humane genoom, innovaties chipontwikkeling, en krachtiger algoritmen die ook op een grafische processor kunnen draaien. In combinatie met de industriële grootschalige processen en historische data hebben ze de stormachtige ontwikkelingen op het gebied van kunstmatige intelligentie en farmacochemie gekatalyseerd. De nieuwe technieken vormen geen van allen een 'Zauberkuhel', maar hebben voor nieuwe

inzichten gezorgd en nieuwe toepassingen geïntroduceerd. Kunstmatige intelligentie past in het rijtje, ook deze techniek vormt geen panacee maar zal onverwachte doorbraken katalyseren.

Zo is kunstmatige intelligentie succesvol ingezet schijnbaar onoplosbare problemen op te lossen zoals de vouwing van eiwitten voorspellen op basis van slechts aminozuur sequentie (AlphaFold, ontwikkeld door Google Deepmind).¹¹

Ik hoop dat U nu een begrip heeft van de ontwikkelingen die er plaats hebben gevonden op het gebied farmacochemie en het gebied kunstmatige intelligentie en hoe deze natuurlijk naar elkaar toe groeien. Maar hoe ga ik de computer als copiloot gebruiken in de zoektocht naar nieuwe kandidaat-medicijnen?

Voorspellen van biologische activiteit op basis van heterogene data

Ik begin graag bij 'machine learning' voor patroonherkenning. Simpel gezegd zijn deze modellen te vergelijken met een spam filter. Waar het spamfilter op basis van de aanwezigheid van woorden een inschatting maakt, doen wij dat op basis van eigenschappen van kandidaat geneesmiddelen. U kunt uzelf vast voorstellen dat een email met de woorden 'chemie', 'geaccepteerd', of 'vergadering' relevanter is dan een email met de woorden 'je hebt gewonnen' of 'male performance'. Onze modellen vinden patronen in moleculen die kunnen voorspellen of een kandidaat geneesmiddel interessant is of niet (spam).

De rode draad in mijn carrière is het combineren van diverse en heterogene data om door kunstmatige intelligentie modellen voor patroonherkenning te maken. Mijn unieke expertise is het combineren van deze data uit gerelateerde velden zoals chemie en cheminformatica met biologie en bioinformatica. Dit zorgt voor modellen die brede toepassingen en een hogere betrouwbaarheid hebben dan modellen die alleen naar de chemie of biologie kijken. Innovatieve modellen op basis van publieke data vormen mijn hoofdonderzoekslijn.

Tijdens mijn master leerde ik dat experimenten in het natte lab niet mijn favoriete bezigheid zijn. Ik begon daarom aan een stage op het gebied van computationeel geneesmiddelenonderzoek onder begeleiding van Herman van Vlijmen en Joerg Kurt Wegner bij het farmaceutisch bedrijf Tibotec (nu Janssen Pharmaceutica). Ik dacht naïef dat eventuele mislukte experimenten het resultaat zouden zijn van fouten in mijn code die ik op kon sporen. Dit is natuurlijk niet waar, want *in silico* is wetenschap nog steeds wetenschap, soms werkt het niet. Ik heb wel geleerd dat ik genoot van het sterk interdisciplinaire karakter van computationeel onderzoek en ben daarom gebleven.

Bij Tibotec werkte ik met multidisciplinaire data en hoe deze te combineren via machine learning. In 2006 hebben wij modellen gemaakt die de resistentie effecten van mutaties in HIV op de werking van medicijnen kunnen voorspellen. Dit is een hulpmiddel om kandidaat-medicijnen te selecteren met een brede werking tegen meerdere varianten van een virus. Zo voorkomt men dat het middel snel de werking verliest door resistentie. Van 1 machine learning-model *per aangrijpingspunt* naar een systeem waarin je meerdere eiwitten kunt modelleren in 1 machine learning-model zorgt voor nieuwe mogelijkheden en verbetert de kwaliteit van voorspellingen.¹² De techniek heet proteochemometrisch modeleren. Het gebruik groeit door betere algoritmen en snellere chips.

Met financiering van Janssen hebben we deze techniek verder onderzocht tijdens mijn promotie aan de Universiteit Leiden onder begeleiding van Ad IJzerman, Herman van Vlijmen, en Andreas Bender. De techniek uit mijn stage bleek ook geschikt om de verschillen tussen binding van een molecuul aan een menselijk eiwit of de tegenpool in een diemodel te verklaren.¹³ Maar bleek ook geschikt om op grote schaal te voorspellen hoe mutaties de effectiviteit van combinatie therapieën kunnen beïnvloeden.¹⁴

Deze basis ben ik tijdens mijn postdoc verder uit gaan bouwen maar de techniek passen wij nog steeds toe in mijn groep.

Publieke databases veranderen de wereld

Waar ontwikkelingen op het gebied van algoritmen en rekenkracht snel gingen was de beschikbaarheid van data over biologische activiteit een probleem, zeker aan de universiteit. Hoewel er data beschikbaar waren binnen farmaceutische bedrijven was dit in het publieke domein niet het geval. Er was een commerciële database, StarLITE, die werd voor veel geld verkocht en bevatte in wezen de gecombineerde literatuur uit de farmacochemie tijdschriften. De database bevatte niet alleen de structuren van kandidaat-medicijnen en experimentele moleculen maar ook hun biologische activiteit.

Alles veranderde in 2008 toen de StarLITE database door het Europese Bioinformatica Instituut met steun van de Wellcome Trust gekocht werd. Ze zetten deze als ChEMBL op het internet op basis van een open access licentie. Daarmee is iedereen in staat de data te gebruiken en werd het in een klap mogelijk om in het publieke domein baanbrekend onderzoek te doen in de kunstmatige intelligentie en farmacochemie. Beschikbare data in openbare bronnen is dan ook essentieel voor het succes van de kunstmatige intelligentie in de farmacochemie.

Ik denk dat het essentieel is dat wij onze promovendi, maar ook de master en bachelor studenten onderrichten in het gebruik van data uit publieke databases en ook bewustwording over datakwaliteit overbrengen.

Tijdens mijn promotie had ik veel met deze data gewerkt en daarom ben ik in 2015 met mijn gezin naar het Verenigd Koninkrijk verhuisd voor een 3-jarige postdoc bij de ChEMBL database. Ik ben nog altijd dankbaar dat Afke dat avontuur aan wilde gaan. Mijn oudste zoon Max was 1,5 en mijn middelste kind Iris werd 2 maanden na de verhuizing geboren in Cambridge. Het was een drukke maar ongelooflijk interessante tijd.

Onder begeleiding van John Overington heb ik machine learning op heterogene openbare data verder onderzocht. Ik heb de bioactiviteit van agrarische chemicaliën voorspeld

met Mark Forster van het bedrijf Syngenta,¹⁵ en algoritmen ontwikkeld om automatisch de wetenschappelijke literatuur te doorzoeken in een project met George Papadatos. We hebben ook modellen gemaakt die de manier van kandidaat medicijnen aan hun aangrijpingspunten voorspeld. Dit heb ik later in Leiden met mijn eerste promovenda Lindsey Burggraaff verder ontwikkeld.^{16,17} Tot slot, als gast aan de universiteit van Cambridge hebben wij een methode ontwikkeld om de activiteit van antikankermiddelen op de NCI-60 kanker cellijnen te voorspellen met Isidro Cortes-Ciriano en Andreas Bender.¹⁸

Op het snijvlak van farmacochemie en informatica

Veel van mijn onderzoek is door samenwerking met experts van verschillende achtergronden. Er is zeer gedetailleerde kennis nodig en de ontwikkelingen op het gebied van de kunstmatige intelligentie gaan razendsnel. Hierom is het goed de verbindingen tussen onderzoekers aan de instituten van farmaceutische wetenschappen en informatica kort te houden. Bottom up ideeën kunnen tot doorbraken leiden en er zouden middelen moeten zijn om promovendi te laten starten in volledig vrij onderzoek op deze snijvlakken.

Na mijn postdoc keerde ik terug naar Leiden op een functie tussen de instituten voor informatica en farmaceutische wetenschappen. Hier werkte ik met Walter Kusters en Wojtek Kowalczyk. Ik heb het gebruik van geavanceerde machine learning-technieken in combinatie met heterogene data onderzocht. Samen met Bart Lenselink, Brandon Bongers, en Niels ten Dijke waren wij de eersten die een geïntegreerd deep-learning model op alle hoge kwaliteit ChEMBL data publiceerden. Het artikel was getiteld: “Beyond the hype” en dit was op basis van neurale netwerken waarvan de wortels, zoals U nu weet, in de jaren 50 van de vorige eeuw liggen.¹⁹

Dankzij rekenen op grafische kaarten kon ik in 2013 experimenteren met deep-learning wat in het werk uit 2017 resulteerde. Zolang gamers mooiere graphics willen kunnen wij

sneller rekenen. Recent is Andrius Bernataclavus in mijn groep begonnen in een samenwerking met Aske Plaat van het LIACS.

Hierna heb ik een VENI-beurs verkregen bij NWO, waarmee ik toe kon treden tot een tenure track in 2016. Doel was het onderzoeken van de rol van G Eiwit-gekoppelde receptoren in kanker door het karakteriseren van mutaties uit patiënten data. Dit doel wilde ik door een combinatie van bioinformatica, kunstmatige intelligentie, en experimentele validatie bereiken. Het project was erg succesvol. De verwachting was dat mutaties met een direct effect een minderheid zouden vormen. Het omgekeerde bleek, mutaties die geen aantoonbaar effect hadden vormden een minderheid. Dit resulteerde in meerdere artikelen, een grote analyse van het landschap van de mutaties in GPCRs door promovendi Brandon Bongers en Marina Gorostiola Gonzales.²⁰ Maar ook een aantal artikelen die individuele receptoren en mutaties tot in detail onderzochten door Xuesong Wang in samenwerking met Laura Heitman en Ad IJzerman.²¹

Een digitale tweeling

In de heterogene datamodellering verlegde ik toen de focus richting de signaal routes binnen de cel in een samenwerking met Mi Yang en Julio Saez-Rodriguez van de Universiteit van Aachen.²² Binnen het LACDR werkte ik met Liesbeth de Lange, Wilbert de Witte, en Anna Vlot. We combineerden modellen die de interactie van kandidaat-medicijnen op hun aangrijpingspunten voorspellen (werkt het?) met modellen die de verspreiding van kandidaat-medicijnen door het lichaam voorspellen (komt het waar het nodig is?).²³ Nu loopt een project met Coen van Hasselt en Piet Hein van der Graaf waarin promovenda Helle van de Maagdenberg dit verder trekt naar Virtual Drug Discovery.

De volgende stap is naar het organisme, waarvoor ik betrokken ben bij het virtual human 4 safety-project (VHP4Safety), via de Nationale Wetenschapsagenda. Onder leiding van de Univer-

siteit van Utrecht maken we deel uit van een consortium met onder andere Joost Beltman, Daan Geerke, Bob van de Water, Chris Evelo, Egon Willighagen, en promovenda Linde Schoenmaker uit mijn groep. Doel is om diverse computationele benaderingen te combineren voor het voorspellen van toxiciteit.

Uiteindelijk zou het geweldig zijn om een digitale patiënt te maken voor veiligheidstesten van geneesmiddelen zonder proefdieren of betere doseringen in kwetsbare groepen zoals baby's en ouderen. Maar eerst leren we elkaars talen (chem-informatica, farmacokinetiek, toxicologie, enz.) en definiëren we het raamwerk. Ik zie dit project als een natuurlijk doel van de onderzoekslijn waar ik in 2006 mee gestart ben, het modeleren van diverse heterogene data in geavanceerde moleculaire spam filters, en dit vormt de eerste copiloot.

Tekst suggestie voor kandidaat-medicijnen

10 Naast mijn hoofdonderzoekslijn ben ik de afgelopen jaren een tweede lijn gestart. Hierin wordt geen medicijn spam van nuttige medicijnen gescheiden, maar wordt een andere toepassing van kunstmatige intelligentie onderzocht. Een goed voorbeeld is uw telefoon. Bijna iedereen heeft een smartphone, wellicht met uitzondering van een enkele zeer bekende Nederlander op een Nokia. Als U een bericht typt zal de telefoon op basis van de input, nieuwe woorden voorstellen. Bijvoorbeeld 'ik', zorgt voor 'kom', en voor 'er' en 'aan'. Deze suggesties zijn op basis van historische data.

Wij gebruiken modellen die doen met moleculaire data geschreven als een serie tekens. Hier staan letters voor atomen, haakjes voor vertakkingen, en nummers voor ringen (SMILES). Onze algoritmen leren patronen in databases en kunnen dan op basis van de eerste tekens vervolg tekens voorstellen. Daarmee kunnen ze kandidaat-medicijnen voorstellen die nog geen chemicus ooit gemaakt of bedacht heeft. Wij kunnen dit doen met 10.000 nieuwe ideeën, of 1.000.000, of meer. U snapt, onze spamfilters zijn essentieel om het kaf van het koren te scheiden. De combinatie van deze methoden, een

ideeën generator en een filter, maakt het mogelijk de chemische ruimte van 10 tot de 33^e gericht te doorzoeken naar dat ene gewenste molecuul. Deze combinatie vormt de tweede copiloot

De software, Drug Explorer of "DrugEx", is begonnen door een promovendus in mijn groep, Xuhan Liu. Er zijn tot op heden 3 versies uitgebracht, allen vrij te downloaden, en een 4e artikel is net geaccepteerd.²⁴ DrugEx wordt in samenwerking met het instituut voor informatica en het farmaceutisch bedrijf Galapagos verder uitgebreid. We willen rekening houden met meerdere eigenschappen en de betrouwbaarheid van onze modellen. Dit is een project met Michael Emmerich, Pieter Stouten, Bart Lenselink, en postdoc Sohvi Luukkoonen in mijn groep.²⁵ Ook willen we de complexiteit om de verbindingen in het lab te maken meenemen in een samenwerking met Mike Preuss en promovendus Alan Hassan. Om de software te testen is organisch chemicus Jerre Madern in mijn groep gestart. Hij maakt de suggesties van onze computerprogramma's in het laboratorium.

3-dimensionale structuren benutten

De derde en laatste hoofdonderzoekslijn in mijn groep focust op 3-dimensionale structuren van aangrijpingspunten (eiwitten). In mijn promotie heb ik hier kort aan gewerkt voor virale mutanten, en met de komst van Willem Jespers als universitair docent in mijn groep kunnen we verder gericht onderzoek doen. Een van de richtingen is het combineren van data uit dynamische simulaties van eiwitten en hun liganden met de analysemethoden van de kunstmatige intelligentie. Promovenda Marina Gorostiola Gonzalez werkt hieraan en onderzoekt hoe dynamische informatie gebruikt kan worden om betere kunstmatige intelligentie modellen te ontwikkelen in een samenwerking met Laura Heitman. Deze 3^e copiloot, passen we toe in samenwerkingen met Sebastian Pomplun van het LACDR, maar ook met het Stephan Hacker, Steffen Brünle en Yangli Zhou van het LIC, en Alex Vahrmeijer en promovenda Nada Badr van het LUMC.

Interdisciplinariteit in de praktijk

De 3 lijnen ontwikkelen methoden die agnostisch voor ziekte-domein zijn. Het combineren van deze 3 lijnen is een ongelooflijk interessante onderzoeksrichting. Toepassingen gebeuren in samenwerking met experts op een bepaald domein. Op termijn is het doel de combinatie van algoritmen koppelen aan een robot die kan synthetiseren, ingebed in een ander groot project. Dit project is de OncoPreclinical Accelerator for Cancer Treatment (OncoPACT) met Mario van der Stelt. In OncoPACT ben ik verantwoordelijk voor het KI-platform met Tassos Perrakis van het NKI en is Mario van der Stelt verantwoordelijk voor het kleine moleculen consortium. Dit groeifondsprogramma heeft de ambitie om sneller en goedkoper nieuwe effectieve geneesmiddelen tegen kanker te ontwikkelen. Het consortium bevat academische partners, commerciële partners, en meer, en is een unieke samenwerking. Het KI-platform onderzoekt het ontwerp van nieuwe moleculen, het verkrijgen van inzicht in werkingsmechanismen, het testen van combinaties van moleculen, en meer. Ik ben verheugd een nieuwe generatie van gereedschappen te ontwikkelen en deze toe te passen bij de ontdekking van antikanker geneesmiddelen.

Geavanceerde tools voor de eindgebruiker

Om onderzoeksresultaten te valoriseren is het van belang dat chemici en commerciële partijen onze innovaties toe kunnen passen. Ook binnen de faculteit werk ik samen met Mario van der Stelt van het instituut voor chemie. Om wetenschappers met een chemische achtergrond in staat te stellen onze hulpmiddelen te gebruiken, zijn gebruikersinterfaces die toegang geven tot het volledige potentieel van onze gereedschappen essentieel. Anthe Janssen ontwikkelde daarvoor tijdens zijn promotie onder onze begeleiding Drug Discovery Maps,²⁶ een implementatie van patroonherkenning.

Voor onze ideeën generators hebben we Generator User Interface (GenUI).²⁷ Deze is ontwikkeld door Martin Sicho in mijn groep in een samenwerking met Daniel Svodzil van de Univer-

siteit van Praag. GenUI is een interface tussen chemicus en de software. GenUI kan via de browser benaderd worden. Tegelijkertijd onderhoudt mijn groep de grootste bekende openbare dataset die “Papyrus” wordt genoemd, naar de Leiden Papyrus. Deze database is ontwikkeld door promovendus Olivier Bequignon en vrijelijk beschikbaar ter download.²⁸ Op Papyrus kan men algoritmen testen of nieuwe kandidaten ontdekken, maar we hebben ook scripts opgenomen om interne of historische gegevens compatibel te maken met Papyrus.

Interdisciplinaire en toepassingsgerichte hubs

Valorisatie in het BioScience park volgde in 2018 via het Center for Computational Life Sciences (CCLS). CCLS liet me binnen en buiten de muren van de universiteit kijken. Zo kwamen er projecten tot stand met Nathaniel Martin en Matthijs van Haren van het Instituut voor Biologie met de focus op transferases en antibiotica.²⁹ Dit is ook zo in de samenwerking met Meindert Lammers van IBL, Boudewijn Lelieveldt van LUMC, en Thomas Baeck van LIACS. Daar onderzoekt promovenda Rosan Kuin de ontwikkeling van resistentie voor antibiotica in tuberculose. Tot slot maakte postdoc Colin Bournez een model om antibiotische peptiden te vinden in een samenwerking met Bas Zaat van het AMC en Jan Willem Drijfhout van het LUMC.³⁰

Met het Center for Human Drug Research vormde mijn groep via promovendus Hein van der Wall een link in projecten met Koos Burggraaf, Robert-Jan Dool, Rob Zuyker, en Robbert Rissmann.³¹ Doel was het toepassen van kunstmatige intelligentie in vroeg klinische setting om meer uit de data te halen en efficiënter onderzoek te doen.

Met TNO werkt postdoc Manuel Gonzales samen aan fibrose met Roeland Hanemaaijer, Lars Verschuren, en Jennifer McCormack.

Tot slot is er een succesvolle samenwerking met Dorien Peters van het LUMC en Leo Price van Crown Bioscience in een

Marie Curie trainingsnetwerk op het gebied van Autosomaal dominante polycysteuze nierziekte.³² Hierop zijn promovendi David Araripe en Bola Khalil werkzaam en Janssen Pharmaceutica uit België is ook betrokken.

De kruisverbanden zijn veel breder dan ik vandaag kan delen. Mijn leerstoel maakt deel uit van de Leidsche interfacultaire samenwerking Social AI for the Life Sciences (SAILS), waar Joost Batenburg aan het roer staat. De interactie met wetenschappers uit het domein voor de talen, de archeologie, en meer is uitermate interessant. De Leiden Early Drug Discovery & Development hub (LED3) met farmaceuten, chemici, biologen, en informatici is een tweede voorbeeld. Hier levert een interdisciplinaire groep meer op dan de som der delen. CCLS, SAILS, en LED3 zijn door wetenschap gedreven platforms en een bron van inspiratie en ideeën. Ze benadrukken mijn overtuiging dat de beste wetenschap tot stand komt in interdisciplinaire projecten op het snijvlak van de klassieke disciplines, de kunstmatige intelligentie is een verbindingsmiddel.

“Kunstmatige intelligentie zal wetenschappers niet vervangen, maar wetenschappers die Kunstmatige intelligentie gebruiken, zullen degenen vervangen die dat niet doen”.

De volgende generatie

Een deel van deze wetenschappers die AI gebruiken wordt in Leiden opgeleid. Als academicus aan een universiteit is onderwijs verweven met onderzoek. De coronapandemie heeft een ontwikkeling aangewakkerd in het gebruik van moderne hulpmiddelen. Een voorbeeld is het gericht fragmenten (video's) over één onderwerp maken die de diepte ingaan en dan het geplande interactiemoment te gebruiken voor een korte quiz om hun begrip te testen en onduidelijke stukken aandacht te geven. Ik ben benieuwd om te zien hoe het onderwijs zich gaat ontwikkelen nu we het vermogen omarmen om grotere groepen te onderwijzen met behulp van moderne communicatiemiddelen. Tegelijkertijd is het interessant hoe hulpmiddelen

zoals ChatGPT en MidJourney ons zullen verplichten om educatieve programma's en testmethoden opnieuw uit te vinden. Ik zie ChatGPT en consorten niet als een bedreiging, maar als een kans. Een kans om studenten het gebruik van deze gereedschappen aan te leren.

Wel maak ik mij zorgen over het onderwijs en onze kenniseconomie die erop steunt. Ik ben sterk van mening dat het hoger onderwijs voor iedereen toegankelijk moet zijn, en ben ook erg positief over de terugkeer van de basisbeurs. Maar ik maak mij zorgen over de ingeslagen weg waarbij het aantal studenten blijft stijgen en de financiering per student tegelijkertijd daalt. Vergeleken met de ons omliggende landen in Europa doet Nederland het relatief goed als percentage van het BNP, maar dit beeld corrigeert niet voor de aantallen. Gelukkig zijn er maatschappelijke initiatieven die deze ontwikkeling om willen draaien. De kwaliteit van het onderwijs en de toegevoegde waarde van een brede universiteit staan voor mij absoluut bovenaan. Zeker de mogelijkheid voor studenten om praktische ervaring tijdens zowel BSc en MSc op te doen is essentieel. Ik denk dat het heel belangrijk is om, zeker in de bètawetenschappen, die praktische ervaring te garanderen en kwaliteit boven kwantiteit te plaatsen.

Deze ervaring betreft zowel het lab als ervaring met het gebruik van kunstmatige intelligentie. Dankzij een gulle donatie van Chris Oomen zijn wij daar nu actief mee bezig in de bachelor van biofarmaceutische wetenschappen door de introductie van wat wij de 'Digitale Labjas' noemen. Dit zal een persoonlijke omgeving voor de student zijn waarin zij kunnen experimenteren met coderen. Ik hoop dat deze faciliteit als een vliegwiel kan dienen voor een groei van de rol van kunstmatige intelligentie in het curriculum. Zelf zullen wij het gebruik van onze eigen programma's zoals DrugEx, GenUI, en het gebruik van Git onderwijzen.

Afsluitend

'Op het gebied van de observatie begunstigt toeval alleen de voorbereide geest'

Observeren, analyseren, en rapporteren. Wetenschap is meer dan alleen feiten, inzichten zijn dynamisch. De balans tussen politiek en wetenschap is precair. De politiek heeft veel baat bij een solide wetenschappelijke onderbouwing van beleid, maar feiten zouden niet selectief geselecteerd moeten worden om een politiek doel te bereiken. Sociale media en generatieve kunstmatige intelligentie (ChatGPT, Midjourney) zullen op dit vlak een uitdaging vormen die we moeten adresseren als maatschappij.

Mijn passie voor de wetenschap is alleen maar gegroeid. Het pad van mijn carrière berust deels op toeval. Hoe ik bij Tibotec terecht kwam via een toevallige email, hoe ik bij het EBI terecht kwam via een borrel na een symposium, de terugkeer naar Leiden na een gespotte kans. Maar ik durf te stellen dat een voorbereide geest de mogelijkheden zag en ben dankbaar voor de kansen die ik kreeg. Ik kijk uit naar wat er komen gaat. De eerste kandidaat medicijnen ontworpen met kunstmatige intelligentie zijn in klinische tests aangekomen, dit zijn spannende tijden!³³

Dankwoord

Rest mij nog een laatste onderwerp. Graag wil ik al mijn promovendi, postdocs, en studenten bedanken voor het onderzoek, de inspiratie, en het harde werken over de afgelopen jaren. Het is dankzij jullie inspanningen dat het onderzoek wat ik vandaag beschreef mogelijk was. Lia, Nicole, Jaco, Rongfang, en Roelof, ik ben dankbaar voor de bijdragen die jullie hebben geleverd en leveren aan het succes van onze groep. Ik wil graag Laura, Sebastian, Erik, Joost, Micha, Bob en de collega's van de divisie 'Drug Discovery en Safety' bedanken voor de ondersteuning, de samenwerking, en de collegialiteit. We hebben een bijzonder leuke werkomgeving gecreëerd, een bron voor succesvolle wetenschap!

Het LACDR-bureau en dan met name Gitty, Sharon, en Sjoerd wil ik graag bedanken voor de ondersteuning. Ik wil Hubertus Irth, Piet Hein vd Graaf, Thomas Klompmaker, en Suzanne van der Pluijm bedanken voor het vertrouwen dat ze me schonken. Ad, jouw begeleiding en wijze woorden over de afgelopen jaren zijn me zeer dierbaar, dank daarvoor en voor de steun die je altijd gaf. Herman, sinds die eerste email en de overstap naar het computationele onderzoek heb je mij met raad en daad gesteund, dank daarvoor. Douwe, ik wil je bedanken voor 20 jaar wijze raad, maar ook kameraadschap en steun. Andreas, your guidance and the companionship in science have been a great support. John, your cheerfulness, intelligence, and originality have been a source of inspiration. Geacht bestuur der L.P.S.V. Aesculapius, dank voor jullie enthousiasme, de gezelligheid tijdens de BUEX, en de rol die jullie spelen. Ik ben ook het College van Bestuur, het faculteitsbestuur, en allen die een bijdrage hebben geleverd aan mijn benoeming erg dankbaar.

13

Ik wil ik graag mijn ouders en familie bedanken voor de altijd aanwezige steun.

Joris, Iris, Max, wat een lang verhaal he? Jullie hebben er geen idee van, maar het is bijzonder leuk om papa te zijn. Ik wil jullie bedanken dat jullie hier vandaag zijn, maar ook voor alle leuke dingen die we samendoen en jullie openheid en enthousiasme om nieuwe dingen te leren.

Afke als allerlaatste wil ik jou bedanken voor je niet aflatende steun, de reflectie die je me altijd wil bieden, en partnerschap in alles. Zonder jou had ik hier vandaag niet gestaan.

Ik heb gezegd.

Referenties

- (1) Vallery-Radot, R. *The Life of Pasteur*; New York Doubleday; Page & Company, 1915.
- (2) Wouters, O. J.; McKee, M.; Luyten, J. Estimated Research and Development Investment Needed to Bring a New Medicine to Market, 2009–2018. *JAMA* **2020**, 323 (9), 844–853. <https://doi.org/10.1001/jama.2020.1166>.
- (3) Ueber Das Morphium, Eine Neue Salzfähige Grundlage, Und Die Mekonsäure, Als Hauptbestandtheile Des Opiums. *Annalen der Physik*, 1817, 55, 56–89. <https://onlinelibrary.wiley.com/doi/abs/10.1002/andp.18170550104>.
- (4) Snow, J. On the Mode of Communication of Cholera. *Edinb. Med. J.* **1856**, 1 (7), 668–670.
- (5) Brown, A. C.; Fraser, T. R. On the Connection between Chemical Constitution and Physiological Action; with Special Reference to the Physiological Action of the Salts of the Ammonium Bases Derived from Strychnia, Brucia, Thebaia, Codeia, Morphia, and Nicotia. *J. Anat. Physiol.* **1868**, 2 (2), 224–242.
- (6) Hollerith, H. An Electric Tabulating System. *The Origins of Digital Computers: Selected Papers*, 1982, 133–143. https://doi.org/10.1007/978-3-642-61812-3_9.
- (7) Meyer, H. Zur Theorie Der Alkoholnarkose. *Arch. Für Exp. Pathol. Pharmacol.* **1899**, 42 (2), 109–118. <https://doi.org/10.1007/BF01834479>.
- (8) Overton, C. E. *Studien Über Die Narkose: Zugleich Ein Beitrag Zur Allgemeinen Pharmakologie*; G. Fischer, 1901.
- (9) Hansch, C.; Maloney, P. P.; Fujita, T.; Muir, R. M. Correlation of Biological Activity of Phenoxyacetic Acids with Hammett Substituent Constants and Partition Coefficients. *Nature* **1962**, 194 (4824), 178–180. <https://doi.org/10.1038/194178b0>.
- (10) McCarthy, J.; Minsky, M.; Rochester, N.; Shannon, C. A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, 1955. <http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html> (accessed 2023-05-10).
- (11) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Židek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, 596 (7873), 583–589. <https://doi.org/10.1038/s41586-021-03819-2>.
- (12) van Westen, G. J. P.; Wegner, J. K.; IJzerman, A. P.; van Vlijmen, H. W. T.; Bender, A. Proteochemometric Modeling as a Tool to Design Selective Compounds and for Extrapolating to Novel Targets. *Med Chem Commun* **2011**, 2 (1), 16–30. <https://doi.org/10.1039/C0MD00165A>.
- (13) van Westen, G. J. P.; van den Hoven, O. O.; van der Pijl, R.; Mulder-Krieger, T.; de Vries, H.; Wegner, J. K.; IJzerman, A. P.; van Vlijmen, H. W. T.; Bender, A. Identifying Novel Adenosine Receptor Ligands by Simultaneous Proteochemometric Modeling of Rat and Human Bioactivity Data. *J. Med. Chem.* **2012**, 55 (16), 7010–7020. <https://doi.org/10.1021/jm3003069>.
- (14) van Westen, G. J. P.; Hendriks, A.; Wegner, J. K.; IJzerman, A. P.; van Vlijmen, H. W. T.; Bender, A. Significantly Improved HIV Inhibitor Efficacy Prediction Employing Proteochemometric Models Generated From Antivirogram Data. *PLOS Comput. Biol.* **2013**, 9 (2), e1002899. <https://doi.org/10.1371/journal.pcbi.1002899>.
- (15) Gaulton, A.; Kale, N.; van Westen, G. J. P.; Bellis, L. J.; Bento, A. P.; Davies, M.; Hersey, A.; Papadatos, G.; Forster, M.; Wege, P.; Overington, J. P. A Large-Scale Crop Protection Bioassay Data Set. *Sci. Data* **2015**, 2 (1), 150032. <https://doi.org/10.1038/sdata.2015.32>.
- (16) van Westen, G. J. P.; Gaulton, A.; Overington, J. P. Chemical, Target, and Bioactive Properties of Allosteric Modulation. *PLoS Comput Biol* **2014**, 10, e1003559. <https://doi.org/10.1371/journal.pcbi.1003559>.

- (17) Burggraaff, L.; van Veen, A.; Lam, C. C.; van Vlijmen, H. W. T.; IJzerman, A. P.; van Westen, G. J. P. Annotation of Allosteric Compounds to Enhance Bioactivity Modeling for Class A GPCRs. *J. Chem. Inf. Model.* **2020**, *60* (10), 4664–4672. <https://doi.org/10.1021/acs.jcim.0c00695>.
- (18) Cortés-Ciriano, I.; van Westen, G. J. P.; Bouvier, G.; Nilges, M.; Overington, J. P.; Bender, A.; Malliavin, T. E. Improved Large-Scale Prediction of Growth Inhibition Patterns Using the NCI60 Cancer Cell Line Panel. *Bioinforma. Oxf. Engl.* **2016**, *32* (1), 85–95. <https://doi.org/10.1093/bioinformatics/btv529>.
- (19) Lenselink, E. B.; ten Dijke, N.; Bongers, B.; Papadatos, G.; van Vlijmen, H. W. T.; Kowalczyk, W.; IJzerman, A. P.; van Westen, G. J. P. Beyond the Hype: Deep Neural Networks Outperform Established Methods Using a ChEMBL Bioactivity Benchmark Set. *J. Cheminformatics* **2017**, *9* (1), 45. <https://doi.org/10.1186/s13321-017-0232-0>.
- (20) Bongers, B. J.; Gorostiola González, M.; Wang, X.; van Vlijmen, H. W. T.; Jespers, W.; Gutiérrez-de-Terán, H.; Ye, K.; IJzerman, A. P.; Heitman, L. H.; van Westen, G. J. P. Pan-Cancer Functional Analysis of Somatic Mutations in G Protein-Coupled Receptors. *Sci. Rep.* **2022**, *12* (1), 21534. <https://doi.org/10.1038/s41598-022-25323-x>.
- (21) Wang, X.; Jespers, W.; Bongers, B. J.; Habben Jansen, M. C. C.; Stangenberger, C. M.; Dilweg, M. A.; Gutiérrez-de-Terán, H.; IJzerman, A. P.; Heitman, L. H.; van Westen, G. J. P. Characterization of Cancer-Related Somatic Mutations in the Adenosine A2B Receptor. *Eur. J. Pharmacol.* **2020**, *880*, 173126. <https://doi.org/10.1016/j.ejphar.2020.173126>.
- (22) Yang, M.; Simm, J.; Lam, C. C.; Zakeri, P.; van Westen, G. J. P.; Moreau, Y.; Saez-Rodriguez, J. Linking Drug Target and Pathway Activation for Effective Therapy Using Multi-Task Learning. *Sci. Rep.* **2018**, *8* (1), 8322. <https://doi.org/10.1038/s41598-018-25947-y>.
- (23) Vlot, A. H. C.; de Witte, W. E. A.; Danhof, M.; van der Graaf, P. H.; van Westen, G. J. P.; de Lange, E. C. M. Target and Tissue Selectivity Prediction by Integrated Mechanistic Pharmacokinetic-Target Binding and Quantitative Structure Activity Modeling. *AAPS J.* **2017**, *20* (1), 11. <https://doi.org/10.1208/s12248-017-0172-7>.
- (24) Liu, X.; Ye, K.; van Vlijmen, H. W. T.; IJzerman, A. P.; van Westen, G. J. P. An Exploration Strategy Improves the Diversity of de Novo Ligands Using Deep Reinforcement Learning: A Case for the Adenosine A2A Receptor. *J. Cheminformatics* **2019**, *11* (1), 35. <https://doi.org/10.1186/s13321-019-0355-6>.
- (25) Luukkonen, S.; Meijer, E.; Tricarico, G.; Hofmans, J.; Stouten, P.; van Westen, G.; Lenselink, E. Large-Scale Modeling of Sparse Kinase Activity Data. *ChemRxiv* **2023**. <https://doi.org/10.26434/chemrxiv-2023-p764w>.
- (26) Janssen, A. P. A.; Grimm, S. H.; Wijdeven, R. H. M.; Lenselink, E. B.; Neeffes, J.; van Boeckel, C. A. A.; van Westen, G. J. P.; van der Stelt, M. Drug Discovery Maps, a Machine Learning Model That Visualizes and Predicts Kinome-Inhibitor Interaction Landscapes. *J. Chem. Inf. Model.* **2019**, *59* (3), 1221–1229. <https://doi.org/10.1021/acs.jcim.8b00640>.
- (27) Sicho, M.; Liu, X.; Svozil, D.; van Westen, G. J. P. GenUI: Interactive and Extensible Open Source Software Platform for de Novo Molecular Generation and Cheminformatics. *J. Cheminformatics* **2021**, *13* (1), 73. <https://doi.org/10.1186/s13321-021-00550-y>.
- (28) Béquignon, O. J. M.; Bongers, B. J.; Jespers, W.; IJzerman, A. P.; van der Water, B.; van Westen, G. J. P. Papyrus: A Large-Scale Curated Dataset Aimed at Bioactivity Predictions. *J. Cheminformatics* **2023**, *15* (1), 3. <https://doi.org/10.1186/s13321-022-00672-x>.
- (29) Tehrani, K. H. M. E.; Wade, N.; Mashayekhi, V.; Bröchle, N. C.; Jespers, W.; Voskuil, K.; Pesce, D.; van Haren, M. J.; van Westen, G. J. P.; Martin, N. I. Novel Cephalosporin Conjugates Display Potent and Selective Inhibition of Imipenemase-Type Metallo- β -Lactamases. *J. Med. Chem.* **2021**, *64* (13), 9141–9151. <https://doi.org/10.1021/acs.jmedchem.1c00362>.

- (30) Bournez, C.; Riool, M.; de Boer, L.; Cordfunke, R. A.; de Best, L.; van Leeuwen, R.; Drijfhout, J. W.; Zaat, S. A. J.; van Westen, G. J. P. CalcAMP: A New Machine Learning Model for the Accurate Prediction of Antimicrobial Activity of Peptides. *Antibiotics* **2023**, *12* (4). <https://doi.org/10.3390/antibiotics12040725>.
- (31) van der Wall, H. E. C.; Doll, R. J.; van Westen, G. J. P.; Koopmans, I.; Zuiker, R. G.; Burggraaf, J.; Cohen, A. F. The Use of Machine Learning Improves the Assessment of Drug-Induced Driving Behaviour. *Accid. Anal. Prev.* **2020**, *148*, 105822. <https://doi.org/10.1016/j.aap.2020.105822>.
- (32) Booi, T. H.; Leonhard, W. N.; Bange, H.; Yan, K.; Fokkelman, M.; Plugge, A. J.; Veraar, K. A. M.; Dauwerse, J. G.; van Westen, G. J. P.; van de Water, B.; Price, L. S.; Peters, D. J. M. In Vitro 3D Phenotypic Drug Screen Identifies Celastrol as an Effective in Vivo Inhibitor of Polycystic Kidney Disease. *J. Mol. Cell Biol.* **2020**, *12* (8), 644–653. <https://doi.org/10.1093/jmcb/mjz029>.
- (33) Wills, T. AI Drug Discovery: Assessing the First AI-Designed Drug Candidates to Go into Human Clinical Trials. September 23, 2022. <https://www.cas.org/resources/cas-insights/drug-discovery/ai-designed-drug-candidates> (accessed 2023-05-10).



- 2020 - Heden Hoogleraar Kunstmatige Intelligentie en Farmacochemie (Universiteit Leiden)
- 2016 - 2020 Universitair Docent Computatoneel Geneesmiddelenonderzoek (Universiteit Leiden)
- 2015 - 2016 Postdoc aan de Universiteit Leiden
- 2012 - 2015 Postdoc aan het Europese Bioinformatica Instituut (Cambridge, Verenigd Koninkrijk)
- 2007 - 2012 Promotie aan de Universiteit Leiden in samenwerking met Janssen Pharmaceutica
- 2001 - 2007 Studie Biofarmaceutische Wetenschappen aan de Universiteit Leiden

Gerard van Westen (Leiden, 1983) studeerde BioFarmaceutische Wetenschappen in Leiden. Na zijn studie promoveerde hij in Leiden op een project in samenwerking met Janssen Pharmaceutica. Tijdens zijn promotie zette hij de eerste stappen op het gebied van kunstmatige intelligentie en ontwikkelde hij methoden voor het gecombineerd modelleren van biologisch en chemische data. Hierna deed hij een postdoc in Cambridge (Verenigd Koninkrijk) waar hij machine learning toepaste op heterogene publieke data bij het Europese Bioinformatica Instituut.

In Leiden combineert hij het veld van kunstmatige intelligentie met het geneesmiddelenonderzoek. Hij is wereldwijd expert in machine learning op het modelleren van gecombineerde chemische en biologische data. In zijn groep worden taal gebaseerde modellen gebruikt om nieuwe moleculaire structuren te ontwerpen (genereren) met gewenste activiteit- en toxiciteitsprofielen die vervolgens worden gesynthetiseerd in het natte lab. Bovendien worden structuur gebaseerde benaderingen gebruikt om een mechanistisch inzicht te krijgen in de moleculaire interactie tussen kandidaat geneesmiddelen en aangrijpingspunten. Hier worden dynamische structuur methoden gecombineerd met kunstmatige intelligentie om nieuwe kandidaat geneesmiddelen te ontwerpen en te verbeteren. Als laatste gebruiken ze modellen gebruikt om bioactiviteits-spectra te genereren om aangrijpingspunten en bijwerkingen van kandidaat geneesmiddelen te identificeren.

