Proteins in harmony: Tuning selectivity in early drug discovery

Burggraaff, L.

Samenvatting voor Leken

Samenvatting voor leken

Dit proefschrift beschrijft hoe computers kunnen worden ingezet in het ontdekken en ontwikkelingen van nieuwe medicijnen. Dat het ontwikkelen van medicijnen van groot belang is moge duidelijk zijn. Er zijn namelijk nog steeds ziektes waarvoor er geen medicijnen beschikbaar zijn of de medicijnen die er zijn leiden tot ernstige bijwerkingen.

Om te kunnen begrijpen hoe een nieuw medicijn ontwikkelt kan worden, moeten we eerst weten hoe een medicijn zijn werking uitoefent (**Hoofdstuk 1**). De potentiële medicijnen die in dit proefschrift aan bod komen vallen onder de groep 'kleine moleculen'. Dit is dezelfde klasse als bijvoorbeeld paracetamol en sildenafil (Viagra), maar ook suikers en aminozuren worden beschouwd als kleine moleculen. Dit soort medicijnen zorgen over het algemeen voor een effect in het lichaam door een interactie aan te gaan met specifieke eiwitten. Het menselijk lichaam bestaat uit heel veel verschillende soorten eiwitten, die allemaal hun eigen functie hebben. Sommige eiwitten zijn in staat signalen door te geven aan de rest van het lichaam, terwijl andere eiwitten belangrijk zijn voor de vertering van voedsel. Je kunt je voorstellen dat bij het ontwikkelen van een nieuw medicijn het dus belangrijk is te weten welk effect je wilt bereiken en welk eiwit je daarvoor moet beïnvloeden. Een pijnstiller richt zich bijvoorbeeld op eiwitten die betrokken zijn bij het doorgeven van een pijnsignaal.

Maar hoe weet een medicijn de juiste eiwitten te vinden? En hoe weet een paracetamol dat jij buikpijn hebt en geen kiespijn? In het kort: dat weet het niet. Wat dat betreft zijn medicijnen een beetje dom. Ze gaan waar ze kunnen gaan en worden uiteindelijk door ons lichaam weer uitgescheiden. Onderweg komen de medicijnen verschillende eiwitten tegen en als ze toevallig een bijpassend eiwit vinden kunnen ze daar een interactie mee aangaan. Of een medicijn-eiwit combinatie een interactie oplevert hangt af van verschillende factoren, maar grofweg kan het worden vergeleken met het passen van een sleutel in een slot. In het geval van de paracetamol past het dus op het pijneiwit – een eiwit dat op meerdere plekken in het lichaam voorkomt. Een paracetamol weet echter niet waar het moet zijn en grote kans dat het dus óók aan de pijneiwitten bindt die te maken hebben met kiespijn in plaats van alleen de pijneiwitten in je buik. Dit merk je alleen niet omdat je op dat moment geen kiespijn ervaart en het pijnstillend effect dus niet opvalt.

Tijdens het medicijnontwikkelingsproces proberen we medicijnen zo goed mogelijk te laten passen op de relevante eiwitten. Dit heet ook wel het selectief maken van een medicijn. Wanneer medicijnen namelijk ook binden aan eiwitten waar ze beter niet aan kunnen binden krijg je bijwerkingen. Het binden van een medicijn aan eiwitten die bijvoorbeeld het hart ontregelen kan dramatische gevolgen hebben.

Met de computer kunnen we tot op zeker hoogte simuleren of een stofje (klein molecuul) past op het relevante eiwit. Bovendien kunnen we ook simuleren of het stofje bindt op eiwitten die bijwerkingen veroorzaken. Die laatste stofjes zijn dus minder geschikt als medicijn. Met de computer kunnen er miljoenen stofjes per dag worden getest, waarvan vervolgens alleen de beste worden gekozen om te dienen als potentieel medicijn.

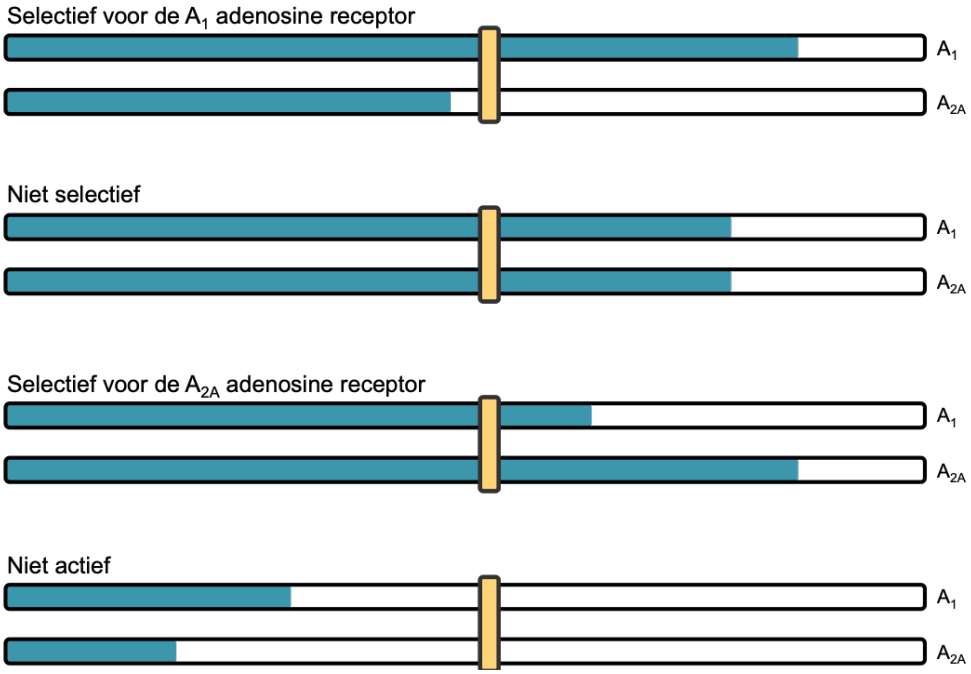
Op welke manieren de computer kan worden ingezet om de stofjes te testen staat beschreven in **Hoofdstuk 2**. Kortom zijn er twee hoofdmethodes: 1) het gebruik van 'machine learning' waarbij informatie wordt gebruikt van reeds geteste stofjes en 2) het gebruik van driedimensionale (3D) eiwitmodellen. Beide methodes worden in het algemeen gebruikt om de bioactiviteit van een stofje voor een eiwit te voorspellen. De bioactiviteit van een stofje weergeeft hoe goed het stofje zal binden aan een eiwit. Een stofje kan dus een andere bioactiviteit hebben voor eiwit A dan voor eiwit B. Heeft een stofje een hele lage bioactiviteit, dan word het stofje als inactief beschouwd voor het betreffende eiwit. Met de eerste methode, machine learning, wordt de computer getraind om patronen te herkennen in data. Vervolgens kan het gegenereerde model worden toegepast op nieuwe data om deze een waarde toe te kennen aan de hand van eerder geleerde trends. Dit klinkt abstract, maar in principe bestudeert het computermodel in zeer korte tijd duizenden stofje-eiwit combinaties. Hierdoor weet het welke stofjeseigenschappen (afmetingen, oplosbaarheid, etc.) het beste geschikt zijn voor een bepaald eiwit. Als er vervolgens een nieuw stofje door het model wordt gehaald, kan het model aan de hand van wat het al heeft gezien dus zeggen wat de bioactiviteit van het nieuwe stofje zal zijn. Om een machine learning model te kunnen trainen moet er dus op voorhand al informatie beschikbaar zijn van reeds in het lab geteste stofje-eiwit combinaties. Dit zijn namelijk "de feiten" waar het model conclusies op baseert. Een nieuw stofje dat door het model een voorspellende waarde wordt toegewezen hoeft uiteraard niet alvorens op bioactiviteit te zijn getest in het lab. Het voordeel van een machine learning model is dat het relatief snel is in vergelijking met 3D eiwitmodellen. Hoewel je dus wel voldoende data nodig hebt voordat je een nuttig model kan maken, kan het model als het eenmaal getraind is toe worden gepast op oneindig veel stofjes. De tweede modeleringstechniek om potentiële medicijnen mee te kunnen vinden is het simuleren van stofje-eiwit combinaties in 3D. Hierbij wordt er een 3D eiwitvisualisatie gemaakt en worden de stofjes er één voor één ingestopt. Hoe beter het stofje in het eiwit past, hoe hoger de bioactiviteit. Hoewel deze methode in het algemeen langzamer is dan machine learning modellen, is het vaak wel nauwkeuriger. Dit komt doordat de informatie van het eiwit in deze modellen meegenomen wordt om de bioactiviteit te voorspellen.

In **Hoofdstuk 3** wordt de machine learning methode toegepast om medicijnen te vinden voor de behandeling van diabetes type II. Een populair aangrijpingspunt voor de behandeling van deze ziekte is een transporteiwit genaamd SGLT2. SGLT2 bevindt zich in de nieren en is verantwoordelijk voor het terug transporteren van suikers in de urine naar het bloed. In een gezonde situatie is dit een handig systeem van ons lichaam dat ervoor zorgt dat er geen nuttige voedingstoffen verloren gaan. Echter, in het geval van diabetes type II zijn de bloedsuikerwaarden te hoog. Door SGLT2 met een medicijn te remmen, wordt er minder suiker terug het bloed in gestopt wat gunstige gevolgen heeft voor de bloedsuikerspiegels. Helaas hebben deze SGLT2 remmers op de langdurige termijn wel nadelige effecten als gevolg van de verhoogde suikerniveaus in de nieren. Bovendien moet het medicijn dat als een pilletje kan worden ingenomen, het hele lichaam doorreizen om bij zijn bestemming te komen. Daarom richt **Hoofdstuk 3** zich op het vinden van stofjes die binden op een alternatief doeleiwit om de bloedsuikerspiegels te verlagen. Het doeleiwit in dit hoofdstuk lijkt heel erg op SGLT2 en heet dan ook SGLT1. Hoewel SGLT1 zich ook in de nieren bevindt, is het ook aanwezig in de dunne darm. Omdat SGLT1 en SGLT2 heel erg op elkaar lijken en ook een soortgelijke functie hebben, zijn ze in staat om ook ongeveer dezelfde soort stofjes te binden. Daardoor is het lastig om een stofje selectief te maken voor SGLT1. In deze situatie is het echter niet zozeer nodig om een stofje selectief te maken voor

SGLT1 ten opzichte van SGLT2. Doordat SGLT1 zich ook in de dunne darm bevindt kan je ervoor kiezen om een stofje selectief maken voor de darmen waardoor het niet in staat is om SGLT1 en SGLT2 in de nieren te bereiken. Op deze manier zou je dus de suikeropname vanuit de darmen naar het bloed kunnen beperken en tegelijkertijd nierschade kunnen voorkomen. Een stofje selectief maken voor de darmen is mogelijk doordat je er eigenlijk voor zorgt dat het stofje het lichaam niet inkomt. Het maagdarmkanaal behoort namelijk nog tot ons uitwendig milieu. Je zou de darmen kunnen vergelijken met een groot filter: alles wat er doorheen komt is opgenomen door het lichaam en bevindt zich in het inwendig milieu. In het geval van een darm-selectief stofje zou dit kunnen betekenen dat het stofje te groot is om door het filter te passen. In **Hoofdstuk 3** wordt er met machine learning gezocht naar allerlei verschillende soorten stofjes die SGLT1 kunnen remmen. Deze stofjes dienen als een mooi startpunt voor het maken van selectieve remmers. Er wordt in het hoofdstuk vooral aandacht besteed aan het vinden van diverse stofjes – stofjes die niet veel op elkaar, of op bekende SGLT remmers lijken. Dankzij deze diversiteit is de mogelijkheid om een stofje te vinden dat darm-selectief gemaakt kan worden namelijk groter.

In **Hoofdstuk 3** ligt de focus op selectiviteit door een medicijn alleen te laten gaan waar het werkzaam moet zijn. Echter, in de meeste gevallen is een dergelijke afbakening van het verspreidingsgebied van een medicijn niet mogelijk. Hetzelfde geldt voor de casus die wordt besproken in **Hoofdstuk 4**. Dit hoofdstuk omvat twee eiwitten die bij veel processen in ons lichaam betrokken zijn: de A_1 en A_{2A} adenosine receptoren. Hoewel deze eiwitten gelijkenissen vertonen, hebben ze een tegenovergestelde werking. De één verlaagt het niveau cAMP in de cel en de ander verhoogt juist de hoeveelheid cAMP. De rol van cAMP is het doorgeven van signalen in de cel. Hierdoor worden andere eiwitten aangestuurd en ontstaat er een signaleringscascade. Met machine learning wordt er in **Hoofdstuk 4** een model gemaakt dat stofjes kan onderscheiden voor de A_1 en A_{2A} adenosine receptoren. In plaats van zoals op de traditionele manier de bioactiviteit van stofjes op beide eiwitten te voorspellen, voorspelt dit model de selectiviteit tussen de twee eiwitten. Vervolgens kunnen de stofjes gekozen worden die aan het gewenste profiel voldoen: A_1-selectief, A_{2A}-selectief of actief op beide eiwitten. **Figuur 1** laat zien hoe het verschil in bioactiviteit zich vertaalt naar selectiviteit. Hoewel het model in staat is verschillen in selectiviteit te detecteren, houdt het echter geen rekening met algehele bioactiviteit. Zo zou het kunnen dat een stofje A_1-selectief wordt voorspeld, maar eigenlijk helemaal niet bioactief is op de A_1 adenosine receptor (zoals het onderste voorbeeld in **Figuur 1**). Daarom is het selectiviteitsmodel gecombineerd met een 3D eiwitstructuur aanpak. Met 3D eiwitmodellen van zowel de A_1 en A_{2A} adenosine receptor konden de stofjes worden gepast in beide eiwitten. De stofjes die A_1-selectief werden voorspeld, maar niet in het 3D model van de A_1 adenosine receptor pasten konden er nu uitgefilterd worden. Andersom geldt dit natuurlijk ook voor stofjes die selectief zijn voor de A_{2A} adenosine receptor.

In **Hoofdstuk 5** wordt er ook gebruik gemaakt van zowel machine learning en 3D eiwitmodellen. Deze worden echter op een andere manier toegepast dan in **Hoofdstuk 4**, namelijk om het bioactiviteitsprofiel voor meerdere eiwitten te voorspellen in plaats van de selectiviteit tussen twee eiwitten. De eiwitten in **Hoofdstuk 5**, de zogenaamde kinases, zijn veelal betrokken bij complexe ziektes zoals kanker. De kaft van dit proefschrift weergeeft een vis die tegen de school inzwemt. Dit illustreert een eiwit dat zich anders gedraagt dan zou moeten, waardoor er een ziektebeeld ontstaat. In het geval van kanker zwemmen er meerdere vissen de verkeerde



Figuur 1. Verschillen in bioactiviteit weerspiegelen de selectiviteit tussen de A_1 en A_{2A} adenosine receptoren. Hoe groter de blauwe balk, hoe hoger de bioactiviteit. De minimaal benodigde hoeveelheid bioactiviteit om een effect te bereiken is weergeven in geel.

kant uit. Er zijn dan meerdere medicijnen nodig om elke vis te corrigeren. Het nadeel van het gebruik van meerdere medicijnen tegelijkertijd is dat deze medicijnen elkaar in de weg kunnen zitten, waardoor er ongewenste effecten ontstaan. **Hoofdstuk 5** beschrijft de zoektocht naar één medicijn dat meerdere eiwitten, of vissen, tegelijkertijd kan aansturen zonder de 'normale' vissen te beïnvloeden. Hiervoor worden meerdere machine learning en 3D eiwitstructuur modellen gebruikt om de bioactiviteit van stofjes te voorspellen voor een hele set kinases. Als resultaat krijg je een voorspelt bioactiviteitsprofiel voor elk stofje dat je door de modellen haalt. Vervolgens kunnen de stofjes met het meest gunstige profiel worden geselecteerd voor verder (lab)onderzoek.

Hoofdstuk 6 gaat in op een concept dat veel invloed kan hebben op het modelleren van selectiviteit, namelijk de bindingplaats van een stofje op een eiwit. Een eiwit heeft een hoofdbindingplaats waar een lichaamseigen stofje aan kan binden. Deze hoofdbindingplaats is vaak ook de plek waar een medicijn aan bindt. Echter, een eiwit kan ook meerdere bindingplaatsen hebben, welke tevens geschikt zouden kunnen zijn voor een medicijn om interactie mee aan te gaan. Vaak hebben de bindingplaatsen op een eiwit andere eigenschappen (bijvoorbeeld qua grote) wat betekent dat er verschillende soorten stofjes binden aan de verschillende bindingplaatsen. Als al deze stofjes in een computermodel worden gestopt zonder onderscheid te maken tussen de bindingplaatsen, kan dit de resultaten negatief beïnvloeden. **Hoofdstuk 6** laat zien dat wanneer de bindingplaatsinformatie meegenomen wordt in een machine learning model, dit ten gunste is van de nauwkeurigheid van voorspellingen. Helaas is bindingplaatsinformatie nog niet makkelijk

te verkrijgen en kan de verbetering dus niet zomaar worden doorgevoerd. Echter, met het 'text-mining' protocol dat in **Hoofdstuk 6** wordt beschreven kunnen de eerste stappen worden gezet naar het classificeren van de bindingplaatsen van stofjes. Door middel van text-mining kan de bindingplaatsinformatie namelijk uit wetenschappelijke publicaties worden gehaald, waarna het gebruikt kan worden om de voorspellingen van modellen te verbeteren.

Zoals in deze samenvatting is geschetst, zijn er meerdere manieren om de selectiviteit van medicijnen te modelleren en te voorspellen. Elke casus vraagt om een andere en eigen aanpak. Hoewel er dus geen eenduidige lijn is te trekken over hoe selectiviteit het beste gemodelleerd kan worden, valt er wel wat te zeggen over wat het succes van computermodellen bepaald. **Hoofdstuk 7** beschrijft onder meer dat de hoeveelheid data (informatie over bekende stofjes) belangrijk is, met name voor machine learning modellen. Deze modellen moeten namelijk een patroon leren herkennen en als er niet genoeg data is om van te leren worden de voorspellingen minder nauwkeurig. Bovendien speelt de diversiteit van data ook een rol. Een grote variabiliteit aan stofjes geeft meer kennis over wat wel en wat niet past in een eiwit. Het is dus heel nuttig om niet alleen veel data te hebben, maar ook om data beschikbaar te hebben waar veel informatie uitgehaald kan worden.

De potentiële medicijnen die met computermodellen worden geselecteerd worden eerst verder onderzocht voordat ze als medicijn beschikbaar komen. Hoewel we met computermodellen tijd kunnen winnen, duurt dit proces gemiddeld 10-15 jaar. Dit komt doordat we er zeker van moeten zijn dat de medicijnen echt veilig zijn voordat ze op de markt komen. Na het simuleren van stofjes op de computer, worden de stofjes getest in het lab. Vervolgens worden de meest succesvolle stofjes getest op dieren en uiteindelijk op vrijwilligers. Het kan zo zijn dat er ondanks het zorgvuldige werk er ergens in dit proces toch nog ernstige onvoorziene bijwerkingen aan het licht komen. Op het moment dat deze worden ontdekt betekent dit vaak het einde van de ontwikkeling van het specifieke stofje. Dat is natuurlijk zonde. Daarom is het blijven ontwikkelen van modellen die accurate voorspellingen kunnen doen van groot belang.