



Universiteit  
Leiden  
The Netherlands

## Interactive evolutionary algorithms and data mining for drug design

Lameijer, E.M.W.

### Citation

Lameijer, E. M. W. (2010, January 28). *Interactive evolutionary algorithms and data mining for drug design*. Retrieved from <https://hdl.handle.net/1887/14620>

Version: Corrected Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/14620>

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

## Samenvatting

Het ontwerpen van geneesmiddelen is moeilijk. Voor vele ziektes (zoals multiple sclerose) bestaan er nog geen geneesmiddelen, zijn de huidige geneesmiddelen onvoldoende om alle gevallen te genezen (kanker), of hebben bestaande geneesmiddelen bijwerkingen die we zouden willen verminderen (bijvoorbeeld bij de behandeling van HIV). Hoofdstuk 1 leidt dit proefschrift in en bespreekt waarom het moeilijk is geneesmiddelen te ontwerpen, hoe computers ons zouden kunnen helpen bij het ontwerp van geneesmiddelen, en het belang van goede samenwerking tussen computer en geneesmiddelenonderzoeker. Ook bespreekt hoofdstuk 1 enkele kernbegrippen die in de rest van dit proefschrift terugkomen, zoals *datamining* en *docking*. Tenslotte bevat het een overzicht van de andere hoofdstukken.

De computermethode die we het meest gebruikt hebben, die van de evolutionaire algoritmen, wordt besproken in hoofdstuk 2. Hoofdstuk 2 geeft naast een inleiding van de theorie achter evolutionaire algoritmen een overzicht van de verschillende toepassingen van evolutionaire algoritmen in geneesmiddelontwerp. Deze toepassingen zijn onder te verdelen in het voorspellen van de eigenschappen van moleculen (zoals het afleiden van relaties tussen de structuur van een molecuul en zijn activiteit) en het ontwerpen van geheel nieuwe moleculen. De verschillende toepassingen die in de literatuur bekend zijn worden besproken en met elkaar vergeleken, alvorens mogelijke ontwikkelingen worden geschetst en suggesties worden gegeven voor verder onderzoek en mogelijke toekomstige toepassingen.

Hoofdstuk 3 beschrijft het onderzoek dat ik gedaan heb met een tweede computermethode, *datamining* ('data delven'). Bij datamining probeert men met behulp van een computer patronen te ontdekken in een grote database. Hoofdstuk 3 beschrijft het vinden en analyseren van patronen in de NCI database, een database van chemische verbindingen die ten tijde van ons onderzoek iets meer dan 250,000 moleculen bevatte. Van deze database wilden we de diversiteit bestuderen, omdat hoe diverser een set moleculen is, hoe groter de kans is dat het uittesten ervan een goed kandidaat-geneesmiddel oplevert. We splitsten alle moleculen op in ringen en niet-ring-structuren, en ontdekten dat de database enigszins 'saai' was. Hoewel er duizenden verschillende ringstructuren en niet-ringstructuren in voorkwamen, vormden de tien meest voorkomende ringstructuren, die nog geen 0,1% van de totale ringdiversiteit vertegenwoordigden, 62% van alle ringen, terwijl er duizenden ringstructuren waren die maar in één of twee moleculen werden gebruikt. Ook bleek dat bepaalde structuren veel vaker dan statistisch verwacht samen voorkwamen en 'superfragmenten' vormden.

Aan de andere kant leken andere structuren elkaar te 'vermijden'. Dit dataminen leverde lijsten op met zeldzame fragmenten en ongebruikelijke fragmentcombinaties, die door chemici gebruikt zouden kunnen worden om nieuwe moleculen te ontwerpen. Zulke nieuwe moleculen, toegevoegd aan de bestaande molecuulbibliotheken, zouden de kans op het vinden van nieuwe geneesmiddelen een stuk groter kunnen maken.

In hoofdstuk 4 kom ik terug op de andere methode die we gebruikt hebben: evolutionaire algoritmen. Hoewel er al vele evolutionaire algoritmen bestonden om nieuwe moleculen te ontwerpen (hoofdstuk 2 geeft daarvan een overzicht), hadden al die methoden het probleem dat wetenschappers (nog) niet goed kunnen berekenen hoe 'goed' een molecuul zou zijn als mogelijk geneesmiddel. Daarom besloten we de kwaliteit van de moleculen op een andere manier te bepalen, namelijk met een interactief evolutionair algoritme. Dat betekent dat de gemaakte moleculen niet door de computer worden beoordeeld, maar door chemici. Zo konden we de ervaring en chemische intuïtie van de chemici gebruiken, die anders moeilijk te achterhalen zijn. Het resulterende programma noemden we de "Molecule Evuator". Hoofdstuk 4 bespreekt de details van hoe we de moleculen codeerden, de mutaties waarmee de moleculen veranderd werden, en de aanpassingen die we moesten doen om het interactieve evolutionaire algoritme bruikbaar te maken voor chemici, en de chemici in staat te stellen hun creativiteit optimaal te gebruiken.

Een van de problemen in de beginfase van de Molecule Evuator was dat het programma nog te vaak moleculen produceerde die moeilijk te synthetiseren waren, of zo instabiel waren dat ze uit elkaar zouden vallen. Om dat op te lossen hebben we de hoofdmethode uit hoofdstuk 3, datamining, toegepast. We gebruikten allereerst datamining om te achterhalen hoe vaak bepaalde groepen van 1, 2, 3 en 4 atomen voorkwamen in de World Drug Index (een geneesmiddelen-database die ten tijde van ons dataminen ongeveer 32000 verbindingen bevatte), en gebruikten die gegevens om de mutaties zó aan te passen dat bepaalde zeldzame substructuren (zoals O-F) veel minder vaak gemaakt werden. We testten de methode vervolgens uit op de fragmenten die we in de NCI hadden gevonden (hoofdstuk 3), en het bleek dat de door datamining aangepaste mutaties inderdaad vaker een ander bestaand fragment produceerden dan de oorspronkelijke mutaties. Hoofdstuk 5 geeft de gedetailleerde informatie over het dataminen, beschrijft hoe de mutaties precies werden aangepast, en bespreekt de precieze resultaten.

Een van de lacunes die naar voren kwam uit het literatuuronderzoek voor hoofdstuk 2 was dat evolutionaire algoritmen in geneesmiddelontwerp doorgaans niet met elkaar vergeleken werden. Zeker in ons veld, het ontwerp van nieuwe

geneesmiddelmoleculen, zijn er vele methodes, maar het was niet duidelijk welke methode de beste is, ofwel welke methode het effectiefst geneesmiddelmoleculen kan optimaliseren. Maar dat is wel een belangrijke vraag, omdat een goede evaluatie van een kandidaat-geneesmiddel veel tijd en geld kost, of dat nou gebeurt met een geavanceerd computermodel of met experimenten. Een evolutionair algoritme dat maar 1000 moleculen hoeft te produceren voor een bepaalde verbetering, is daardoor veel nuttiger/economischer dan een algoritme dat voor dezelfde prestatie 10,000 moleculen nodig heeft. In hoofdstuk 6 onderzoeken we één van de belangrijkste kwesties bij evolutionaire algoritmen voor geneesmiddelontwerp: is evolutie die gedaan wordt met kleine stapjes (atoom voor atoom veranderen) effectiever dan evolutie die grote sprongen maakt (dus hele stukken molecuul ineens verplaatst of verandert)? De kwaliteit van elk molecuul evalueerden we met een docking-programma, dat het molecuul in het HIV-enzym reverse transcriptase paste en de interactie-energie berekende, om een benadering te krijgen van hoe de evolutie zich zou gedragen als we de moleculen echt zouden synthetiseren en testen. Atoom-gebaseerde en fragment-gebaseerde evolutie leken het ongeveer even goed te doen, wel leerden we een aantal lessen waarmee we onze oorspronkelijke algoritmen kunnen verbeteren. Het verloop van het experiment staat samen met onze discussie en aanbevelingen in hoofdstuk 6.

Het door ons ontwikkelde evolutionaire algoritme zou samen met de resultaten van het dataminnen in staat moeten zijn chemici te helpen bij het ontwerpen van nieuwe moleculen. Maar werkt het ook in de praktijk? Om dat uit te testen lieten we de Molecule Evuator 300 moleculen genereren, die door chemici werden beoordeeld en aangepast. Uit dit proces kwamen 34 'interessante' moleculen, die we opzochten in de databases van Beilstein en de ACS om te kijken of deze moleculen of varianten erop al gemaakt waren. Van de moleculen die zowel nieuw waren als er 'geneesmiddelachtig' uitzagen, kozen de chemici een aantal uit om te synthetiseren. Uiteindelijk leverde dit acht kleine, geheel nieuwe moleculen op. Deze werden op 83 eiwitten uitgetest. Vier verbindingen bleken biologisch actief te zijn, onder andere op de  $\alpha$ -adrenerge receptoren. Hoofdstuk 7 bespreekt dit experiment, inclusief de structuren en de syntheses. Het vinden van nieuwe, biologisch actieve verbindingen toonde aan dat onze methodes nuttig kunnen zijn voor het ontwerpen van nieuwe kandidaat-geneesmiddelen: de Molecule Evuator kan samen met de expertise van chemici geheel nieuwe en toch werkzame stoffen ontwerpen.

Dit proefschrift sluit af met hoofdstuk 8. Het eerste deel hiervan beschrijft de conclusies die ik uit het in dit proefschrift beschreven onderzoek getrokken heb. Het tweede deel van hoofdstuk 8 bevat mijn gedachten over de verdere uitbreidingen en

toepassingen voor de Molecule Evuator en evolutionaire algoritmen in geneesmiddelontwerp, en een aantal mogelijkheden voor toekomstige software voor het helpen ontwerpen van geneesmiddelen.