Universiteit
Leiden
The Netherlands

Transforming data into knowledge for intelligent decision-making in early drug discovery

Paricharak, S.A.

Citation
Paricharak, S. A. (2017, February 9). *Transforming data into knowledge for intelligent decision-making in early drug discovery*. Retrieved from https://hdl.handle.net/1887/45874

Version:	Not Applicable (or Unknown)
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/45874

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

Cover Page

Universiteit Leiden

The handle http://hdl.handle.net/1887/45874 holds various files of this Leiden University dissertation

Author: Paricharak, S.A.
Title: Transforming data into knowledge for intelligent decision-making in early drug discovery
Issue Date: 2017-02-09

Samenvatting

Dit proefschrift beschrijft verschillende analyses van data uit de disciplines chemie, biologie en biochemie met het doel om de efficiëntie van experimentele testen voor het ontdekken van nieuwe geneesmiddelen te verbeteren en om nieuwe inzichten te verkrijgen in het mechanisme van bioactiviteit van moleculen. De recente toename in de hoeveelheid publieke data heeft geleid tot nieuwe kansen op het gebied van bioactiviteitmodellering, en de rol die de cheminformatics en de bioinformatics hierin spelen wordt toegelicht.

Hoofdstuk één beschrijft het belang van de computational drug discovery. De principes van bioactiviteitmodellering worden in detail uitgelegd, gevolgd door een inleiding tot de onderwerpen die in dit proefschrift aan bod komen.

Hoofdstuk twee is een recensie over data-gedreven methodes die gebruikt worden bij het ontwerpen van molecuulsets voor experimentele testen, hit triage en bioactiviteitmodellering in high-throughput screening. High-throughput screening is een techniek die regelmatig wordt toegepast in de farmaceutische industrie om de activiteitsprofielen van molecuulsets te verkennen, met als doel de identificatie van mogelijk actieve stoffen voor verder onderzoek. De opmerkelijke vooruitgang op het gebied van bioactiviteitmodellering sinds de recente introductie van grootschalige moleculaire similarity metrics gebaseerd op bioactiviteitsprofielen wordt in het bijzonder besproken.

In *Hoofdstuk drie* wordt het belang van de chemische ruimte voor bioactiviteitmodellering onderzocht. Moleculen kunnen beschreven worden als een reeks eigenschappen (descriptor) die computers kunnen gebruiken om de gelijkenis tussen moleculen vast te stellen. Het concept van de chemische ruimte wordt vaak toegepast bij het selecteren van structureel diverse moleculen, met het doel het aantal actieve stoffen in biologische experimenten te optimaliseren. De mate waarin de descriptoren die in dit onderzoek gebruikt werden correleerden bij het vaststellen van moleculaire diversiteit wordt uitgebreid uitgelegd. Descriptoren die gebaseerd zijn op atoomtopologie vertonen grote overeenkomsten in hun beoordeling van moleculaire diversiteit, terwijl de op vorm gebaseerde descriptoren weinig overeenkomsten vertoonden met andere descriptortypen. Tenslotte werd gevonden dat de descriptor “Bayes Affinity Fingerprints”, die gebaseerd is op voorspelde bioactiviteitsprofielen van moleculen, het meest effectief is voor het selecteren van molecuulsets die divers zijn in de bioactiviteitsruimte.

In *Hoofdstuk vier* is de toepassing van computertechnieken om op systematische wijze moleculen te prioriteren beschreven, met als doel de efficiëntie van experimentele testen te verbeteren ten opzichte van high-throughput screening. De methode die in dit hoofdstuk beschreven wordt bestond uit de iteratieve selectie van moleculen die chemische en biologische gelijkenis vertoonden met de in eerdere iteraties geïdentificeerde actieve moleculen. De methode werd op retrospectieve wijze gevalideerd op Novartis high-throughput screening data en leidde tot een grote verbetering van efficiëntie in diverse assays: door slechts 1% van het aantal stoffen te testen werden consequent diverse molecuulsets teruggevonden die behoorden tot de top 0.5% qua activiteit. Het gebruik van deze methode kan mogelijk leiden tot aanzienlijke besparingen in tijd en middelen.
In *Hoofdstuk vijf* wordt de constructie van een "informer compound set" met behulp van data-gedreven methodes besproken. Deze informer set verschaft de meeste informatie over welke ongeteste stoffen uit een grote molecuulset het beste getest kunnen worden in een volgende ronde, ongeacht het biologische target. Het concept van *active learning* werd gebruikt voor de afleiding van de informer set, waardoor de hoeveelheid informatie van de set wordt vergroot. Een retrospectieve validatie van de informer set werd uitgevoerd op publieke high-throughput screening data, en een verbetering in vroege herkenning van actieve stoffen werd bereikt in 38 van de 46 assays.
Hoofdstuk zes beschrijft een case study over de analyse van het mechanisme van bioactiviteit van mogelijk actieve stoffen tegen malaria, welke ontdekt werden in fenotypische testen bij GlaxoSmithKline. De toepassing van twee machine learning methodes (Bayesian target prediction en proteochemometrics modeling) voor de gelijktijdige voorspelling van polyfarmacologie en affiniteit wordt beschreven. Het target prediction algoritme identificeerde 534 dihydrofolate reductase inhibitoren, terwijl de proteochemometrics modeling techniek 25 inhibitoren identificeerde, met een overlap van 23 moleculen tussen beide methoden.
Tenslotte worden er in *Hoofdstuk zeven* algemene conclusies getrokken uit het onderzoek dat in dit proefschrift beschreven wordt gevolgd door mijn toekomstperspectief over het vroege stadium van geneesmiddelenonderzoek in de academie en de farmaceutische industrie.