



Computacionele Chemie in Geneesmiddelenontwikkeling

Kenneth Goossens

Kenneth Goossens



Academische Achtergrond

- Master Farmaceutische wetenschappen
- PhD Computatieve Chemie (Universiteit Antwerpen)
- PostDoc Universiteit Antwerpen & Janssen Pharmaceutica



Professionele Ervaring

- Onderzoeker Galapagos (3 jaar)
- Onderzoeker SandboxAQ (sinds 2025)



Publicaties & Patenten

- 13 Wetenschappelijke papers
- 2 Verleende patenten
- 2 Lopende patentaanvragen



Inhoud

1

Introductie

2

Kernmethodes in computationele chemie

Kwantummechanische methoden

Moleculaire mechanica

Virtueel screenen

Machine learning & AI

3

Case studies

4

Conclusie

Voor wie is computationele chemie?

Ideale achtergrond

Studenten met interesse in **chemie, biochemie, farmacie (en computers)**. Interdisciplinaire carrière waar **wetenschap en technologie samenkomen** en je vaak in contact komt met experimentalisten.

Creatieve denkers

Je houdt van **programmeren en modelleren (maar voorkennis is geen vereiste)**. Het vermogen om abstract te denken, complexe systemen te visualiseren **en probleemoplossend te denken** is essentieel. Je spendeert het merendeel van je tijd aan het **analyseren** van simulaties en het **begrijpen** van een eiwit en hoe het zich gedraagt, en elk project is een nieuwe puzzel.

Teamspeler met impact

Je wilt bijdragen aan **snellere en efficiëntere geneesmiddelenontwikkeling**. Als computationele chemist heb je snel een impactvolle rol in projectteams en is **communicatie en een goede vertrouwensband** met projectleiders essentieel.

Niet geschikt voor...

Wie **experimenteel werk in het labo** wil doen of niet van een bureaujob houdt. Indien je **niet met onzekerheid om kan** - modellen blijven altijd benaderingen en hebben experimentele validatie nodig. Deze functie vereist ook **zelfvertrouwen en overtuigingskracht**, gezien je andere teams moet overtuigen om experimenten uit te voeren.



Wat doet de computationele chemist?

Over het algemeen zijn er sterk uiteenlopende profielen die onder "computationele chemie" vallen. Er is **geen objectieve definitie** van verschillende rollen gezien dit vaak **afhankelijk is van de structuur van het bedrijf**. Vaak vult een computationele chemist ook een **combinatie van taken uit die onder verschillende profielen vallen**.



Moleculair Modeller

Ontwerpt en optimaliseert potentiële geneesmiddelen door verscheidene tools toe te passen in de context van onderzoeksprojecten. Sterke focus op structuur-gebaseerde methoden en **werkt meer in 3D**. Nauwe samenwerking met **medicinale chemisten en biochemisten**.



Cheminformaticus

Selecteert en beheert datasets, ontwikkelt, workflows en methodes voor het computationele team. Past 2D-gebaseerde modellen toe in onderzoeksprojecten. Doorgaans een sterke focus op ligand-gebaseerde methoden en **werkt meer in 2D**. Nauwe samenwerking met **medicinale chemisten en ADME-PK**.



Data Scientist

Analyseert en visualiseert datasets. **Ontwikkelt voorspellende modellen** (machine learning/AI) en **past deze toe** om voor het voorspellen van activiteit, veiligheid van moleculen, ADME eigenschappen, en meer. Nauwe samenwerking met **ADME-PK**.



Bioinformaticus

Analyseert biologische data (genomica, proteomica, transcriptomica) en onderzoekt biologische pathways om inzichten te krijgen in ziekten. De bioinformaticus speelt een **belangrijke rol bij het identificeren van nieuwe targets** en het opstellen van **in vivo** experimenten (identificeren van biomarkers). Nauwe samenwerking met **biologen**.

Rol van computationele chemie in de industrie

Computationele chemie maakt geneesmiddelontwikkeling sneller en efficiënter



**Snellere en gerichtere
ideeënontwikkeling**



**Lagere kosten om
medicijn op de markt
te brengen**



**Biochemie, biologie
en medicinale chemie
winnen aan inzicht**



**Betere kansen op
veilige klinische trials**



Versnelling van medicijnontwikkeling

Simulaties stellen onderzoekers in staat om nieuwe medicijnen **jaren sneller** te ontwikkelen door virtueel te testen voordat synthese in het labo plaatsvindt.



Kostenreductie

Vermindert kosten en tijd van laboratoriumexperimenten door **voorspellingen op basis van modellen**, waardoor alleen de meest veelbelovende kandidaten worden gesynthetiseerd en getest.



Moleculair inzicht

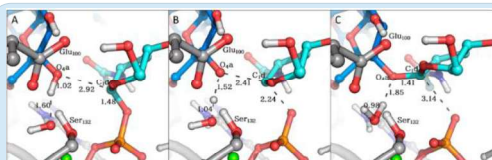
Helpt bij het begrijpen van moleculaire interacties op **atomair niveau**, wat helpt bij het gericht optimaliseren van moleculen.



Interdisciplinaire integratie

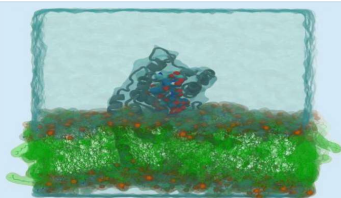
Samenwerking met andere disciplines: intelligent design van biochemische experimenten, bepalen van ideale cellijnen en proefdieren, omics voor het verwerven van inzichten in ziekten, ...

Moderne pijlers van computationele chemie



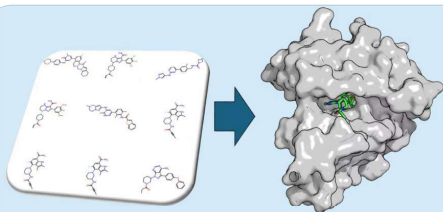
Kwantummechanische methoden

Ab initio en Density Functional Theory (DFT) bieden nauwkeurige berekeningen van **elektronische structuren** (subatomair niveau) en moleculaire eigenschappen. Worden gebruikt voor **nauwkeurige energieberekeningen**.



Moleculaire mechanica & dynamica

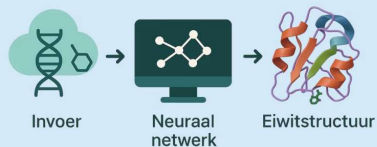
Snelle simulaties van **grotere biomoleculen** en hun gedrag over tijd aan de hand van moleculaire mechanica. Essentieel voor het **begrijpen van eiwitvouwing en binden van liganden**.



Virtueel screenen (docking & ligand-based)

Voorspellen hoe moleculen aan eiwitten binden a.h.v. het **screenen van miljoenen/miljarden moleculen**. Dit vermindert het aantal moleculen dat in het labo getest moet worden, en verhoogt experimentele hit rates.

AlphaFold3



Machine learning & AI

Traditionele **machine learning** en modernere **deep learning-algoritmen** revolutioneren moleculaire modellering en worden vaak geïntegreerd in of **gecombineerd met andere pijlers** in complexe workflows.

Kernmethodes onder de loep

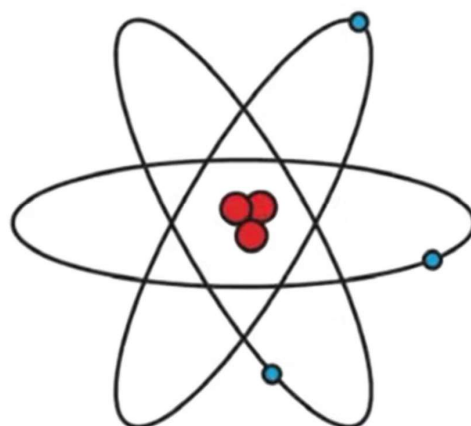
Kwantummechanische (QM) methoden

QM methoden zijn essentieel voor het begrijpen van moleculaire structuren en reactiviteit op **subatomair niveau**, en bieden de **meest nauwkeurige voorspellingen** in de computationele chemie.

Door de **benodigde computationele kracht** zijn toepassingen voor eiwitten relatief beperkt.

Toepassingen:

- Voorspellen en kwantificeren van reactiemechanismen
- Torsieprofielen berekenen
- Moleculaire eigenschappen berekenen (elektronendichtheid, pKa, energetische eigenschappen, ...)

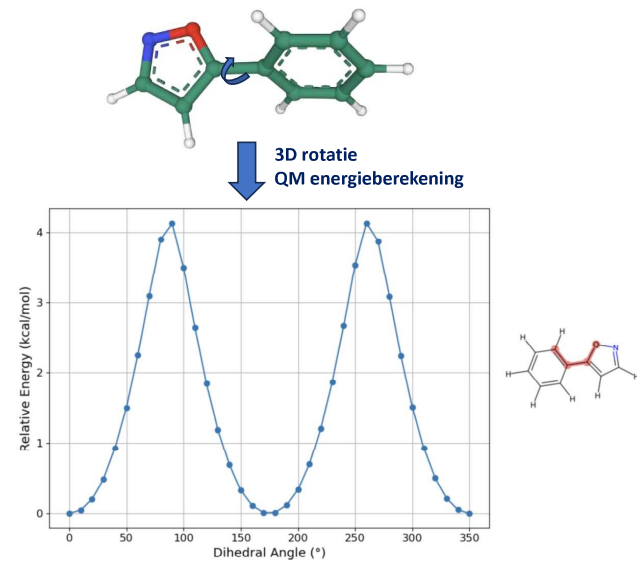
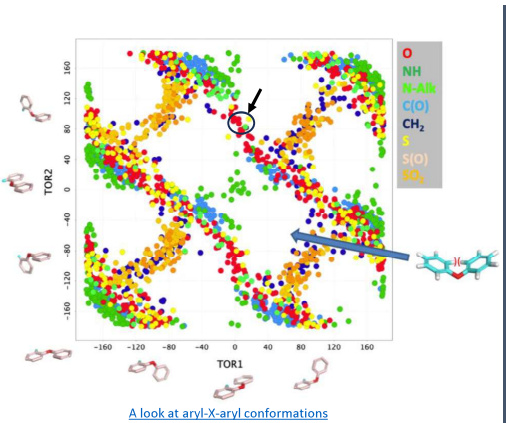
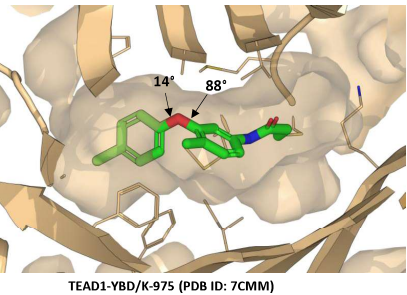


QM Methoden (kwantummechanische methoden)

Torsieprofielen berekenen

Torsieprofielen van moleculen zijn vaak belangrijk voor het begrijpen van **bioactieve conformaties in ligand-eiwitcomplexen** en hoe het mogelijk is om de **affiniteit te moduleren**.

Algemene torsieprofielen zijn af te leiden uit statistische analyse van kristalstructuren, maar in specifieke complexen kunnen **QM torsiescans worden gebruikt om energieprofielen van bindingen te bepalen**. Deze profielen kunnen dan worden vergeleken met de bioactieve conformatie, wat op zijn beurt gebruikt kan worden om modificaties voor te stellen.



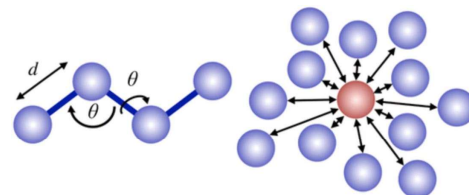
Kernmethodes onder de loep

Moleculaire dynamica: gewone simulaties

Moleculaire Dynamica (MD) simuleert de beweging van atomen en moleculen in een systeem over tijd. Door **Newton's bewegingswetten** toe te passen kunnen we het **dynamische gedrag van biomoleculen op atomair niveau** observeren.

$$\mathbf{r}(t) = \mathbf{r}_0 + \mathbf{v}_0 t + \frac{1}{2} \mathbf{a} t^2$$

$$\mathbf{F}_i = -\nabla_i U(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) \quad m_i \frac{d^2 \mathbf{r}_i}{dt^2} = \mathbf{F}_i$$



Eiwitdynamiek

Hoe bewegen eiwitten en veranderen ze van vorm? Is dit te koppelen aan functie?



Ligandbinding

Visualiseer en kwantificeer hoe liganden binden aan eiwitten.
Kunnen we dit koppelen aan werking van moleculen?



Conformationele Veranderingen

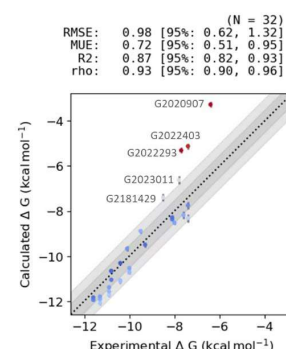
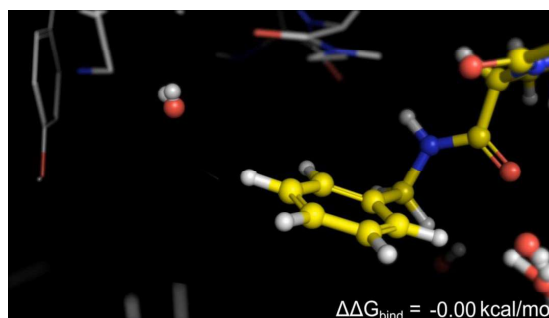
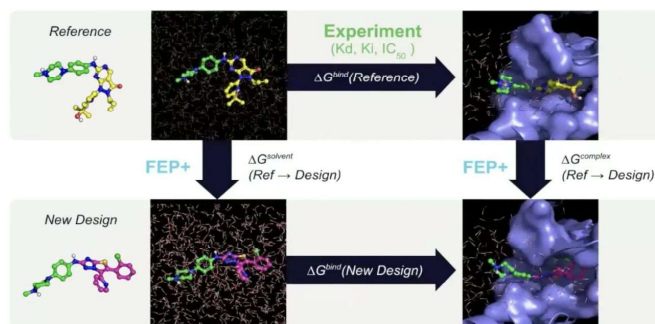
Onderzoek de flexibiliteit van moleculen.
Wat kunnen we leren uit de stabiliteit van een eiwit/complex?

Kernmethodes onder de loep

Moleculaire dynamica: vrije-energie perturbatie

Vrije-energie perturbatie (FEP) methoden zijn geavanceerde technieken die de vrije energievverschillen tussen systemen berekenen. Deze methoden worden frequent gebruikt voor het **voorspellen van relatieve bindingsaffiniteiten** van liganden en zijn een **sleutelfactor in het optimaliseren van moleculen**. Deze berekeningen zijn relatief veeleisend, en kunnen meestal maar op **beperkte schaal** worden toegepast (100-1000 moleculen).

- **Nauwkeurige Bindingsaffiniteiten:** FEP biedt nauwkeurige voorspellingen van ligand-eiwit bindingsenergieën, wat essentieel is voor het selecteren van de meest actieve moleculen.
- **Rationeel ontwerp:** Het maakt rationeel medicijndontwerp mogelijk door te begrijpen hoe kleine wijzigingen in een molecuul de bindingsaffiniteit beïnvloeden en beter te begrijpen waarom bepaalde delen van een molecule essentieel zijn voor activiteit.
- **Toepassing in virtueel screenen:** FEP wordt vaak toegepast in workflows tijdens virtueel screenen als één van de laatste stappen in de selectieprocedure.
- **Retrospectieve validatie van experimenten:** Door FEP berekeningen te vergelijken met experimentele data kunnen we de methode valideren.

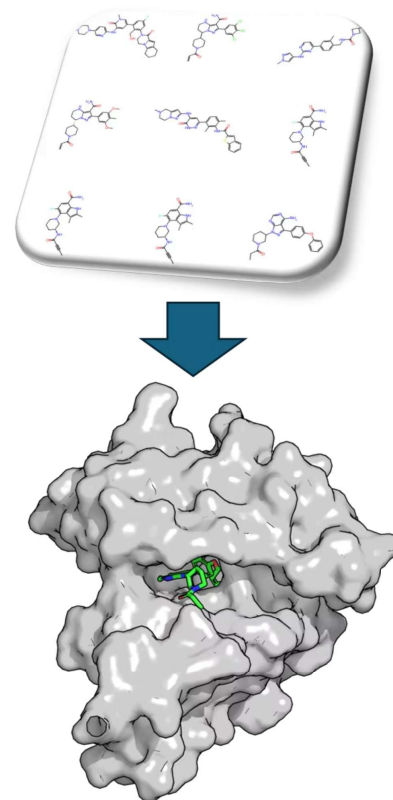


Kernmethodes onder de loep

Structure-based virtueel screenen: Docking

Moleculaire docking is een computationele methode die de optimale **oriëntatie van een ligand in een receptor voorspelt** en een score geeft aan elk ligand. Dit proces is cruciaal voor het begrijpen van moleculaire interacties en speelt een **sleutelrol in de vroege fase van geneesmiddelenontwikkeling**. Omdat dit een snelle methode is, wordt deze vaak op **grote schaal** toegepast (100,000+ moleculen)

- **Voorspelling van binding poses:** Voorspelt nauwkeurig de driedimensionale positie en oriëntatie van een ligand binnen een eiwit, wat essentieel is voor het ontwerpen van moleculen met specifieke interacties.
- **Scoring:** Kwantificeert de interactie tussen het ligand en eiwit met behulp van diverse scoringsfuncties. Deze kunnen dienen als initiële filter, maar zijn meestal niet nauwkeurig genoeg voor verder toepassing.
- **Toepassing in virtueel screenen:** Docking wordt vaak toegepast in workflows tijdens virtueel screenen als één van de eerste stappen gezien de hoge throughput.



Machine learning (ML) & artificiële intelligentie (AI)

Zeer omvangrijk domein binnen geneesmiddelontwikkeling

In de laatste 5-10 jaar is het domein van ML/AI uitgegroeid tot een **onmisbare pijler binnen de industrie**. De **meeste biotech- en farmabedrijven** wijden een deel van hun computationele team toe aan het **onderzoeken, ontwikkelen, integreren en toepassen** van ML/AI.

ML/AI heeft zijn plaats gevonden in zowel 2D- als 3D-gebaseerde methoden

Hoewel de AI-boom is begonnen met voorspellende modellen voor **ADME eigenschappen en activiteit** van moleculen, rijken de toepassingen vandaag veel verder, tot het **voorspellen van 3D structuren, 3D-gebaseerde activiteitsmodellen**, het **opschalen** van **traditionele computationele methoden** zoals docking, MD en FEP, het genereren van nieuwe moleculen en meer.

Combineren van AI met traditionele computationele chemie is de sleutel tot succes

Bedrijven zoals BenevolentAI en Exscientia zijn bekend geworden door hun **"AI-first/only" aanpak**, waarmee een aantal **kritische problemen** aan het licht zijn gekomen. **Vandaag** de dag moet AI nog steeds gezien worden als een **supplement voor traditionele methoden en geen vervanging**. Het is dan ook van belang dat deze methoden worden gebruikt door wetenschappers die de juiste achtergrond hebben om de **resultaten correct te interpreteren**.

Machine learning (ML) & artificiële intelligentie (AI)

Voorbeelden van toepassingen

ADME-T modellen

Essentieel in **gevorderde projecten** om te waarborgen dat moleculen **opneembaar en niet toxisch** zijn in het lichaam. Vaak getraind op interne data en bijna **standaard toegepast** alvorens een molecule gemaakt wordt in het labo.

Activiteitsmodellen

Hoewel een universeel model het doel is, zijn ze vaak **doelwitspecifiek**. Deze modellen maken **snelle voorspellingen** mogelijk op enorme datasets.

Generatieve modellen

Deep-learning modellen die **nieuwe moleculen genereren** die **niet in databanken bestaan**, wat een enorme hoeveelheid relevante moleculen biedt. Er moet geen rekening gehouden worden met chemie zowel een voor- als nadeel kan zijn.

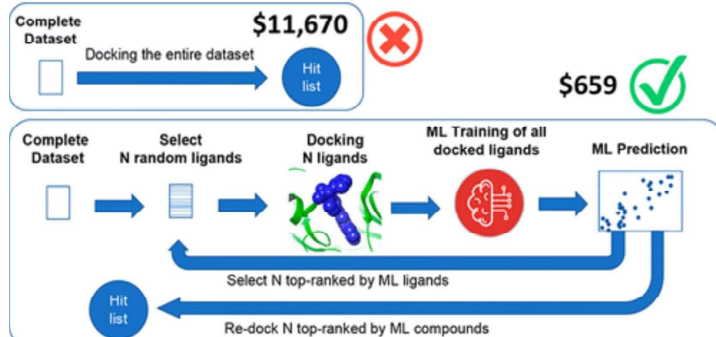
AlphaFold

Een deep learning model dat de **3D-structuur van eiwitten voorspelt** zonder structurele data. Cruciaal voor het **vinden van nieuwe binding sites**.



Active learning

Trainen van een "snel" deep-learning model dat **voorspellingen maakt** over wat een **tragere methode** zou voorspellen. Maakt **computationeel intensieve methoden toegankelijk** voor grote datasets.



Case studies: computationele chemie in de praktijk gebracht

1

Ontwikkeling van mode-of-action hypothese uit de literatuur en toepassing van virtueel screenen op een zeer uitdagend doelwit

2

Ontwikkeling van een structure-based model voor een target zonder structurele data voor het sturen van een vergevorderd project

3

Onsuccesvolle toepassing van scaffold hopping om te ontsnappen uit een IP-competitieve omgeving

Case study 1

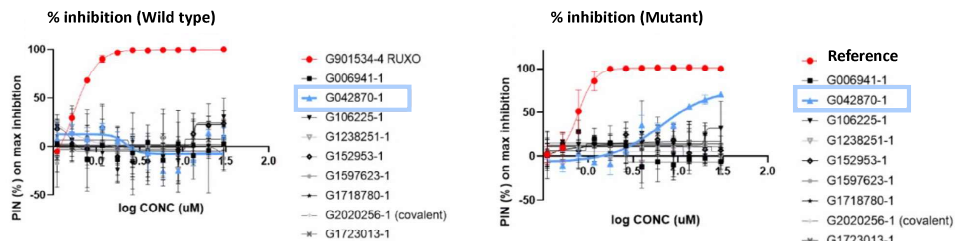
Ontwikkeling van mode-of-action hypothese uit de literatuur en
toepassing van virtueel screenen op een zeer uitdagend doelwit

Case study 1

- **Doelwit:** Domein van een eiwit met regulerend effect op ander domein. Mutatie in dit domein (ver van de natuurlijke binding site) speelt een belangrijke rol in kanker. **Selectieve inhibitie voor de mutant is een vereiste.**
- **Aanvang project**
 - **Geen bestaande selectieve inhibitoren.** Enkele publicaties over niet-selectieve inhibitoren.
 - **Literatuurstudie:** Thermodynamische en mutagene experimenten die inzicht geven over werking.
 - **Literatuurstudie:** Moleculaire dynamicastudies die inzicht geven over werking van autoinhibitie.
 - **Discussie met biologen:** Werkingsmechanisme voorgesteld en uitgewerkt. Welke studies zijn mogelijk om hypothese te bevestigen?
 - **Discussie met biochemisten:** Welke assays kunnen we gebruiken voor screening op dit werkingsmechanisme?
- **Virtueel screenen:** Moleculaire docking van interne verzameling compounds (>600k) en selectie op basis van score, diversiteit en compatibiliteit met hypothese: 144 compounds getest in biochemisch en cellulair assays.

- **G042870 has demonstrated (low potency) activity on HEL (Mutant) without affecting THP-1 (wild type)**

=> Validation of mutant selective activity in isogenic settings needed to confirm data (THP-1 wt vs THP-1 V617F).



Case study 1

- **Hit expansie:** Na identificatie van initiële hit: zoeken naar gelijkaardige compounds
 - Structuur-activiteitsrelatie vastleggen
 - Bevestigen van binding mode
 - Identificeren van actievere compounds
- **Ontdekking van potente, selectieve compounds**
 - ~350 gelijkaardige compounds getest
 - Meest actieve compound met ~100nM IC50 in cellulaire assay en zeer selectief t.o.v. mutant
- **Problemen met verdere ontwikkeling**
 - Complex gedrag in celassays (activiteit afhankelijk van expressie van andere eiwitten)
 - 100% Emax niet bereikt (partieel antagonisme)
 - Inconsistente readout in biochemische assays
 - Kristallografie niet succesvol



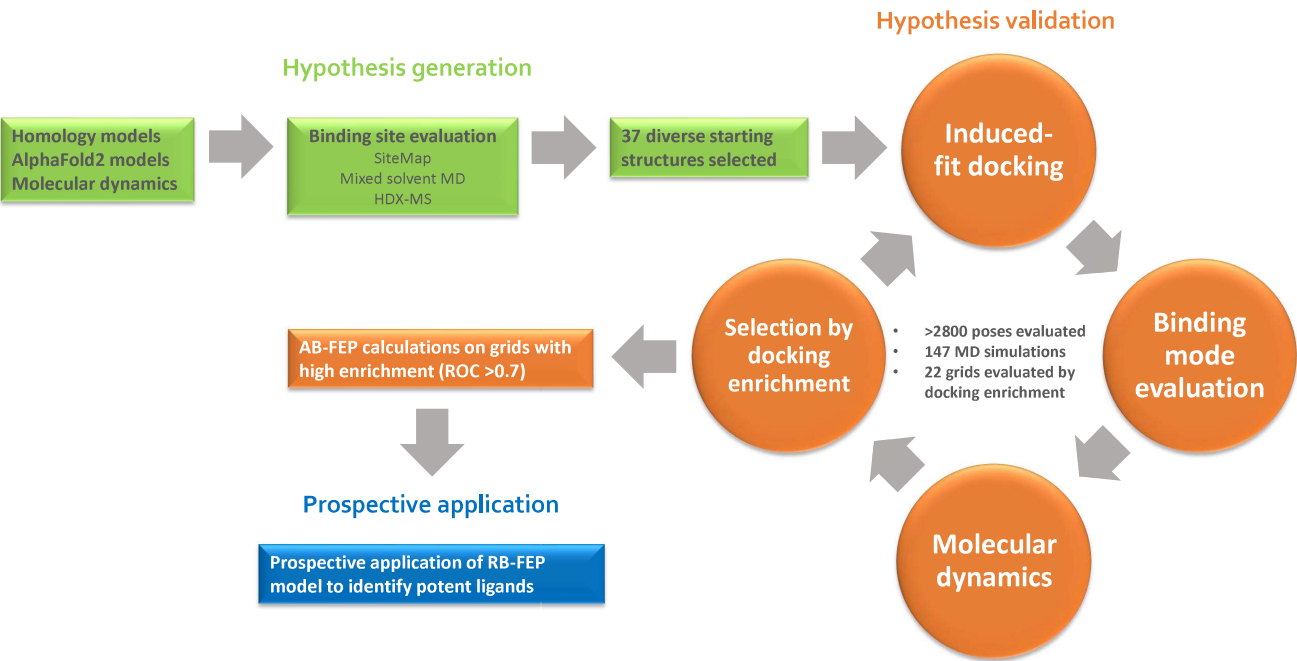
Case study 2

Ontwikkeling van een structure-based model voor een target zonder structurele data voor het sturen van een vergevorderd project

Case study 2

- **Doelwit:** Doelwit in oncologie. Twee verschillende scaffolds gekend uit de literatuur & intern onderzoek.
- **Aanvang project:** Verschillende methoden gebruikt
 - Bioisosteren & reactie enumeratie
 - Ligand-based virtueel screenen
 - ML-based modellen om activiteit te voorspellen
 - Hoe kan dit project worden ondersteund door structure-based drug design?

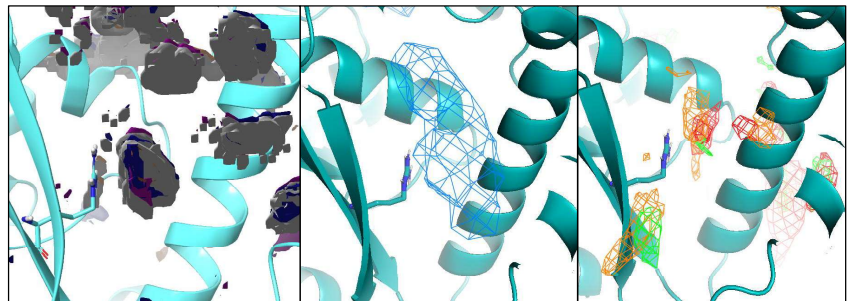
Case study 2



Case study 2

Binding site identification: Consensus method

- Experimental
 - Mutagenesis study ✓
 - HDX-MS ✓
- Empirical
 - SiteMap
- Physics-based
 - **Mixed-solvent MD (custom fragments)** ✓
 - **FragmentScout** ✓
- Deep learning
 - DiffDock
 - **DeepSite** ✓
 - Pocketminer ✓
 - Boltz-1*



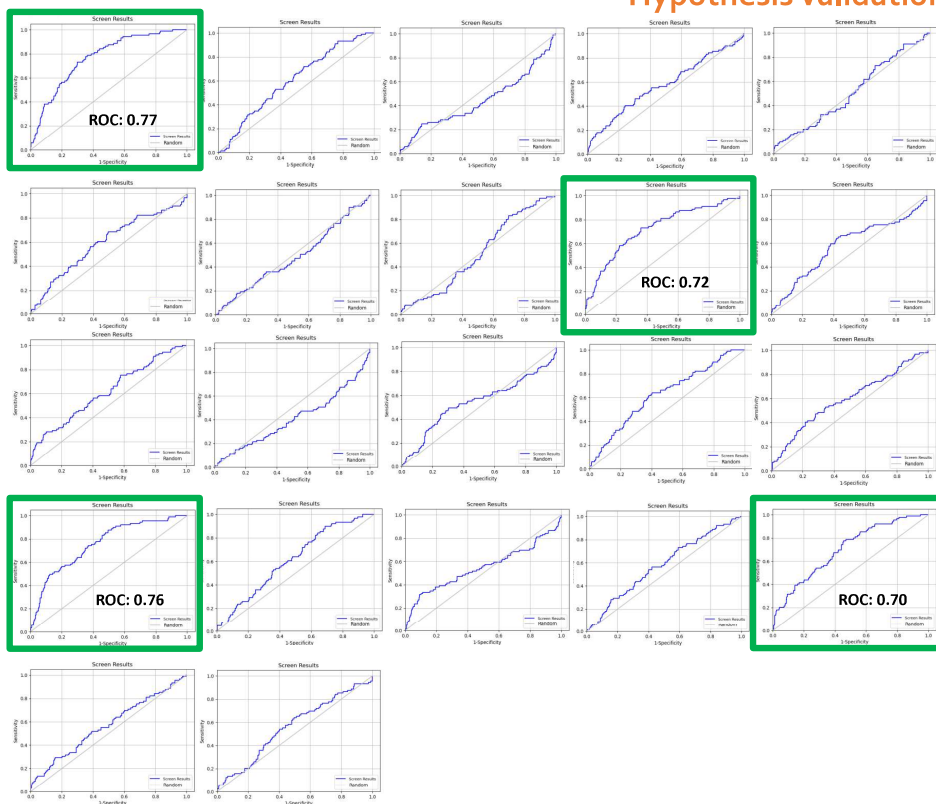
Consensus druggable region between MxMD (left), DeepSite (middle) and FragmentScout (right).

Case study 2

binding mode validation by physics-based approaches

- H2L stage: **~450 compounds with biochemical activity data** ranging from active to inactive
- Can we **use retrospective docking as a validation** for protein conformation?
- **22 binding site conformations tested**, using the 450 available data points for retrospective docking

Hypothesis validation



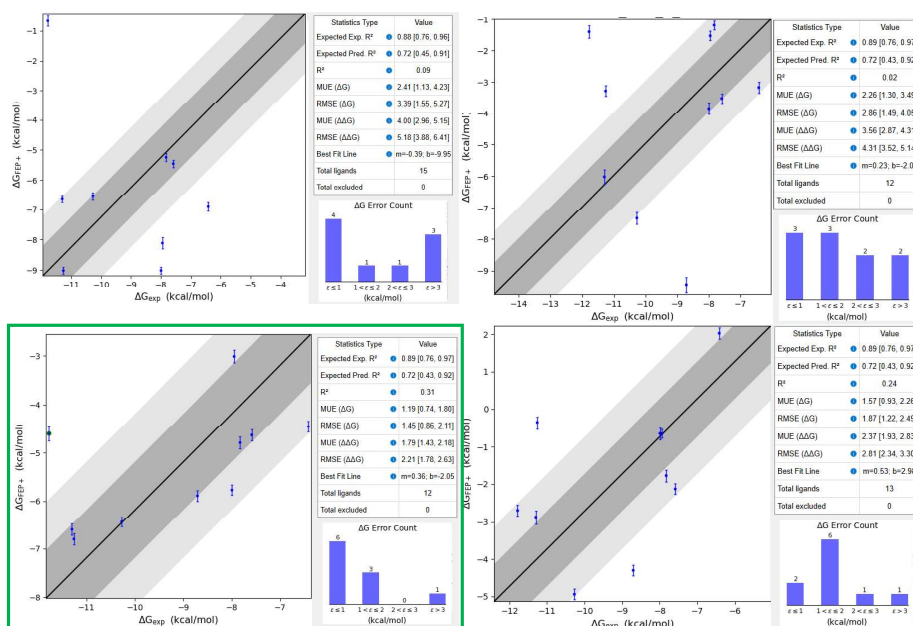
Using AUROC as a metric for actives recovered allows quick identification of promising protein conformations.

Hypothesis validation

Case study 2

binding mode validation by physics-based approaches

- H2L stage: **~450 compounds with biochemical activity data** ranging from active to inactive
- Can we **use retrospective docking as a validation** for protein conformation?
- For a more sensible estimation of reliability in later project stages: **FEP calculations**
 - Binding assay requested for 10 diverse compounds
 - Top 4 grids assessed with FEP



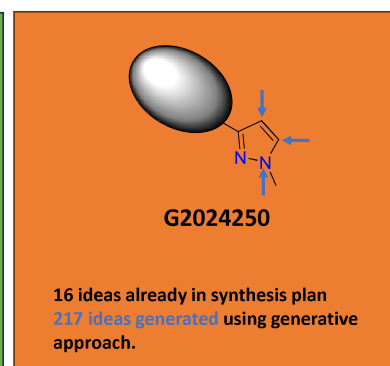
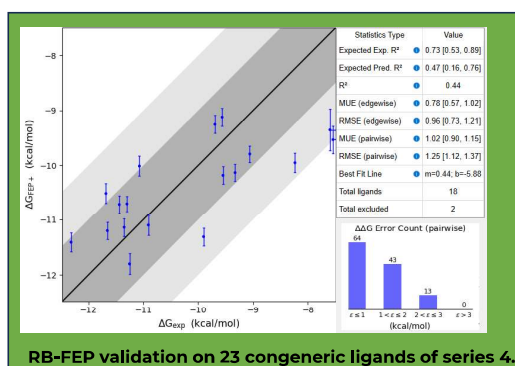
Bottom left grid shows potential for prospective use, with good rank correlation and acceptable MUE/RMSE.

Case study 2

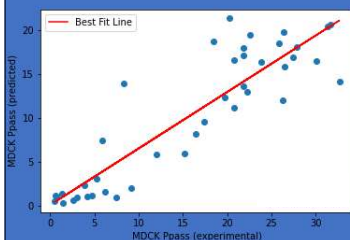
binding mode validation by physics-based approaches

- **LO stage:** lead series picked
- Multiple **generative approaches**
- **ADME model** used for selection based on permeability prediction

Prospective application



109 ideas with pred IC₅₀ ≥ G2024250
 → Evaluate ADME prediction model for key properties in lead series.



P_{passive} predictions for series 4 data points.
 19 compounds selected to propose for synthesis.

16 ideas from manual design: predicted less potent than reference, confirmed by experimental data.
 HIT RATE: 0%

4 ideas from generative approach synthesized. 1 more potent, 2 equipotent, 1 less potent than G2024250.
 HIT RATE: 75% (contribution to patent)

Case study 2

Identificatie van backup series via machine learning modellen

Case study 2

Backup series identification: ML-based activity prediction models

Goal: Identify alternative starting points (scaffolds) with more interesting ADME properties

Curation of biochemical assay data from patents, public datasets, and internal data.

Exploration of deep learning and ML activity prediction models

Prediction of full **internal compound library (>600k compounds)** for selecting compounds to test in **iterative active learning approach**

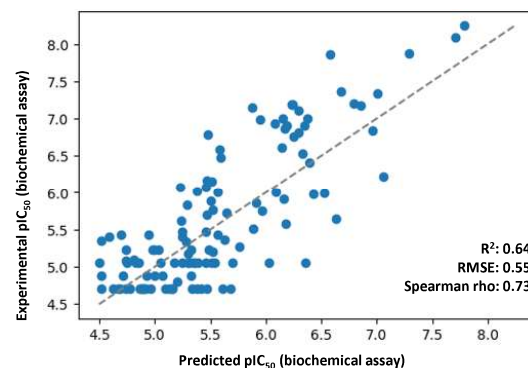
Round 1

- **138 compounds** selected and tested
- **9 active <4 μM** , 45 active <20 μM

Round 2

- **71 compounds** tested
- **12 active <20 μM**
- **Higher hit diversity $\rightarrow$ more opportunities for new starting points**

Efforts discontinued.



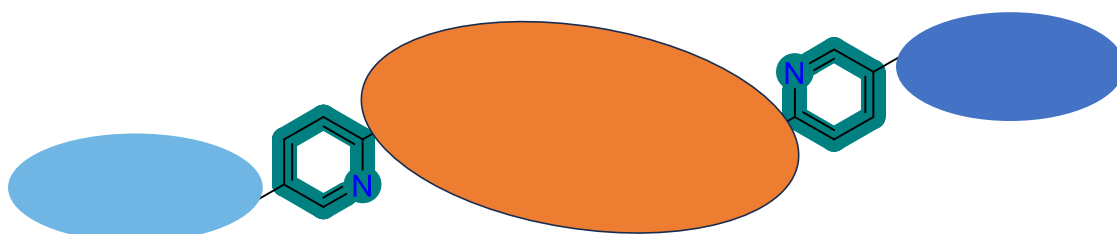
Validation of PyBoost model performance on biochemical activity prediction.

Case study 3

Toepassen van scaffold hopping om te ontsnappen uit een IP-competitieve omgeving

Case study 3

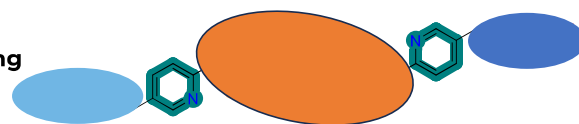
- **Doelwit:** Doelwit in oncologie. Intern onderzoek werd uitgevoerd in een heel competitief landschap: **200+ patenten** op gelijkaardige compounds. **Scaffold is zeer gevoelig aan aanpassingen**, dus moeilijke differentiatie.
- **Aanvang project:** Voornamelijk rationeel design om interne compound scaffold te onderscheiden van de competitie.
 - Kunnen we **computationalele methoden** gebruiken om **nieuwe ideeën te genereren**?
 - Methode 1: **AI** gebruiken voor **scaffold hopping**
 - Methode 2: Volledige **recombinatie** van compounds die gemaakt zijn. Zijn er zaken gemist?



Case study 3

Efficient SAR exploration on backup series

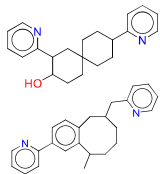
- Goal: **differentiate** from 100s of competitor patents by **replacing biphenyl core**
- Narrow cavity with no hydrogen bonding opportunities: **difficult challenge**
- **Approach: REINVENT to generate core hopping ideas**
- **Multiple iterations to generate more desirable output by tweaking scoring**



X-ray and structure of G2017580 in complex with the target. Orange: scaffold to replace. Green: attachment points to keep.

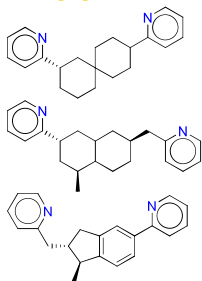
Run 1

Shape/pharmacophore scoring
No restrictions
Wild structures generated



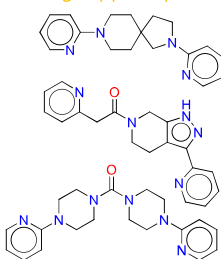
Run 2

Substructure exists in ChEMBL
Spirocycles more in line with chemist designs
Many unstable motifs
Challenging chiral centers



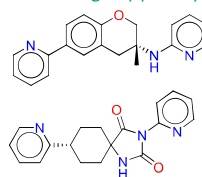
Run 3

Restrict max. ring size
Restrict number of chiral centers
Penalize longer/flexible linkers
Rigidity penalty was too high



Run 4

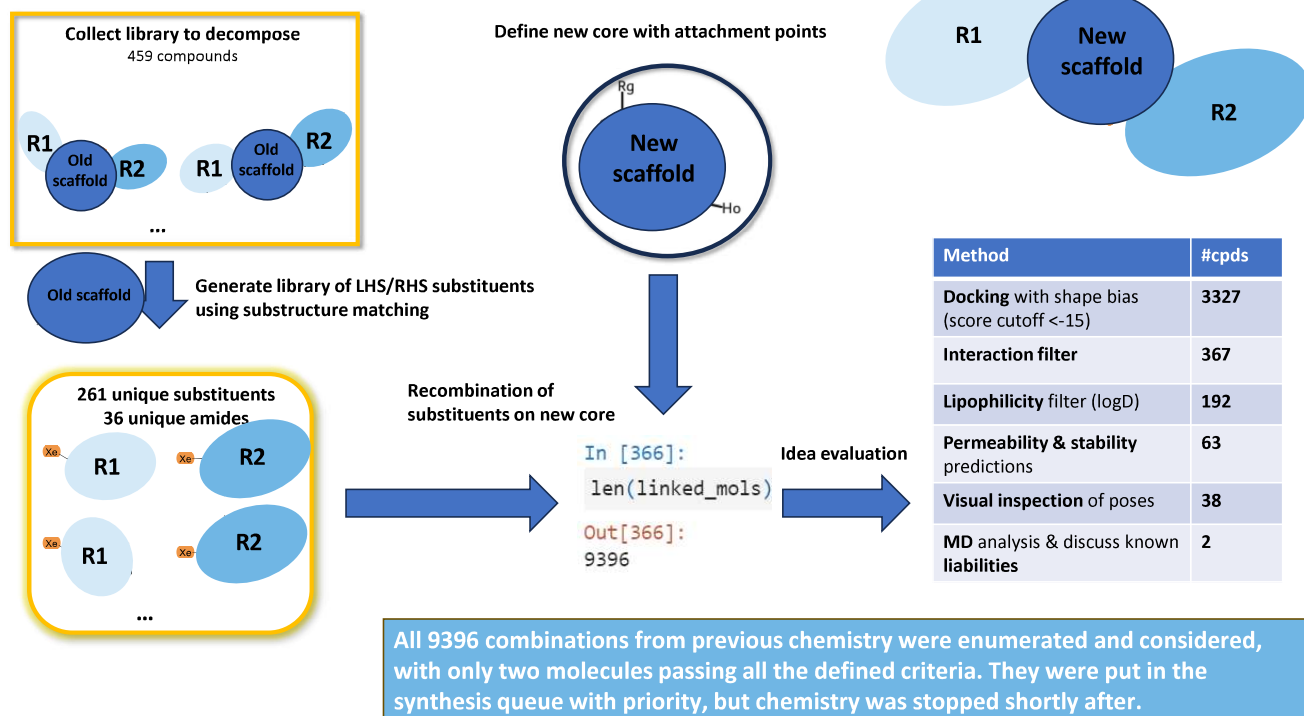
Reduce rigidity penalty



Some spirocycles were synthesized, but efforts on scaffold hopping were quickly stopped because of synthetic challenge and other priorities.

Case study 3

Efficient SAR exploration on backup series



Conclusie: Computationale Chemie in de farmaceutische industrie



Integrale Pijler

Computationale chemie is een **onmisbare pijler** die traditionele methoden en geavanceerde ML/AI combineert voor het onderzoeken en **ondersteunen van uiteenlopende onderzoeksvragen**.



Gediversifieerde Aanpak

Van kwantummechanica en moleculaire dynamica tot ligand-based screening en machine learning. Een **brede waaier aan computationale technieken** biedt oplossingen voor complexe moleculaire uitdagingen.



Versnelling & Innovatie

Computationale methoden **versnellen het identificeren van kandidaat-geneesmiddelen** door het virtueel optimaliseren van eigenschappen. Door constante innovatie blijft computationale chemie zeer relevant.