Generative AI for Drug Design: A powerful tool to support medicinal chemists

Ruel Cedeno*, Yaelle Fischer, Dhoha Triki, Philippe Schambel

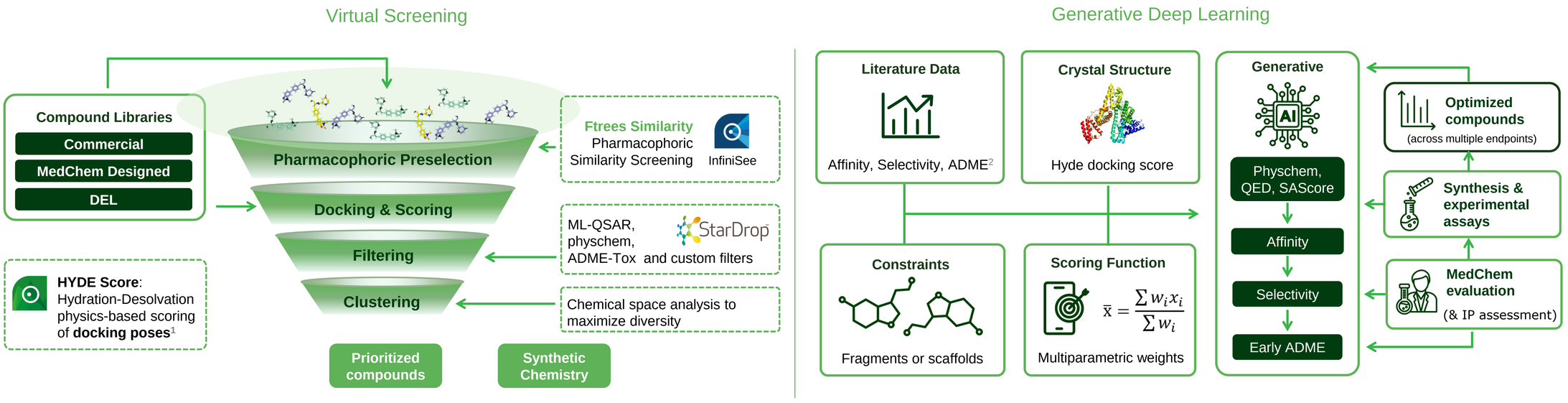
Department of Computational Chemistry, Novalix, 67000 Strasbourg, France

Abstract

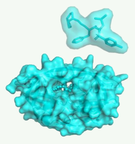
One major challenge in medicinal chemistry is to design compounds that satisfy multiple parameters simultaneously (potency, efficacy, physicochemical and ADME-PK properties, synthesizability, patentability etc). Generative artificial intelligence (AI) has emerged as a promising tool to address this with the help of *de novo* design guided by multiparameter optimization (MPO) algorithms. In this work, by combining the leading opensource packages[1-3] and proprietary softwares[4], we built an automated *in silico* workflow which takes our project data and structural constraints as input and optimized molecules as output. After performing extensive benchmark tests, we demonstrate that such integrated workflow offers outstanding performance in terms of (1) ability to generate known inhibitors in preclinic and clinical trials (2) druglikeness (3) synthesizability (4) structural diversity, using CCR5 inhibitors as a test case study.

Combining the above results, we integrated our generative AI platform in synergy with virtual screening as well as the classical structure-based and ligand-based drug design. Furthermore, with internal access to a wide range of biophysics/biochemical assays, augmented synthetic chemistry, DNA-encoded library (DEL) and PROTAC development systems which can generate the important experimental data, such an AI-integrated platform represents a powerful tool for supporting medicinal chemists in accelerating projects from hit identification to lead optimization, enabling the discovery of new medicines.

Chem-informatics workflow at Novalix Drug discovery

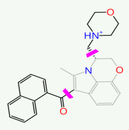


Ligandibility Analysis



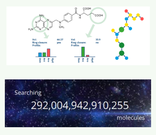
- Binding pocket detection (orthosteric & allosteric site)
- Pharmacophoric modeling via docking of reference compounds
- Homology modeling

Growing & Scaffold Hopping



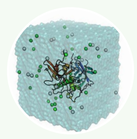
- 3D fragment library search based on pocket constraints
- Improvement of target efficacy & selectivity
- Hit expansion and IP generation

Ligand-based Drug Design



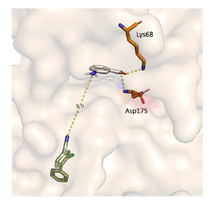
- Pharmacophoric similarity search using known actives
- Useful when target structure is unknown
- Fuzzy pharmacophores for screening billions of compounds

Molecular Dynamics Simulations



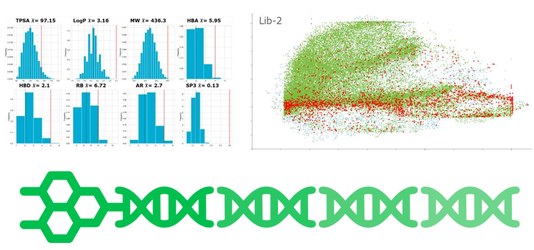
- Protein flexibility studies
- Binding affinity prediction & water thermodynamics
- Residence time assessment with enhanced sampling methods

Structure & Ligand-based Drug Design

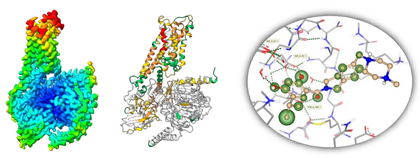


Hit Fragments
K_d > 270µM

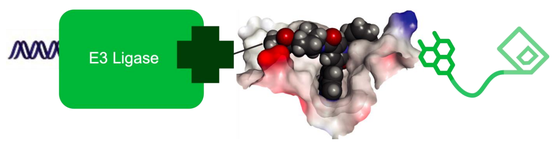
Linked Fragments
K_d = 0.3 µM



Validation of Crystallographic Structures



PROTAC design and data analysis



Support for Biophysics, Synthesis, DEL & PROTAC

Test case study using generative deep learning

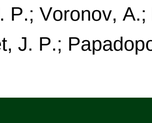
Design of CCR5 Inhibitors and Maraviroc Rediscovery

CCR5 Literature Data

ChEMBL

Known Inhibitors
(1652 with 8 in preclinic)

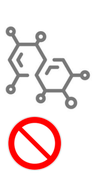
CCR5 Cocrystal



SeeSAR

Hyde Docking Score[1]

Structural Constraints



Scaffolds

Unwanted substructures

Scoring Function

RDKit, LightGBM, scikit learn, CDDD, BAYER

ML-QSAR pIC_{50} model

AstraZeneca

REINVENT 4

Generative Molecular Design[4]

Stage 1	
	480 < MW < 650
	3.5 < logP < 6.5
✓ Physchem	0.4 < Fsp^3 < 0.75
✓ Druglikeness	0.30 < QED < 0.60
✓ Synthesizability	2 < SAScore < 4
Stage 2	
✓ Affinity (Predicted pIC_{50} > 8)	
Stage 3	
✓ Docking Score (Hyde pK_i > 8)	

AI-generated
(N=4870)

Known Inhibitors
(N=1652)

Clinic
(N=8)

Criteria	AI-Generated	CHEMBL
Druglikeness QED_{med}	0.49 ± 0.16	0.38 ± 0.14
Synthesizability SA_{med}	3.20 ± 0.57	3.67 ± 0.72
% Distinct Scaffolds	49.8%	43.5%

Pfizer

Maraviroc
(approved 2007)

Ranked at **Top 1%** by **Hyde Score**

AZD5672
Top 3%

Vicriviroc
Top 3%

INCB009471
Top 3%

Conclusions & Perspectives

- ✓ We described the implementation of deep learning for molecular design and multiparameter optimization as part of the cheminformatics tools at Novalix.
- ✓ By leveraging a unique combination of leading opensource packages and proprietary softwares, we built an *in silico* workflow which demonstrated excellent performance.
- ✓ This AI-integrated workflow can be coupled with diverse research services at Novalix which are customized to the requirements of our clients.

References

1. Schneider, N.; Lange, G.; Hindle, S.; Klein, R.; Rarey, M. A consistent description of HYdrogen bond and DEhydration energies in protein–ligand complexes: methods behind the HYDE scoring function. Journal of Computer-Aided Molecular Design, 2012, 27, 15–29.
2. Fang, C.; Wang, Y.; Grater, R.; Kapadnis, S.; Black, et al. Prospective Validation of Machine Learning Algorithms for Absorption, Distribution, Metabolism, and Excretion Prediction: An Industrial Perspective. Journal of Chemical Information and Modeling, 2023, 63, 3263–3274.
3. Winter, R.; Montanari, F.; Noé, F.; Clevert, D.-A. Learning continuous and data-driven molecular descriptors by translating equivalent chemical representations. Chemical Science, 2019, 10, 1692–1701.
4. Loeffler, H. H.; He, J.; Tibo, A.; Janet, J. P.; Voronov, A.; Mervin, L. H.; et al. Reinvent 4: Modern AI–driven generative molecule design. Journal of Cheminformatics, 2024, 16.
5. Guo, J.; Knuth, F.; Margreitter, C.; Janet, J. P.; Papadopoulos, K.; Engkvist, O.; et al. Link-INVENT: generative linker design with reinforcement learning. Digital Discovery, 2023, 2, 392–408

CONTACT: info@novalix.com