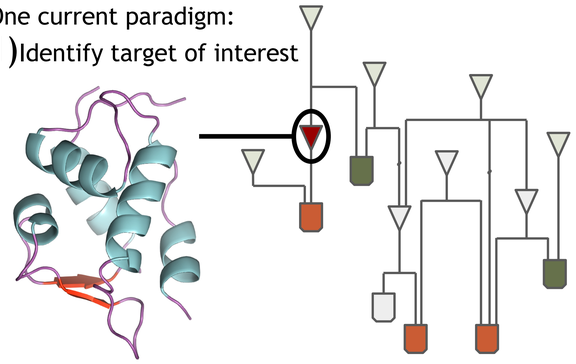
Deep Generative Modeling for Structure-guided *de novo* Drug Design

Sree Vadlamudi* Anna Kriukova, Pascaline Jacquemard, Anthony Martinez, Stéphane Sautet, Alexis Denis, Eric Boursier, Maud Jusot, Nicolas Devaux, Christopher Housseman, Nicolas Martin, Stéphanie Labouille, Yann Gaston-Mathé, Nicolas Do Huu, Quentin Perron, Yann Lamotte, Brice Hoffmann

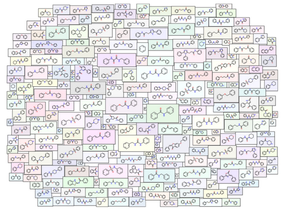
Structurally-enabled computer-aided drug design (CADD)

One current paradigm:

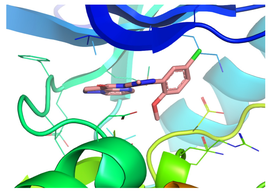
1) Identify target of interest



2) Select a virtual library



3) Screen and rank



A major drawback is that “good” compounds need to *already* be a part of your virtual library.

Our largest virtual libraries are massive! On the order of 10^{10} to 10^{11}, but these pale in comparison to the size of chemical space, estimated at 10^{60}.

Ultra-large virtual libraries: Oleksandr O. Grygorenko, Dmytro S. Radchenko, Igor Dziuba, Alexander Chuprina, Kateryna E. Gubina, Yurii S. Moroz, Generating Multibillion Chemical Space of Readily Accessible Screening Compounds, *iScience*, **2020**, 23, *11*, 101681, https://doi.org/10.1016/j.isci.2020.101681.

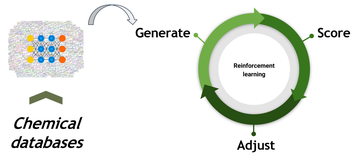
Iktos generative infrastructure

An emerging technology at the intersection of computers and chemistry is AI-based molecule generation which allows a computer to produce new virtual molecules very rapidly.

Our generative infrastructure at Iktos couples various models and assessments to score these virtual molecules with a reward-based feedback system to guide the generation towards optimal properties.

By applying our technology we are able to explore chemical space as we generate new molecules, finding regions with optimal properties and increasing the chances of the “best” virtual molecules being screened.

Generative AI | Reinforcement Learning | Predictors



- QSAR models
- 3D Simulations
  - Docking
  - 3D QSAR
- Retrosynthesis (spaya.ai)
- Generic scores

Generative AI coupled with structure-based evaluation of generated molecules

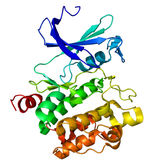
Proof-of-concept project background

Primary goals in collaboration with Oncodesign

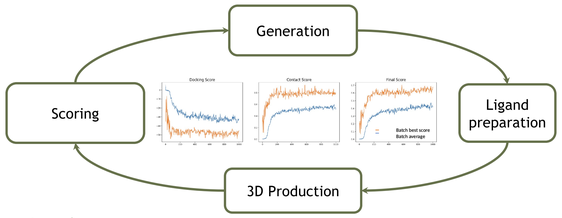
- Build the workflow for a real scenario
- Demonstrate the possibilities of the technology

Project target: Pim-1

- Proto-oncogene serine/threonine-protein kinase Pim-1
- Implicated in multiple human cancers, including prostate cancer, acute myeloid leukemia and other hematopoietic malignancies
- Clinical trial results so far have showed promising anti-cancer activity, but side effects due to insufficient selectivity have proved problematic and research continues to find more potent and selective inhibitors for this target
- Multiple PDB files with good resolution and known ligands



Methodology



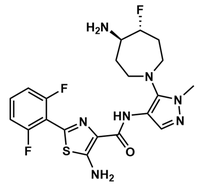
Generation protocol

- Text-based generator
- Based on Recurrent neural network (RNN) with Long Short Term Memory (LSTM)
- Trained on Chembl dataset
- Reference pocket for docking PDB:6NO9
- Docking software: UCSF Dock 6
- Reward function includes:
  - Molecular descriptors:(MW, cLogD, TPSA, # of H-bond donor and acceptor, QED, cPFI)
  - Docking score
  - Contact score
- Murcko scaffolds of all known PIM1 inhibitors forbidden during the generation

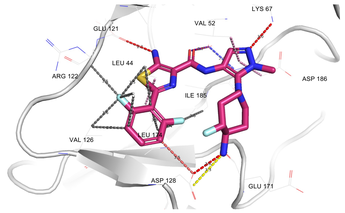
Contact Score

- Guide the generation by rewarding molecules if they form key interactions with the protein (X-ray ligand’s exiting interactions)
- Can be built from:
  - Single PDB file with co-crystallized ligand
  - Multiple PDB files with distinct ligands: frequency of interactions observe in the different PDB files
  - Molecular Dynamics simulation: frequency of interactions observe in the MD
  - Manually tuned with expert knowledge

Ligand: GDC-0339 (genentech)

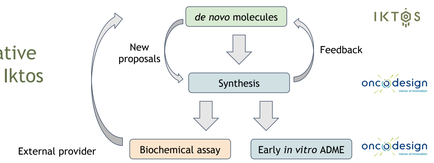


PDB entry: 6NO9



Wang X., Blackaby W., Allen V., Chan G. K. Y., Chang J. H., Chiang P. C., Diène C., Drummond J., Do S., Fan E., Harstad E. B., Hodges A., Hu H., Jia W., Kofie W., Kolesnikov A., Lyssikatos J. P., Ly J., Matteucci M., Moffat J. G., Munugalavadla V., Murray J., Nash D., Noland C. L., Del Rosario G., Ross L., Rouse C., Sharpe A., Slaga D., Sun M., Tsui V., Wallweber H., Yu S. F., Ebens A. J. Optimization of Pan-Pim Kinase Activity and Oral Bioavailability Leading to Diaminopyrazole (GDC-0339) for the Treatment of Multiple Myeloma. *J. Med. Chem.*, **2019**, 62, *4*, 2140, https://doi.org/10.1021/acs.jmedchem.8b01857

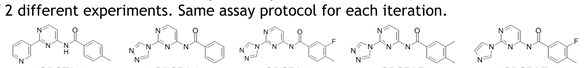
Cooperative iterative process between Iktos and Oncodesign



Results

First iteration

- Synthesis of five compounds
- Biochemical assay on 3 PIM Kinase isoforms @ 10µM and 1µM. % of inhibition (Relative to DMSO controls). Mean of 2 different experiments. Same assay protocol for each iteration.

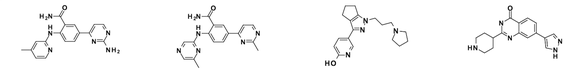


ODS736 (analog)	ODS584 (analog)	ODS586 (generative AI)	ODS585 (generative AI)	ODS587 (analog)
69% @10µM / 32% @1µM (Pim1) 88% @10µM / 22% @1µM (Pim3)	Inactive @10µM	Inactive @10µM	Inactive @10µM	45% @10µM (Pim1) 71% @10µM / 69% @1µM (Pim3)

With five molecules synthesized: from scratch ⇒ ~5µM

Second iteration

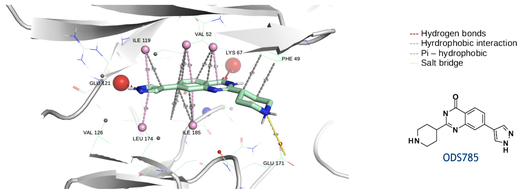
- Synthesis of 4 compounds
- Early in vitro ADME



ODS715 (generative AI)	ODS866 (generative AI)	ODS909 (analog)	ODS785 (generative AI)
72% @10µM / 37% @1µM (Pim1) 82% @10µM / 26% @1µM (Pim3) ChromlogD 1.9 Sol. PBS/Fassif 218/228 Clint r/h <5/9	Inactive @10µM ChromlogD 2.7 Sol. PBS/Fassif 235/247 Clint r/h 6.7/<5	Inactive @10µM ChromlogD 0.4 Sol. PBS/Fassif 250/250 Clint r/h 38/<5	88% @10µM / 50% @1µM (Pim1) 90% @10µM / 54% @1µM (Pim3) ChromlogD 0 Sol. PBS/Fassif 221/250 Clint r/h 54/27

With 4 molecules synthesized: from 5µM ⇒ ~1µM

Taking an in-depth look at the structure-based scoring functions:



ODS785

Contact map overview

PDB	Hydrophobic										HBA					HBD	
	Leu44	Phe49	Val52	Ala65	Ile104	Leu120	Arg122	Val126	Leu174	Ile185	Glu121	Asp128	Glu171	Asp186	Asn172	Lys67	Asp131
6NO9 (GNE)	X						X	X	X	X	X					X	
6MT0 (Amgen)	X	X		X		X		X	X	X						X	
4DTK (AZ)	X	X	X	X				X	X	X		X	X	X		X	
ODS715 - 2nd iteration	X	X	X					X	X	X	X		X	X*			
ODS785 - 2nd iteration		X	X						X		X		X	X		X	
ODS142 - 3rd iteration	X	X	X					X	X		X			X		X	

Molecular Mechanics - Generalized Born/Surface Area Rescoring

- More complex scoring functions based on a force field, including solvation energy
- Better confidence in prediction

Third iteration

- Synthesis of 3 compounds
- ** 10-points dose response IC50

ODS137 (analog)	ODS142 (generative AI)	ODS151 (analog)
Inactive @10µM ChromlogD 2.5 Sol. PBS/Fassif 94/232 Clint r/h 32/7	IC50 = 1.08 µM (Pim1) ** IC50 = 535 nM (Pim3) ** ChromlogD 2.7 Sol. PBS/Fassif 205/239 Clint r/h 103/53	31% @10mM / 26% @1µM (Pim1) 69% @10mM / 33% @1µM (Pim3) ChromlogD 0.6 Sol. PBS/Fassif 157/226 Clint r/h nd/nd

With 3 molecules synthesized: from ~1µM ⇒ ~0.5/1.0µM

Overall: 12 molecules synthesized From scratch ⇒ ~0.5/1.0µM

A successful proof of concept

- 12 molecules synthesized by Oncodesign, from 5 different scaffolds
- 4 molecules with activity < 5µM, from 3 different scaffolds
- 2 molecules with activity < 1µM, from 2 different *novel* scaffolds
- Good preliminary ADME properties (logD, solubility, clearance)
- Possibility to forbid multiple scaffolds during the generation to avoid existing patents
- Spaya synthetic access optimization during the generation
- Multi-Parametric Optimization
- Easy to create diversity around a hit