Deep Learning for *de novo* Drug Design applied to Lead Optimization

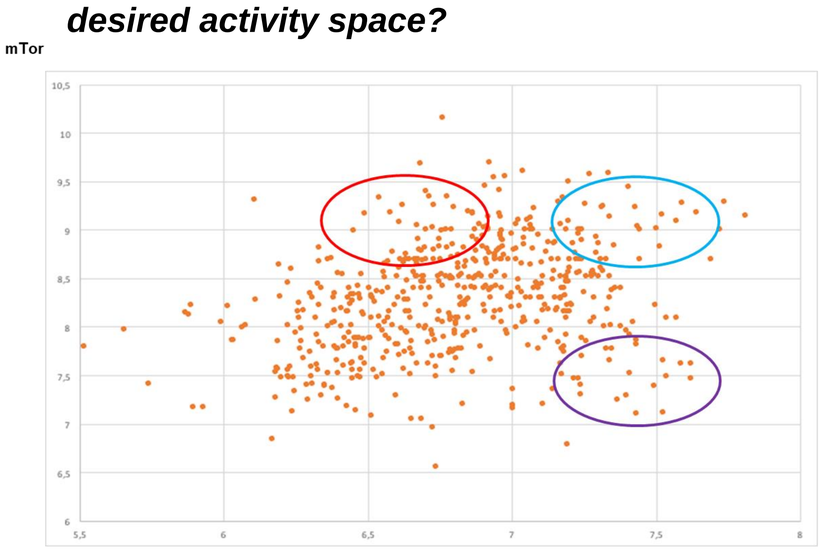
Quentin Perron, Hamza Tajmouati, Nicolas Do Huu

61 rue Blomet, 75015 Paris, quentin.perron@iktos.com

Abstract

- Multi-Parameter Optimization (MPO) is a major issue in small molecules (NCE) drug discovery projects. Recently, promising results have been reported for deep learning generative models applied to de novo molecular design, but until now, the value of this technology for addressing the MPO issue has not been evaluated.
- We present here an *in silico* proof of concept built from data set of 600 molecules measured on PI3K and mTOR[1], that shows that generative models could be efficient tools to design molecules with a predefined multi-objective activity blueprint and meeting drug-likeness criteria.
- We show that not only AI designed molecules have all the targeted characteristics, but also that they are very similar to the molecules that were actually designed by chemists.

How to automatically design molecules in the desired activity space?



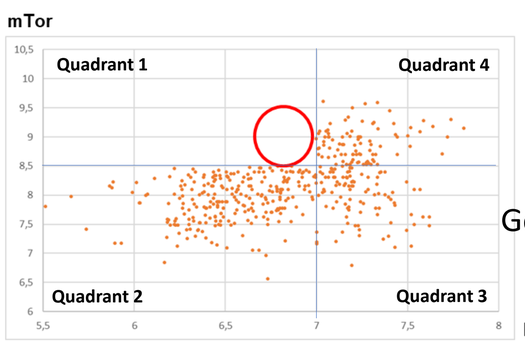
1 *Prediction and generation*

Activity space was divided into 4 quadrants.

Molecules from quadrant 1 (mTOR pKi> 8.5, PI3K pKi<7) were removed from the data set.

QSAR models were trained on the remaining data.

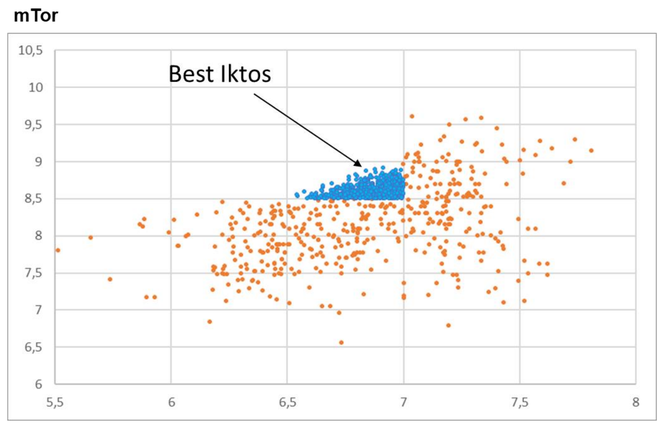
Then a deep generative model was used to generate ~700 molecules with activity predicted to be mTOR>8.5 and PI3K<7.



QSAR
+
Generative model

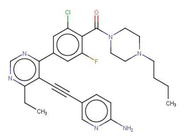
Additional objectives were similarity to the initial data set and QED (drug-likeness metric).
Two molecules generation runs were performed, the first using the Morgan fingerprint representation, the second the Pharmacophore fingerprint.

Morgan fingerprint representation



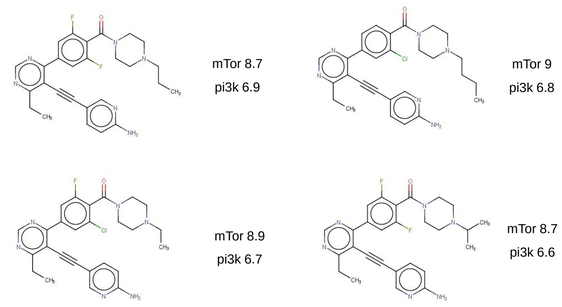
Initial data + Iktos molecules (719)

Best Iktos (predicted)



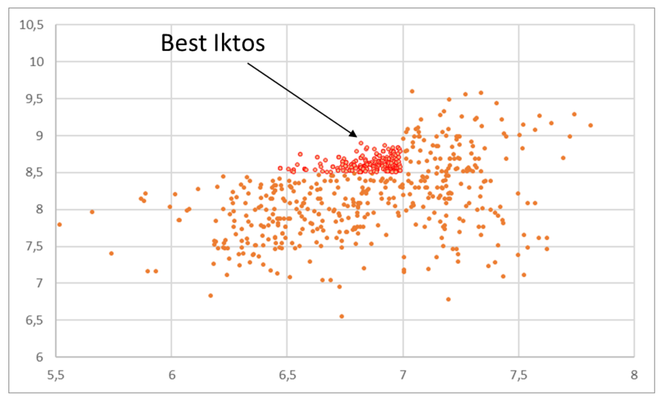
mTor 8.8
pi3k 6.8

Closest removed Molecules (measured)



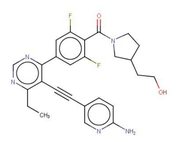
AI designed molecules were found to be highly similar to the original molecules previously removed from the data set.

Pharmacophore fingerprint representation



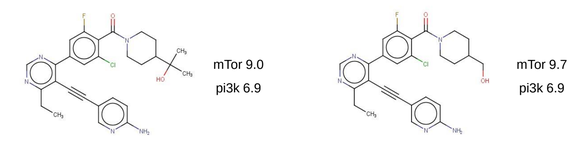
Initial data + Iktos molecules (513)

Best Iktos (predicted)

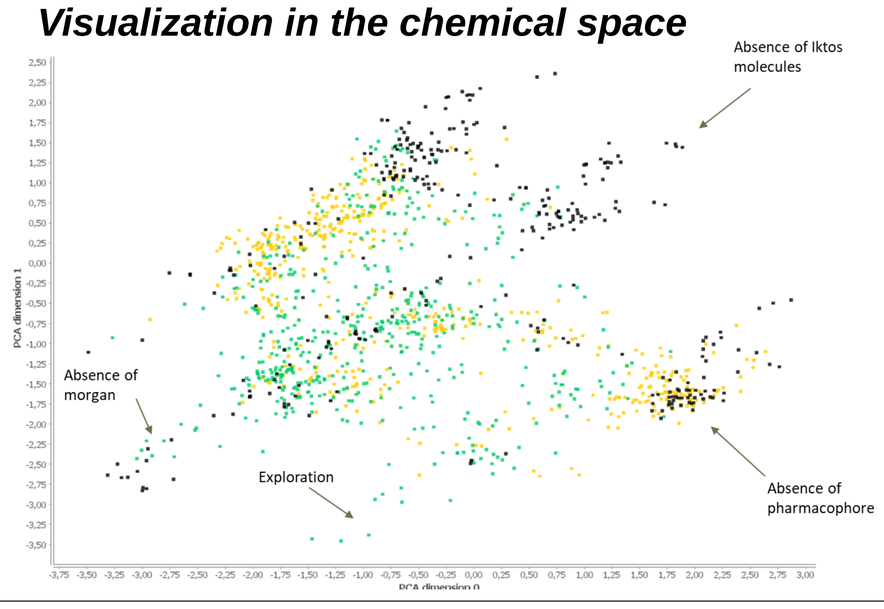


mTor 8.8
pi3k 6.6

Closest removed Molecules (measured)



2 *Visualization in the chemical space*



Initial dataset (removed molecules)

Morgan FP generation

Pharmacophore FP generation

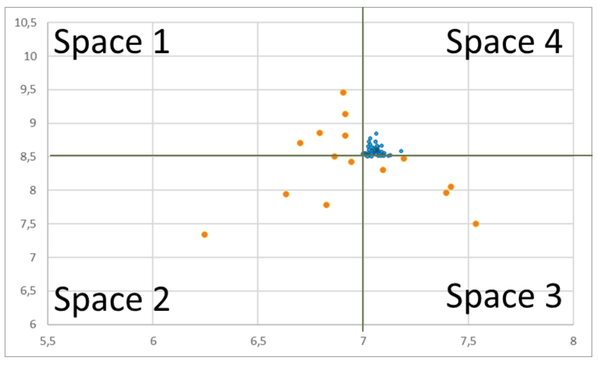
Depending on the molecular representation used for prediction and generation, different parts of the chemical space were covered by the AI designed molecules.

Generative models using either Morgan or Pharmacophore FP, appeared as complementary and capable of covering a large spectrum of optimal solutions.

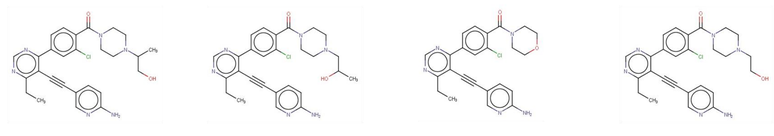
There was an overlap, but also specificities for each approach, compared to what had been explored by the chemists (compounds originally designed, synthesized and tested in the initial data set).

3 *Similar approach using a small data set*

A similar approach was tested with a smaller data set (15 molecules) extracted from the main data set.
5 molecules were drawn randomly from each quadrant 1, 2 and 3. QSAR models were built on these 15 molecules.
Then generative models were used to design molecules predicted to be in Quadrant 4.



87 molecules were generated in quadrant 4.



Predicted	Measured	Predicted	Measured	Predicted	Measured	Predicted	Measured
mTor 8.6	mTor 8.8	mTor 8.8	mTor 8.7	mTor 8.6	mTor 9.3	mTor 8.6	mTor 9.2
pi3k 7.1	pi3k 7.1	pi3k 7.1	pi3k 7.1	pi3k 7.1	Pi3k 7.2	pi3k 7.1	pi3k 7.1
1st position		3rd position		11th position		29th position	

In the first 30 propositions, 4 molecules were found to be identical to molecules previously removed.

1 – J. A. Engelman, Nature Reviews Cancer, 2009, 9, 550-562; A. Carnero, Expert Opin. Investig. Drugs, 2009, 18, 1265-1277 and P. Liu et al., Nature Reviews Drug Discovery, 2009, 8, 627-64. National Center for Biotechnology Information. PubChem BioAssay Database; AID=1257783