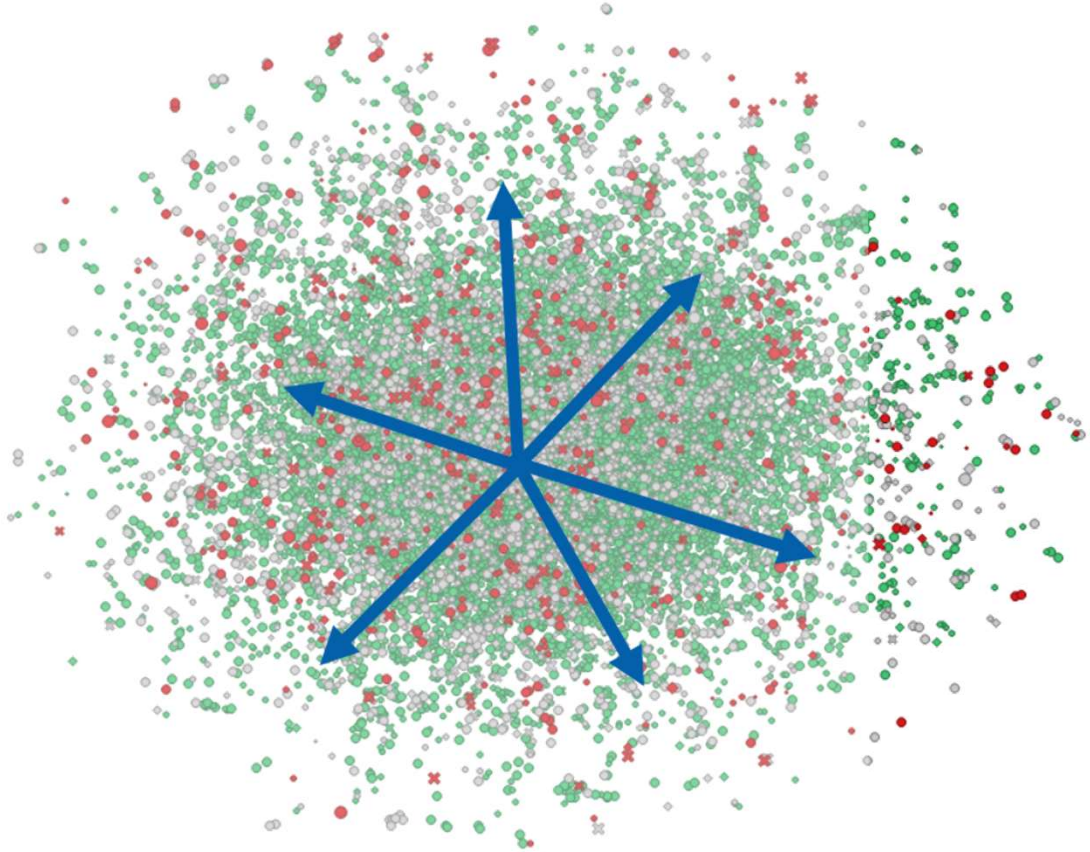
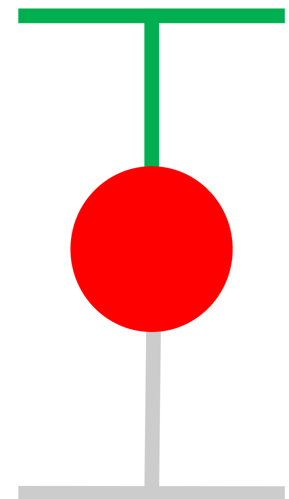
NIBR
Global Discovery Chemistry
Computer Aided Drug Discovery



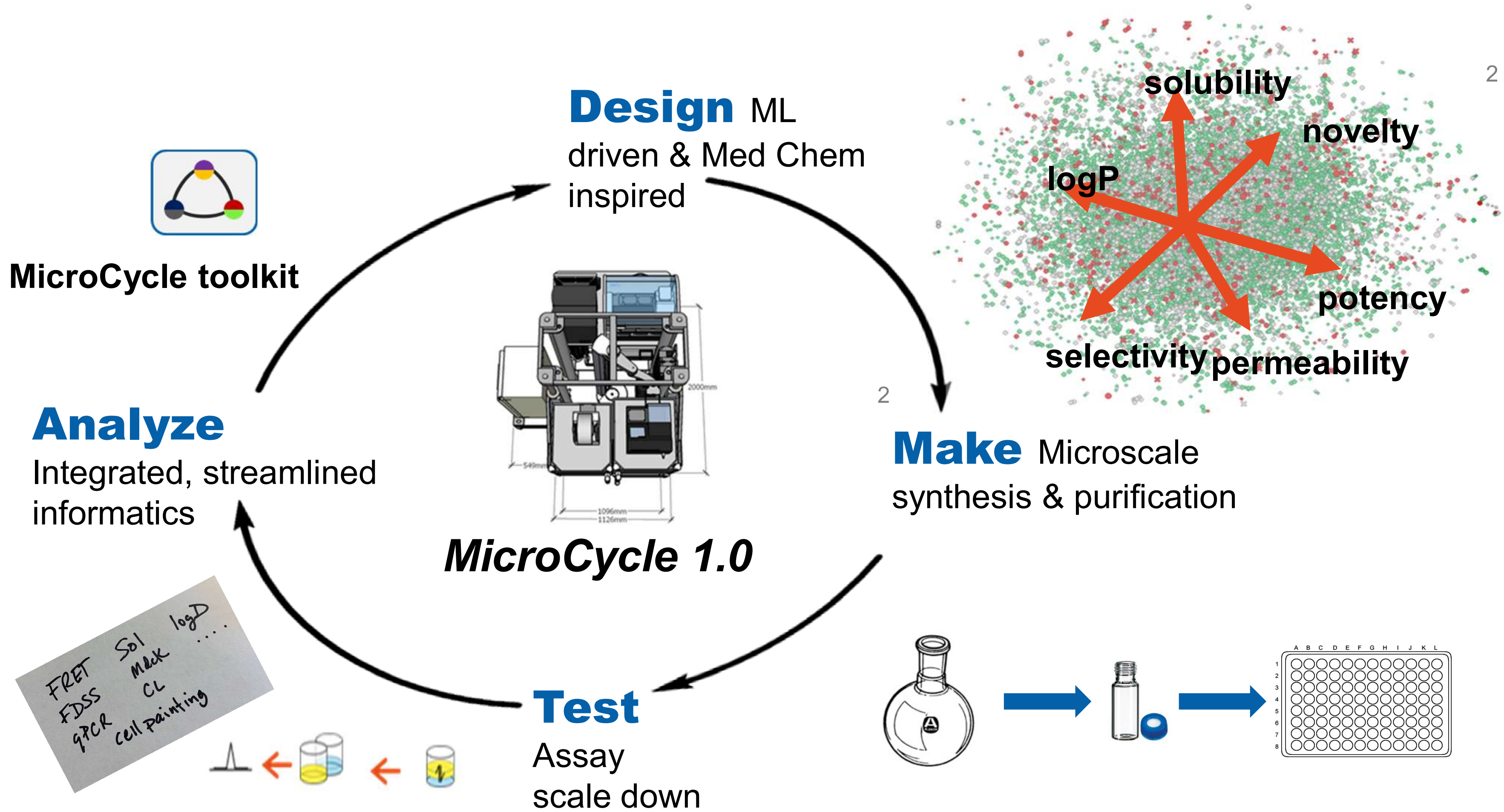


Self driving chemical space exploration

Clayton Springer
Jacob Gora

NOVARTIS | Reimagining Medicine

MicroCycle automated workflow



MicroCycle utilizes state-of-the-art chemistry, assay technology and automation coupled with cutting edge machine learning (ML) to accelerate drug discovery

Lead optimization is significant part of the drug discovery process

<u>PHASE</u>	target to hit	hit to lead	lead opt	preclinical	phase 1	phase 2	phase 3	submission to launch	Launch
chance of progression to next step	**80%**	**75%**	**85%**	**69%**	**54%**	**34%**	**70%**	**91%**	
number needed for each launch	**24.3**	**19.4**	**14.6**	**12.4**	**8.6**	**4.6**	**1.6**	**1.1**	**1**
cost per cycle	**$1**	**$2.50**	**$10**	**$5**	**$15**	**$40**	**$150**	**$40**	
total cost	**$24**	**$49**	**$146**	**$62**	**$128**	**$185**	**$235**	**$44**	**$873**
percent of total cost	**3%**	**6%**	**17%**	**7%**	**15%**	**21%**	**27%**	**5%**	

Paul, S., *Nat Rev Drug Discov* **9,** 203–214 (2010).

NOVARTIS | Reimagining Medicine

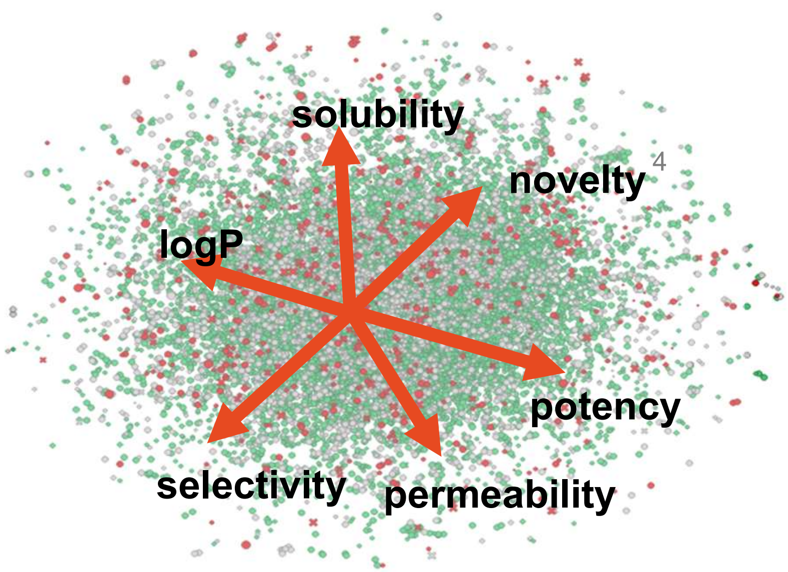
Inspiration from gold mining

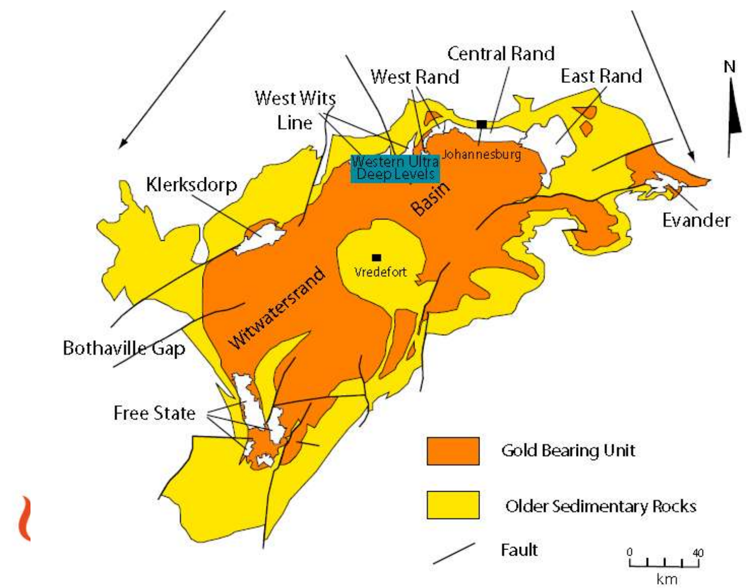
Medicinal Chemistry — "GeoStatistics"

Looking for rare, valuable minerals or drugs

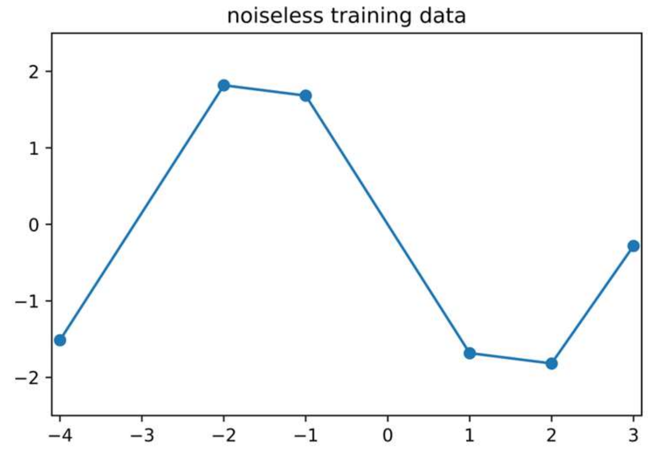
Data is expensive

'regions nearby each other are more similar than those far away'

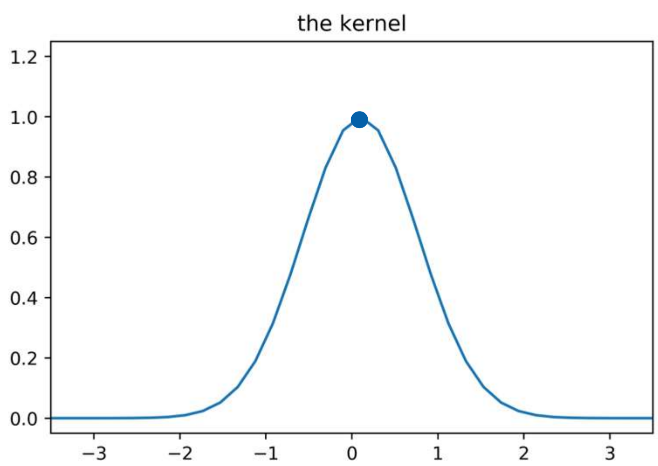




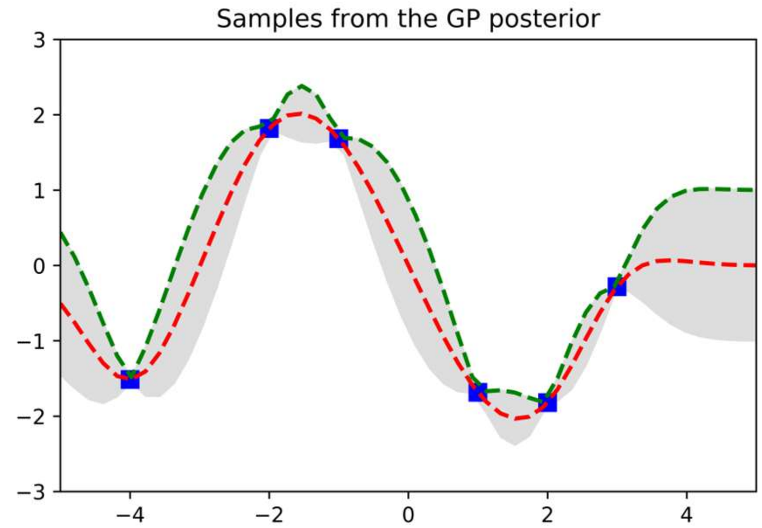
Kriging



Data
Points

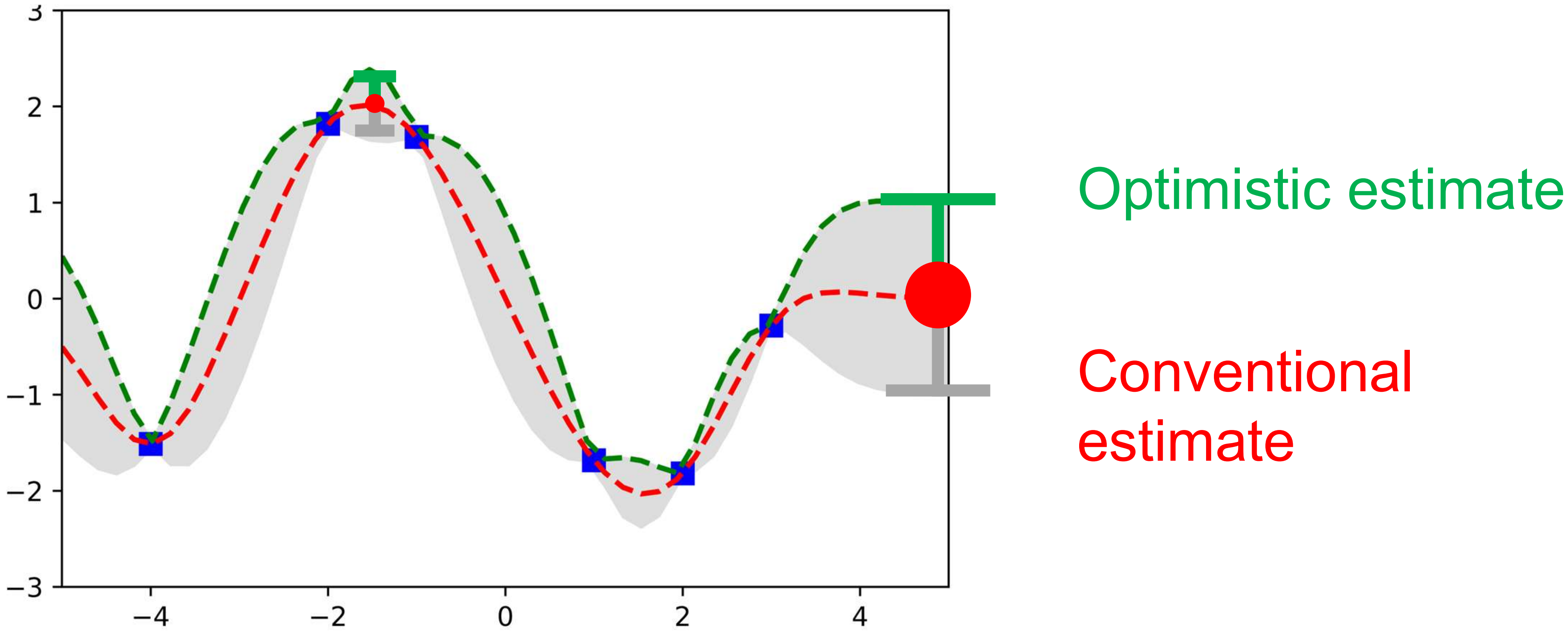


Local
knowledge

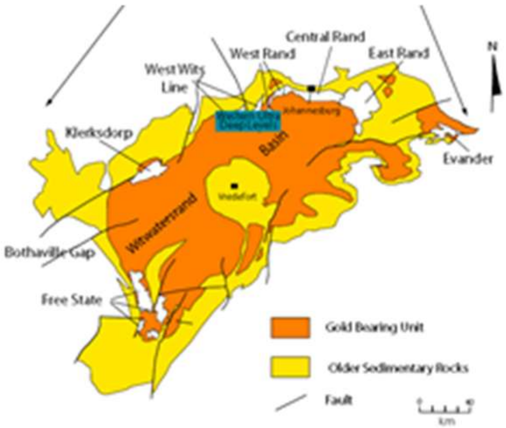


Gaussian
Process Model
combines both

Model based selection strategies

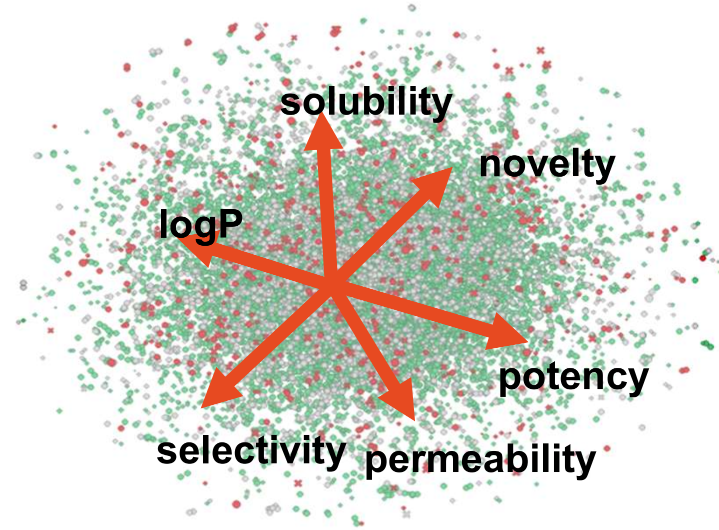


Geostatistics → Drug Discovery

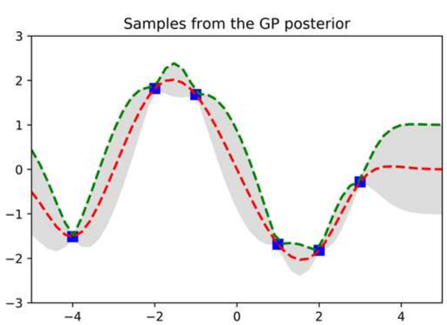


(Lat, Long, Depth)

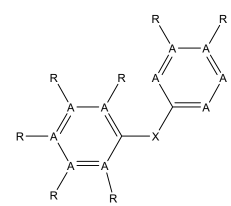
From discrete molecules to 1000s of descriptors

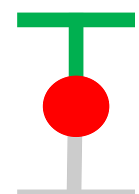


Gaussian processes



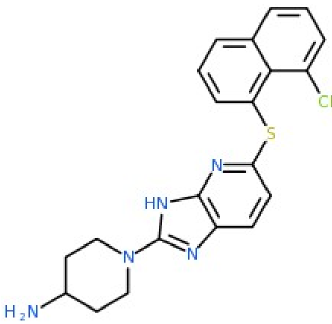
Computational chemistry model with error bars



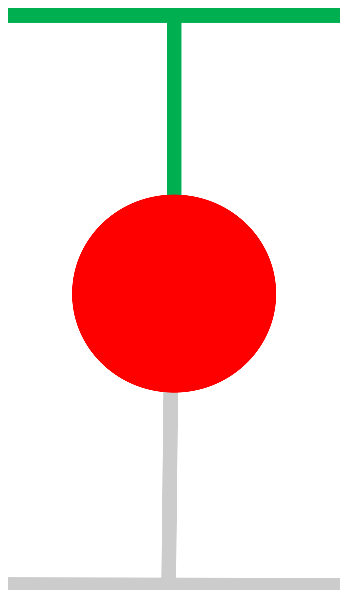


NOVARTIS | Reimagining Medicine

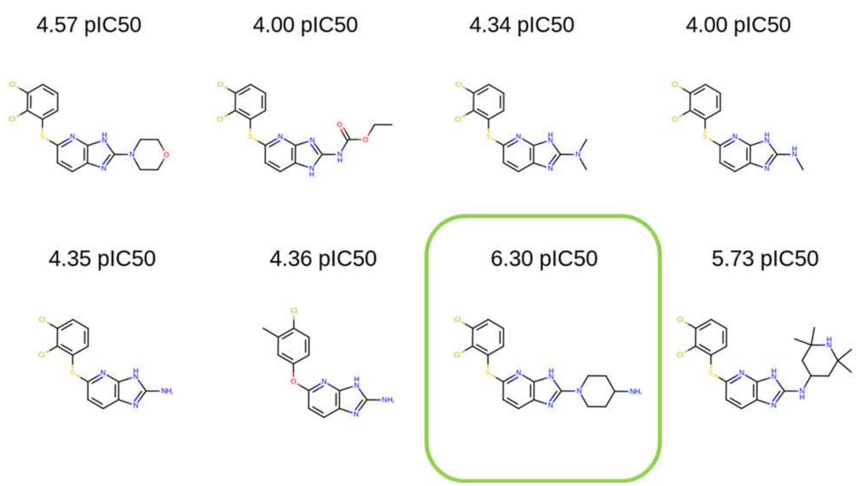
Kriging in chemical space



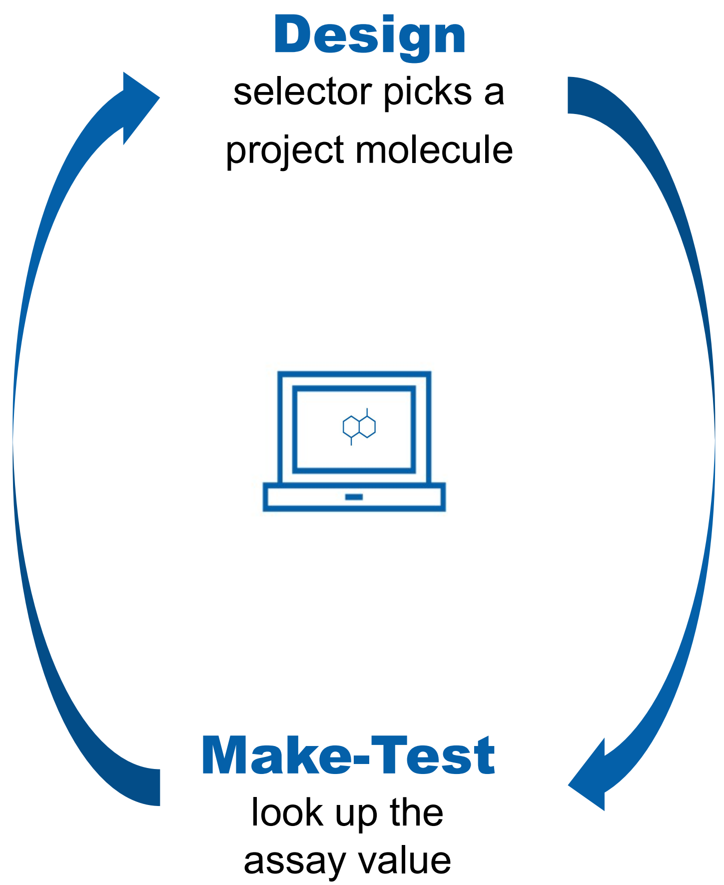
Model:



Evaluate selector with simulated design make test cycle

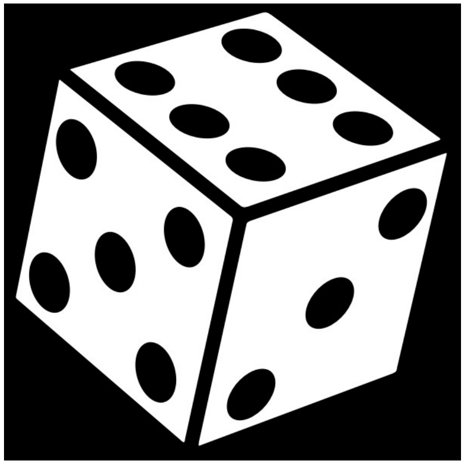


8 Seed data points

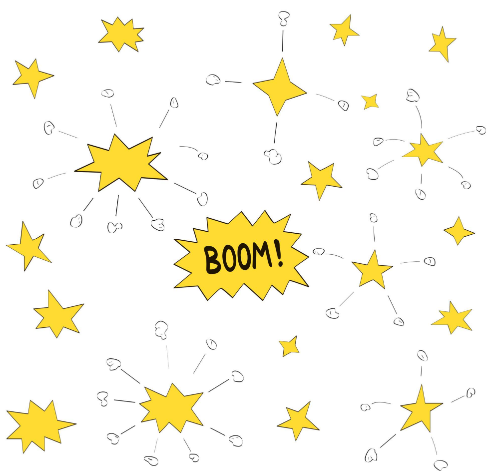


NOVARTIS | Reimagining Medicine

Model free selection strategies

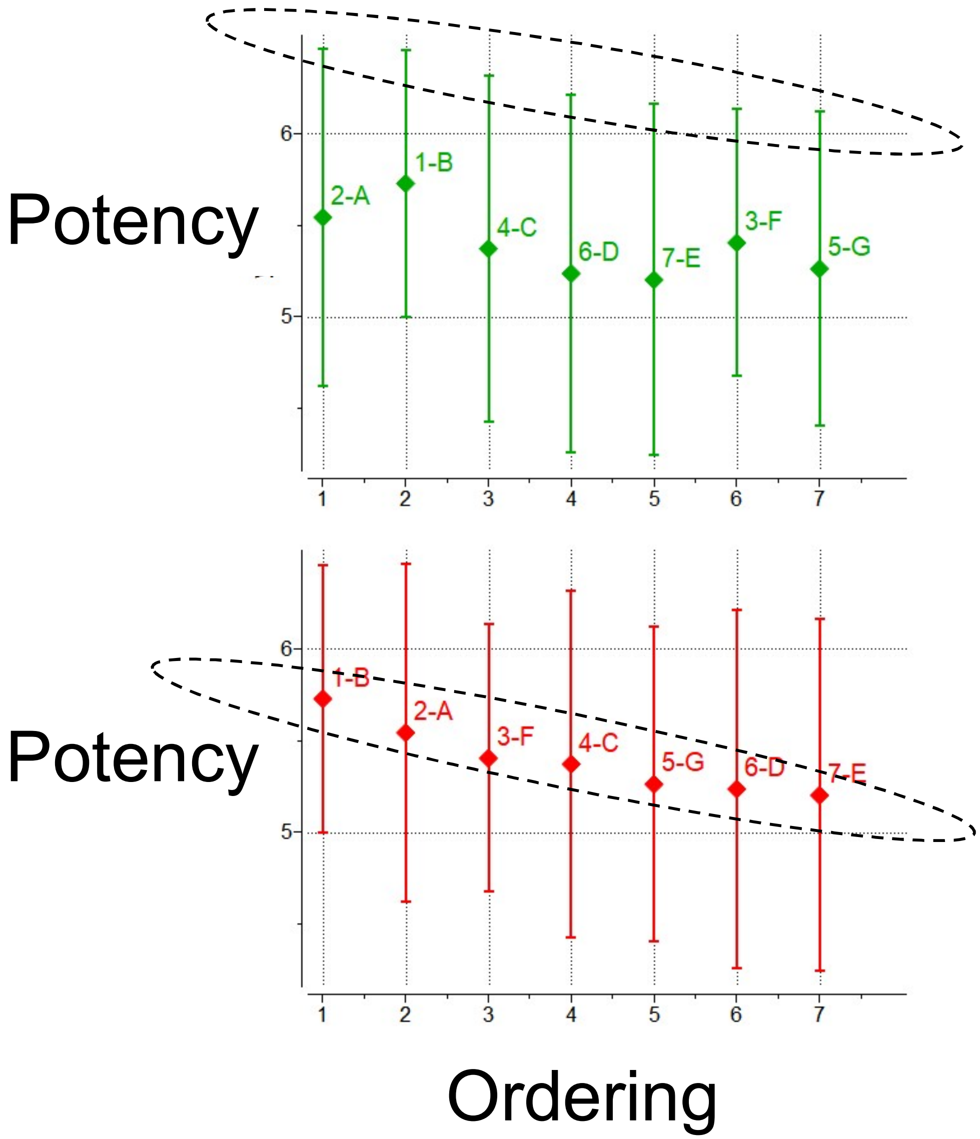


Random picking



Diverse picking

Ranking is different for Optimistic vs Conventional

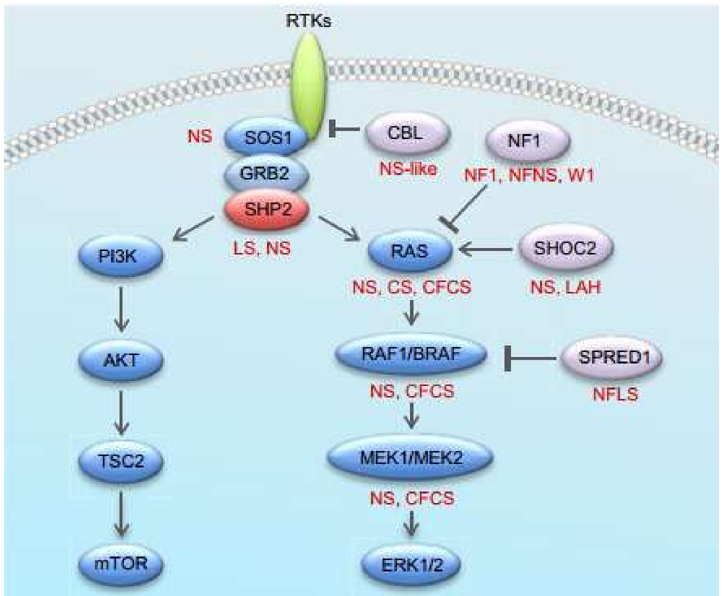


Optimistic ranks by high estimate

Conventional ranks by middle estimate

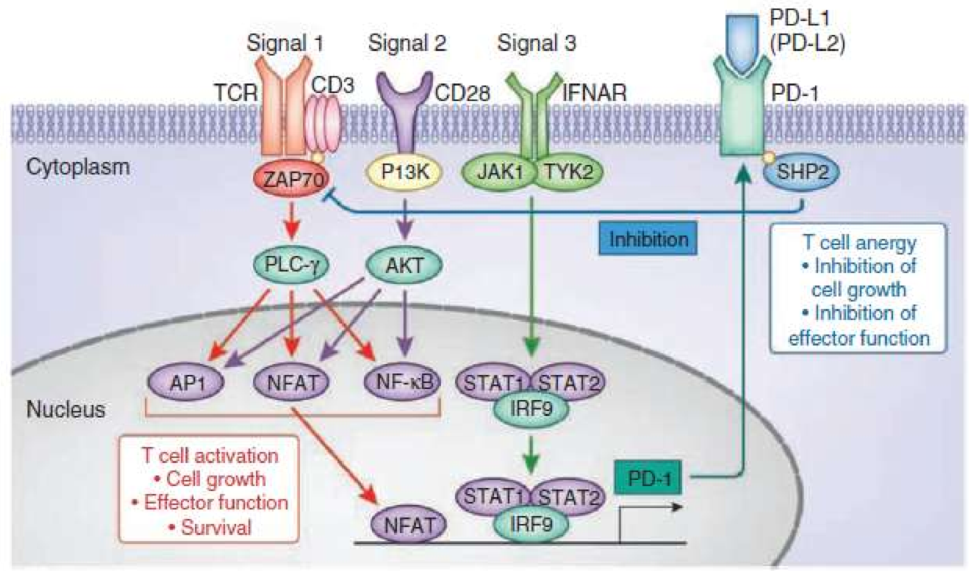
Retrospective analysis of SHP2

- Required for signaling downstream of many receptor tyrosine kinases



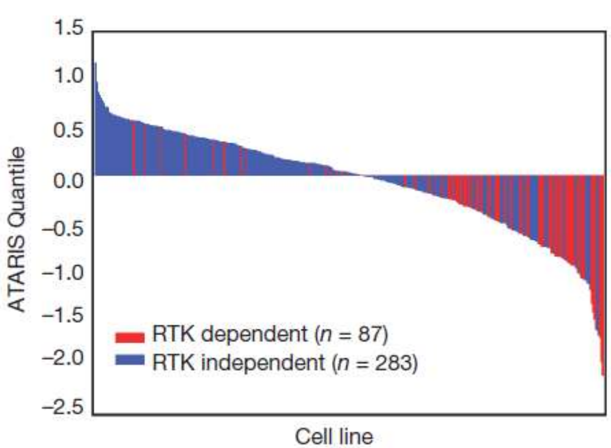
Kontaridis 2011

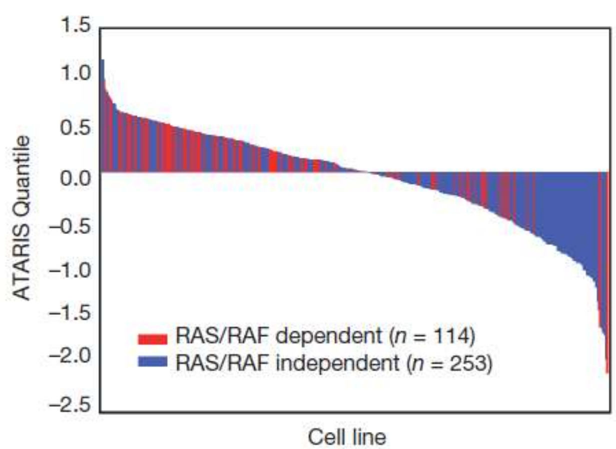
- Downstream transducer of PD-1 signaling



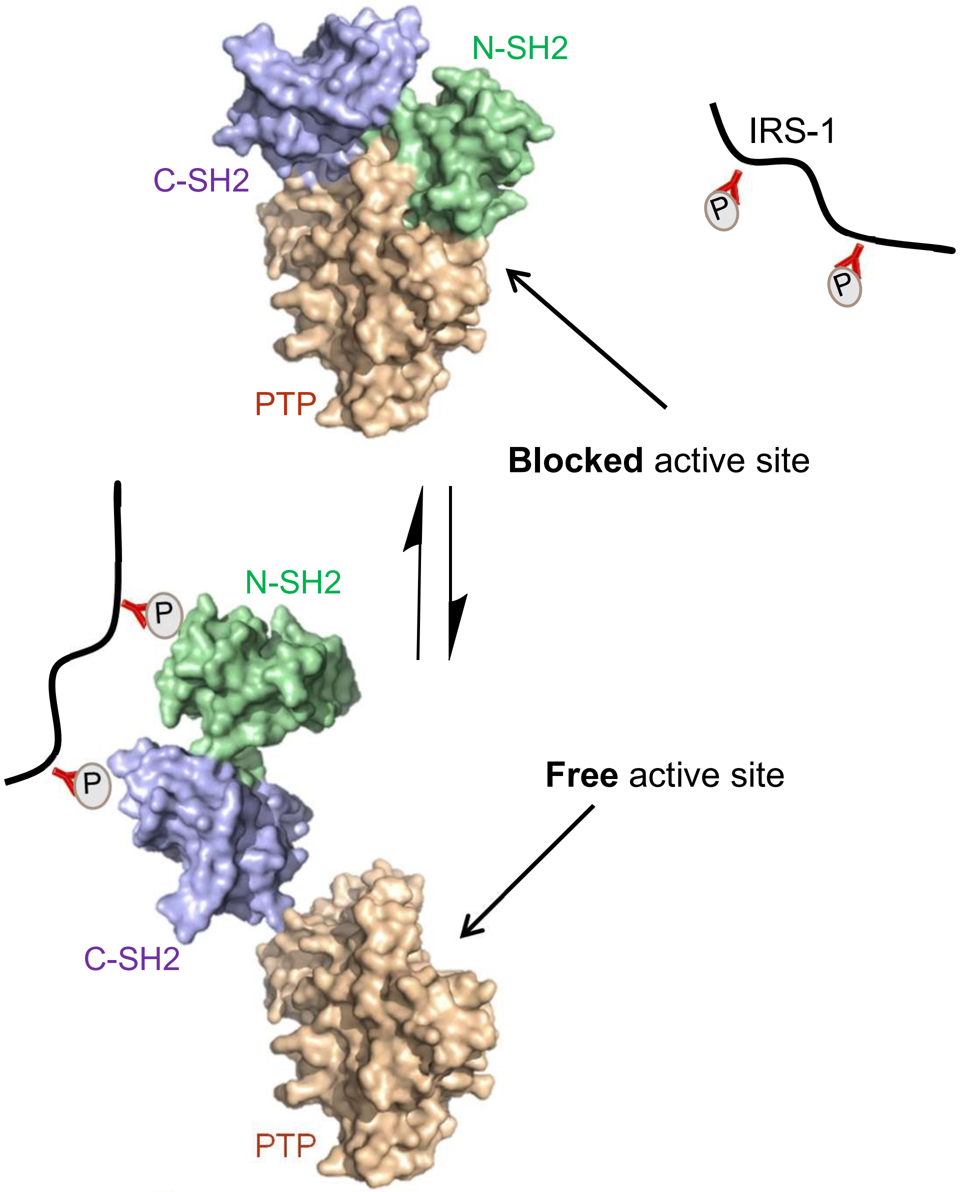
Okazaki et al. 2013

shRNA: SHP2 dependence correlates with RTK dependence



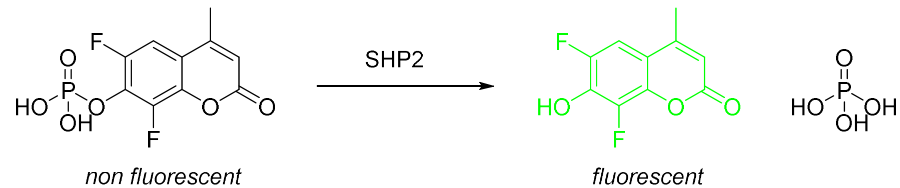


Use SHP2 biochemical IC50 as our data set

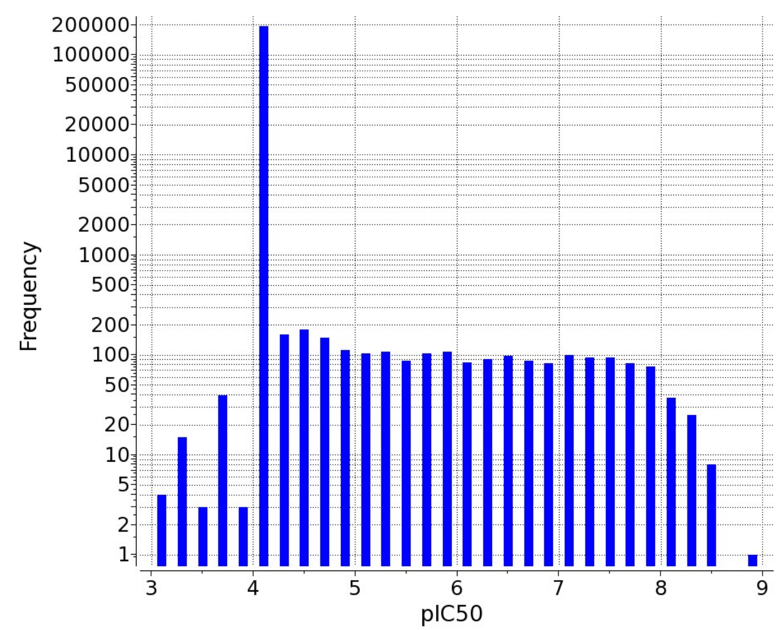


Oncology

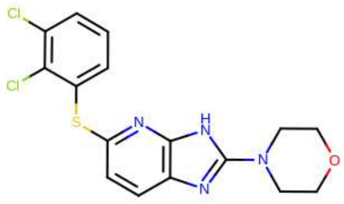
1. DIFMUP Assay: FL vs PTP



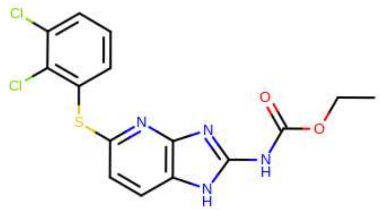
<u>(190+5)K total compounds:</u>
190K: >100uM from screening
5K: compounds with IC50s



Initial hit structures and data



4.57 pIC50

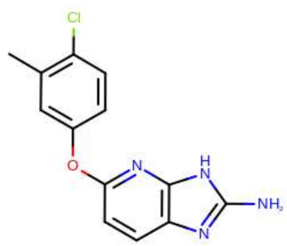


4.00 pIC50

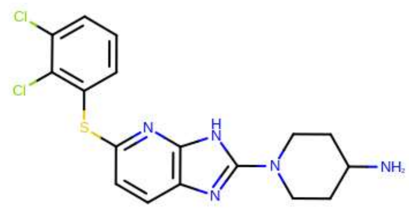
4.34 pIC50

4.00 pIC50

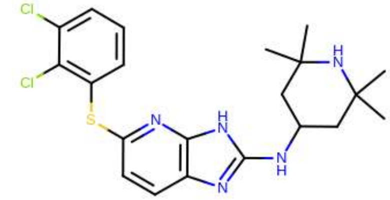
4.35 pIC50



4.36 pIC50



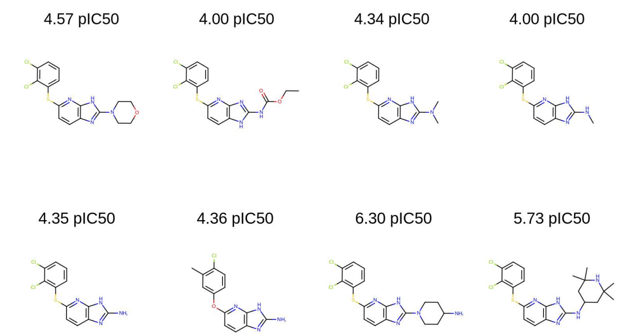
6.30 pIC50



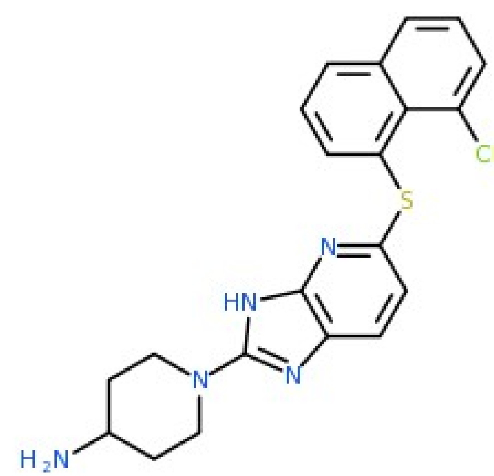
5.73 pIC50

Round #1 –

- Training data has 8 compounds

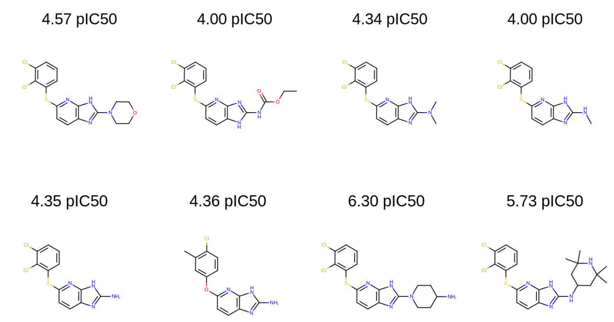


- Build QSAR model
- Predict on (190+5)K chemical space
- Conventional selector picks molecule
- Reveal data: 12 uM
- Next round ...



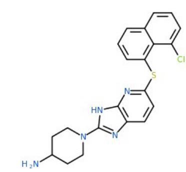
Round #2

1. Training data is now 9 (= 8+1)



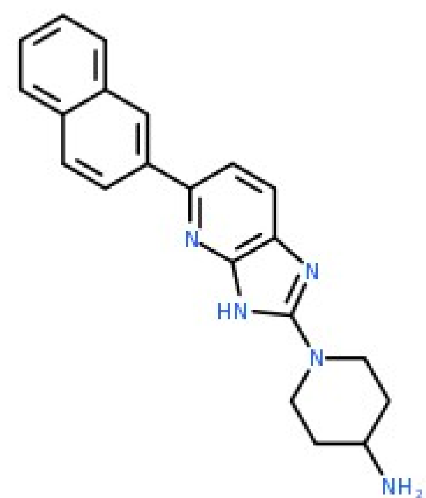
2. Build QSAR model

6.1 pIC50



3. Predict on (190+5)K chemical space

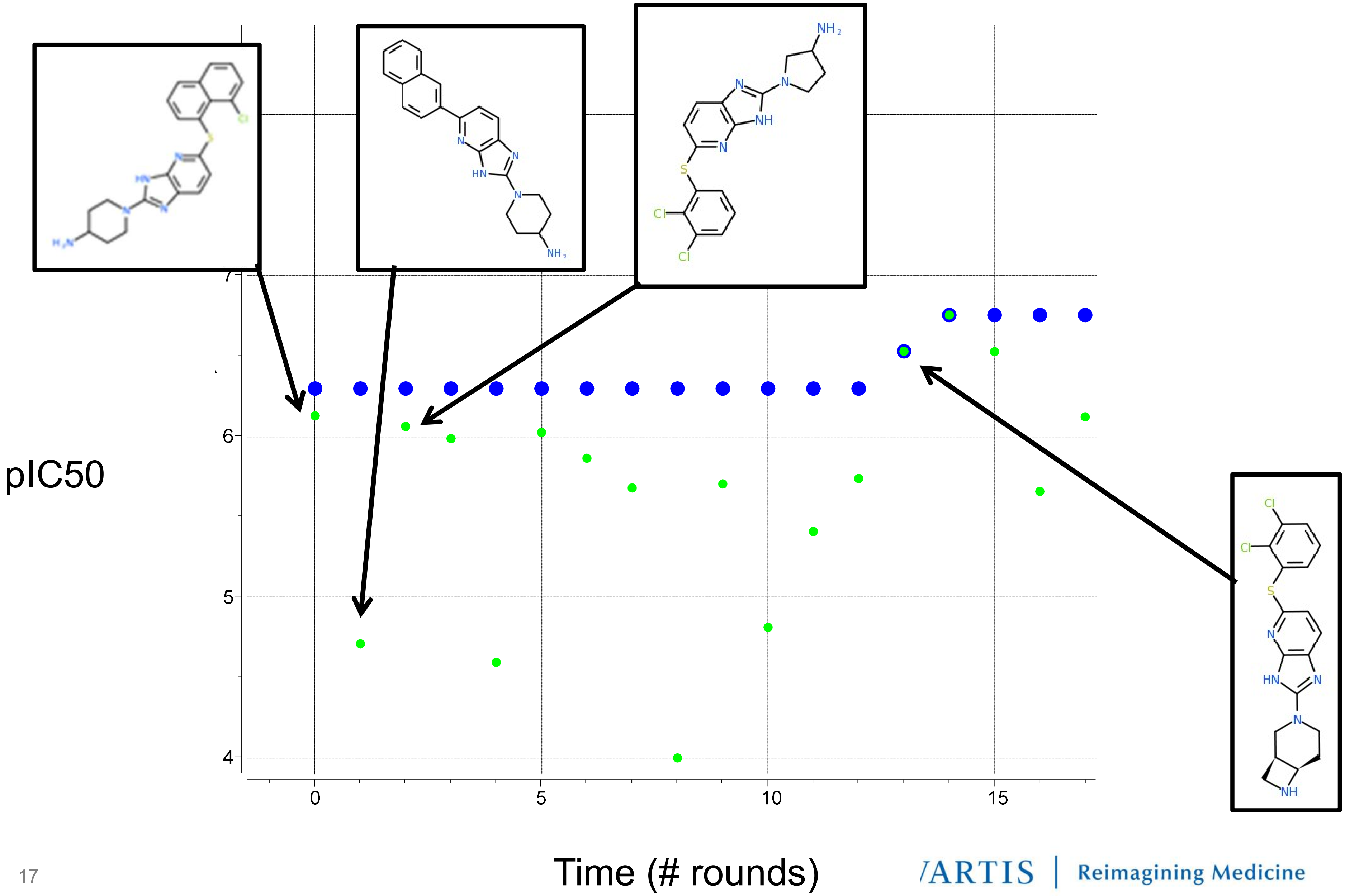
4. Conventional selector picks molecule



5. Reveal data: 20 uM

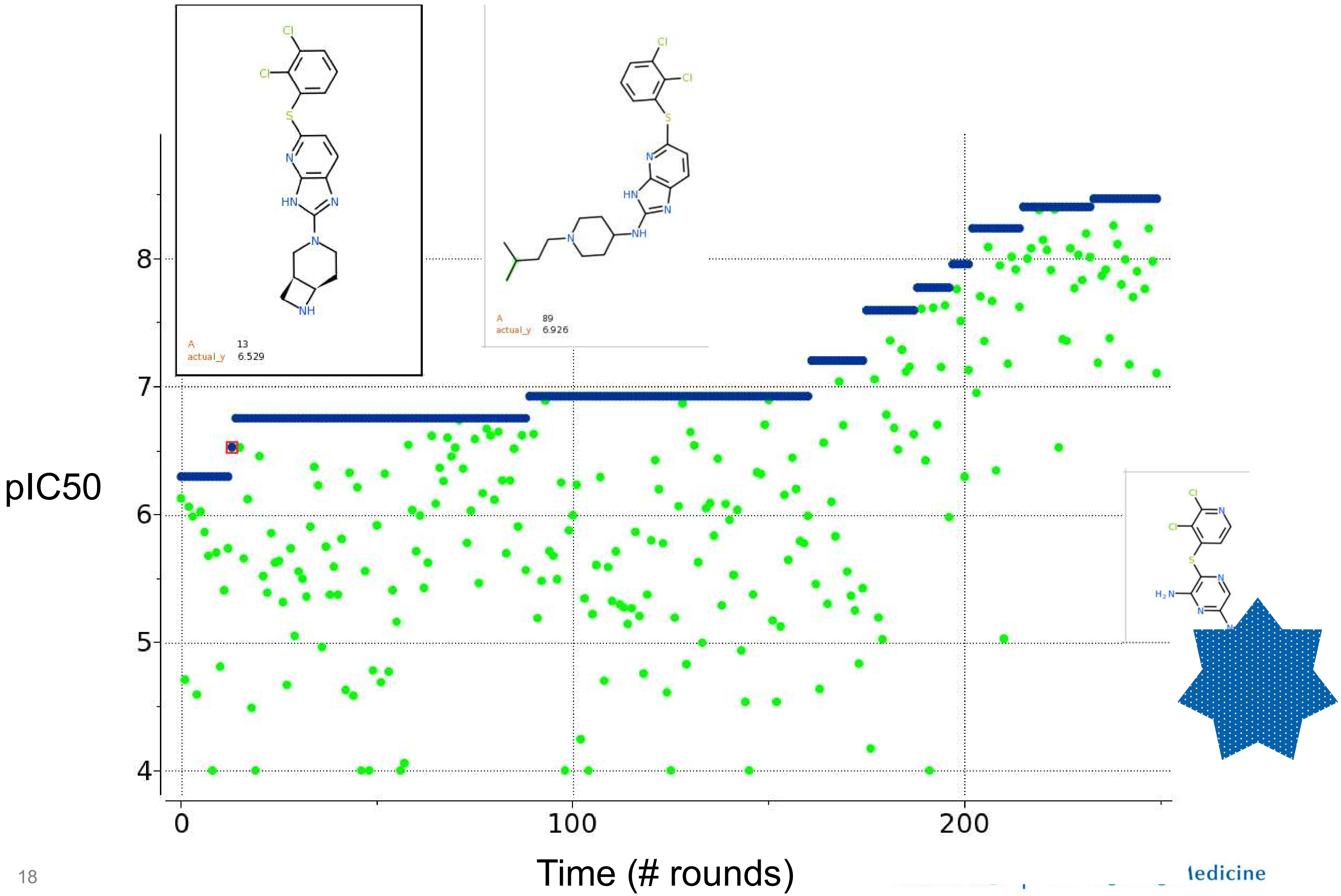
6. Next round ...

Judge by best found so far

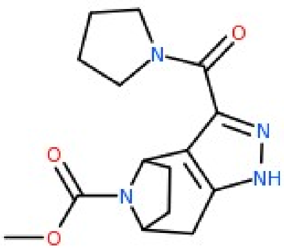
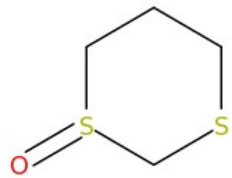
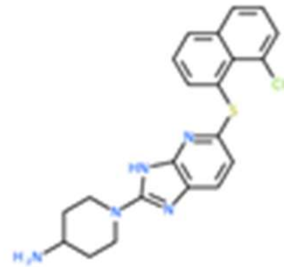
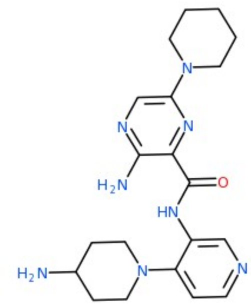
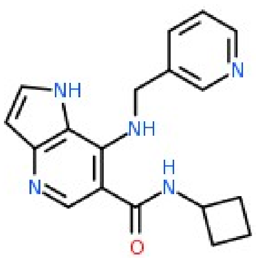
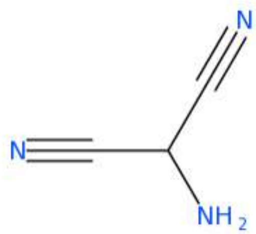
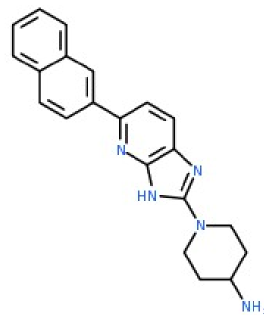
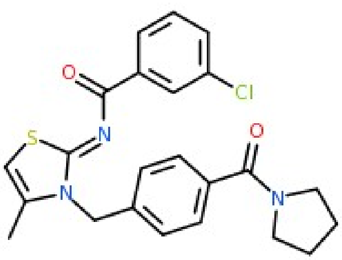
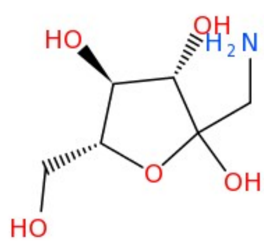
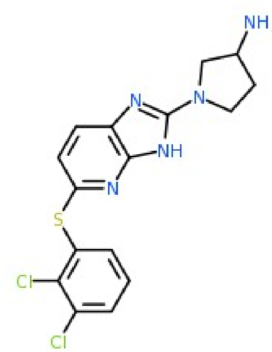
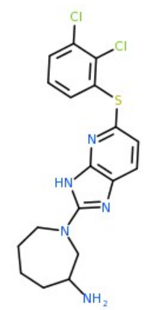


/ARTIS | Reimagining Medicine

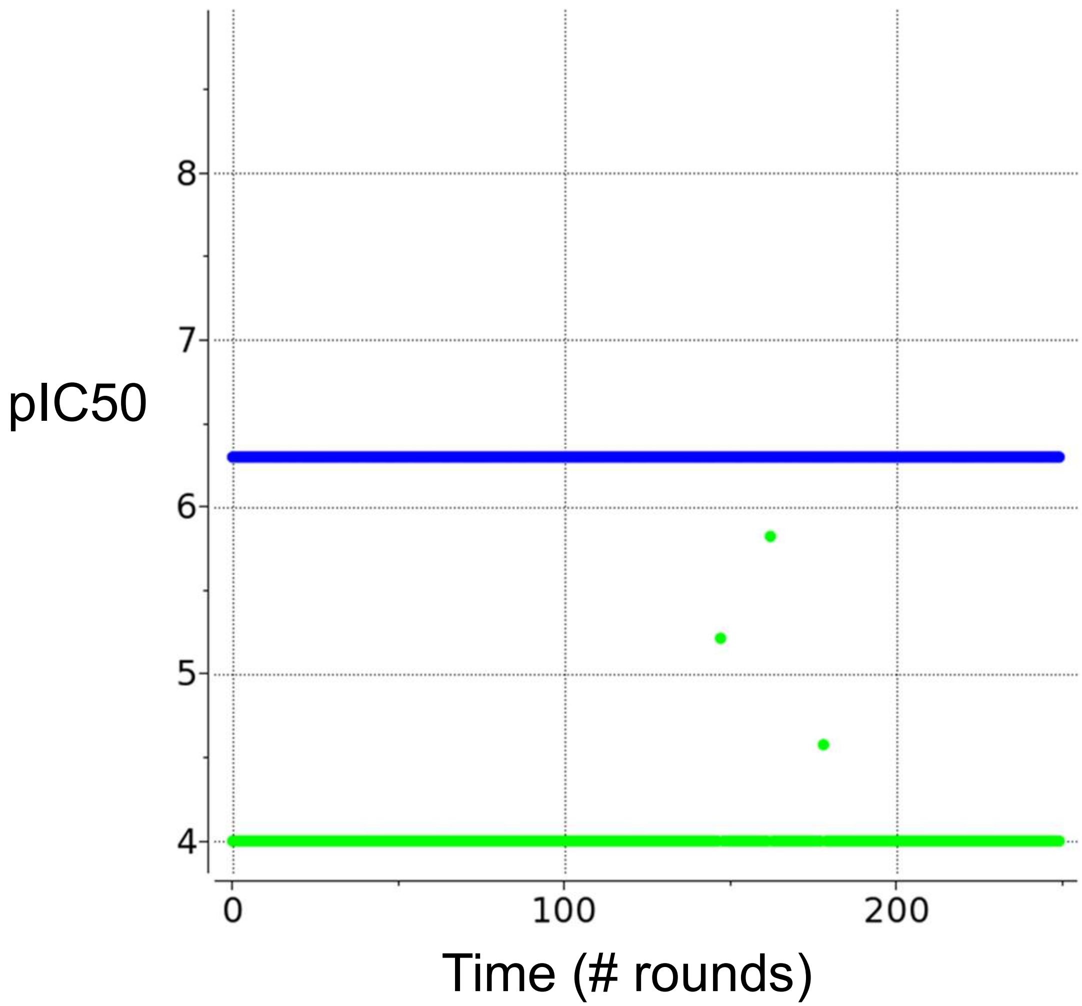
250 rounds of conventional QSAR

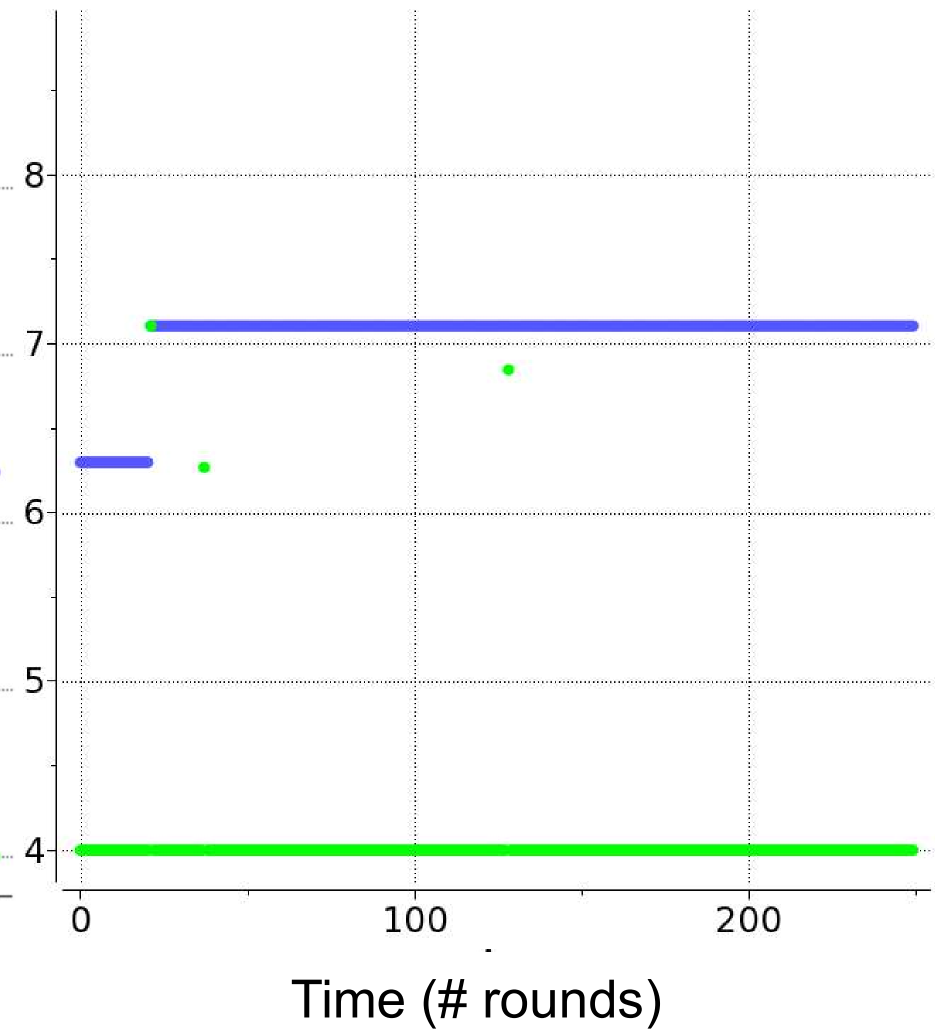


Picks from 4 selectors

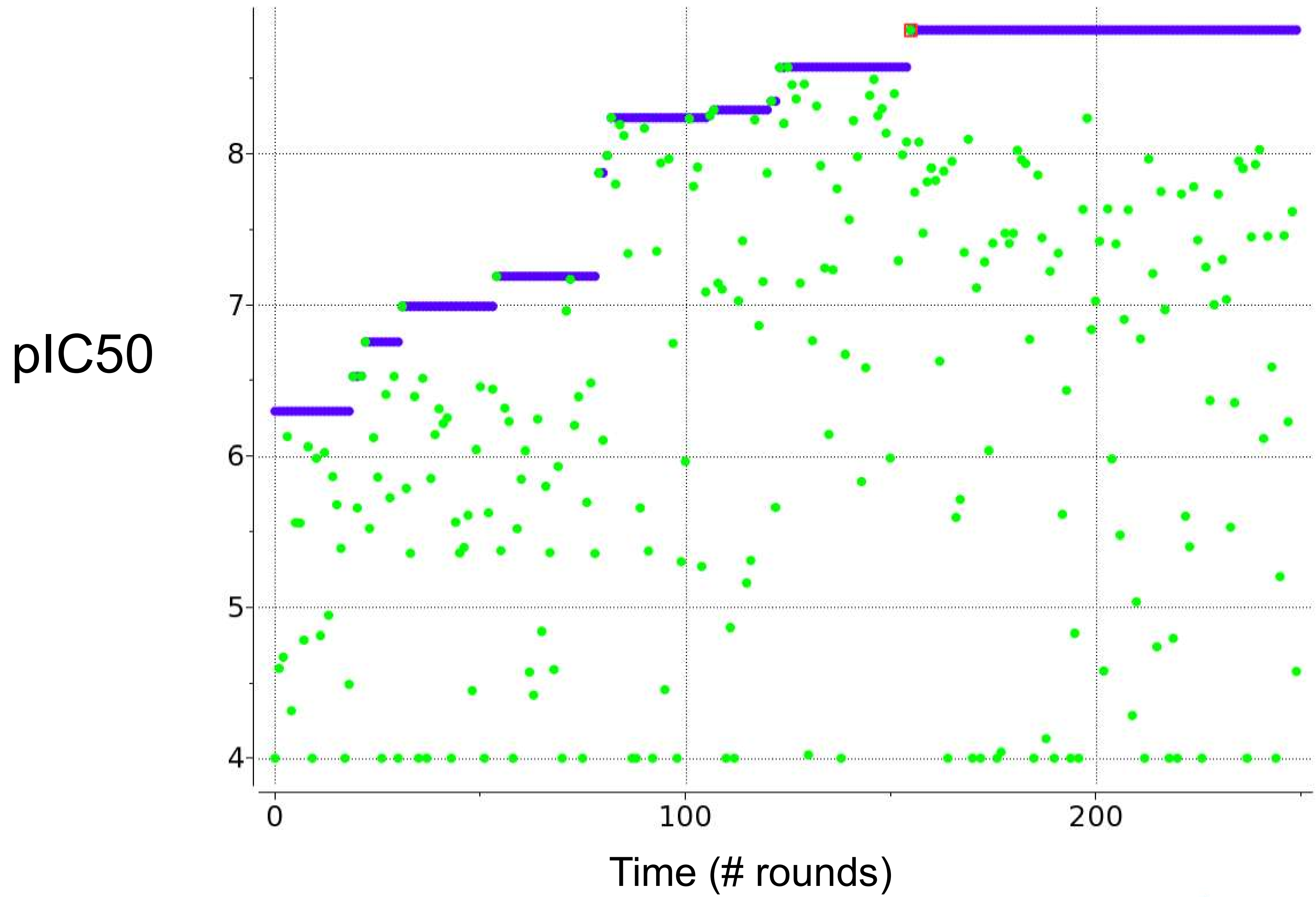
Random	Diverse	Conventional Model	Optimistic Model
			
			
			

250 rounds of diverse, random picks

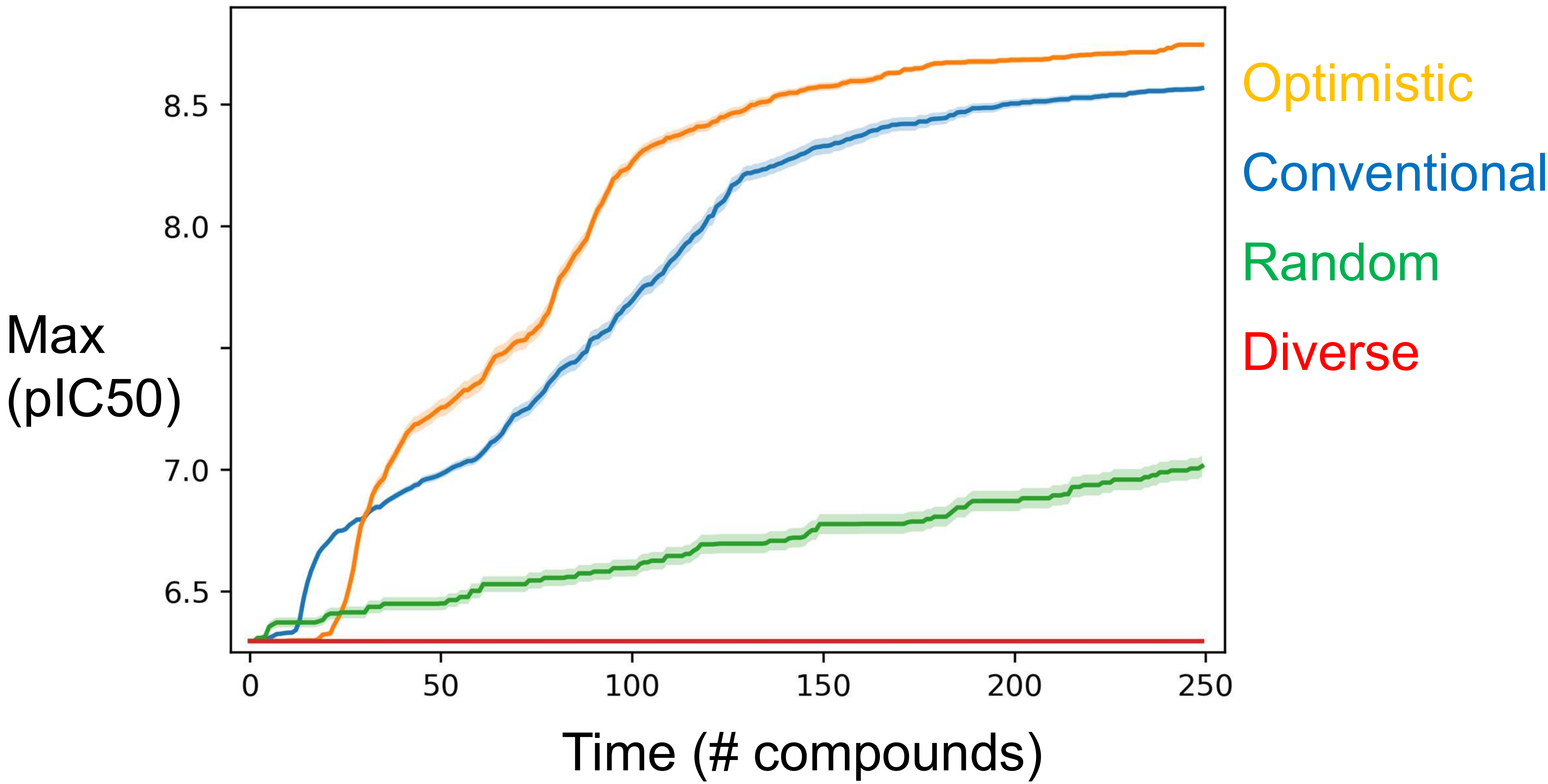




250 picks of optimistic QSAR



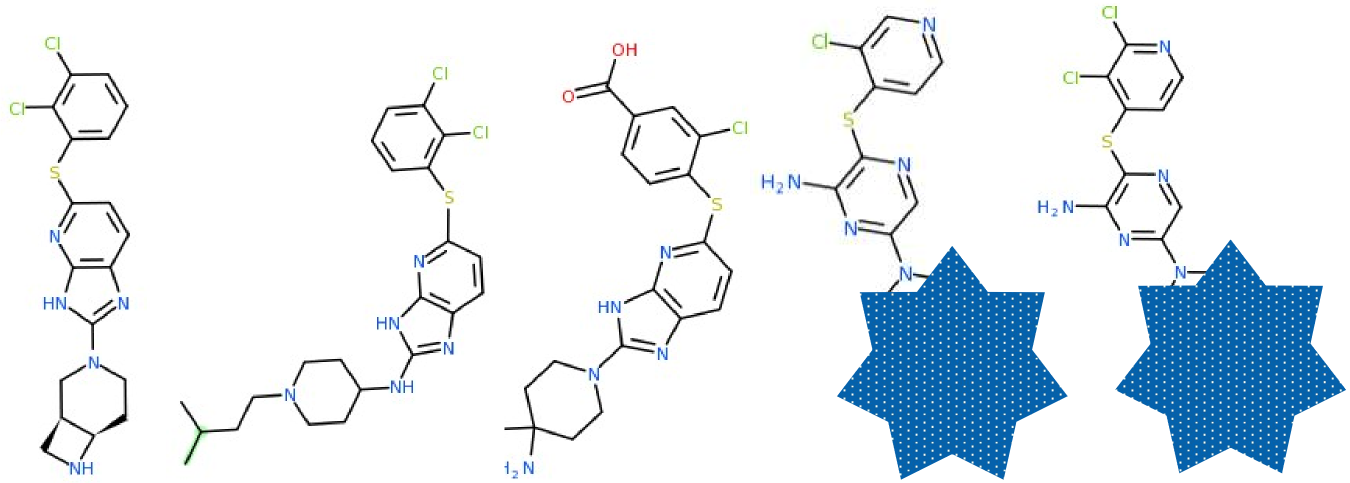
Repeat 99 times



NOVARTIS | Reimagining Medicine

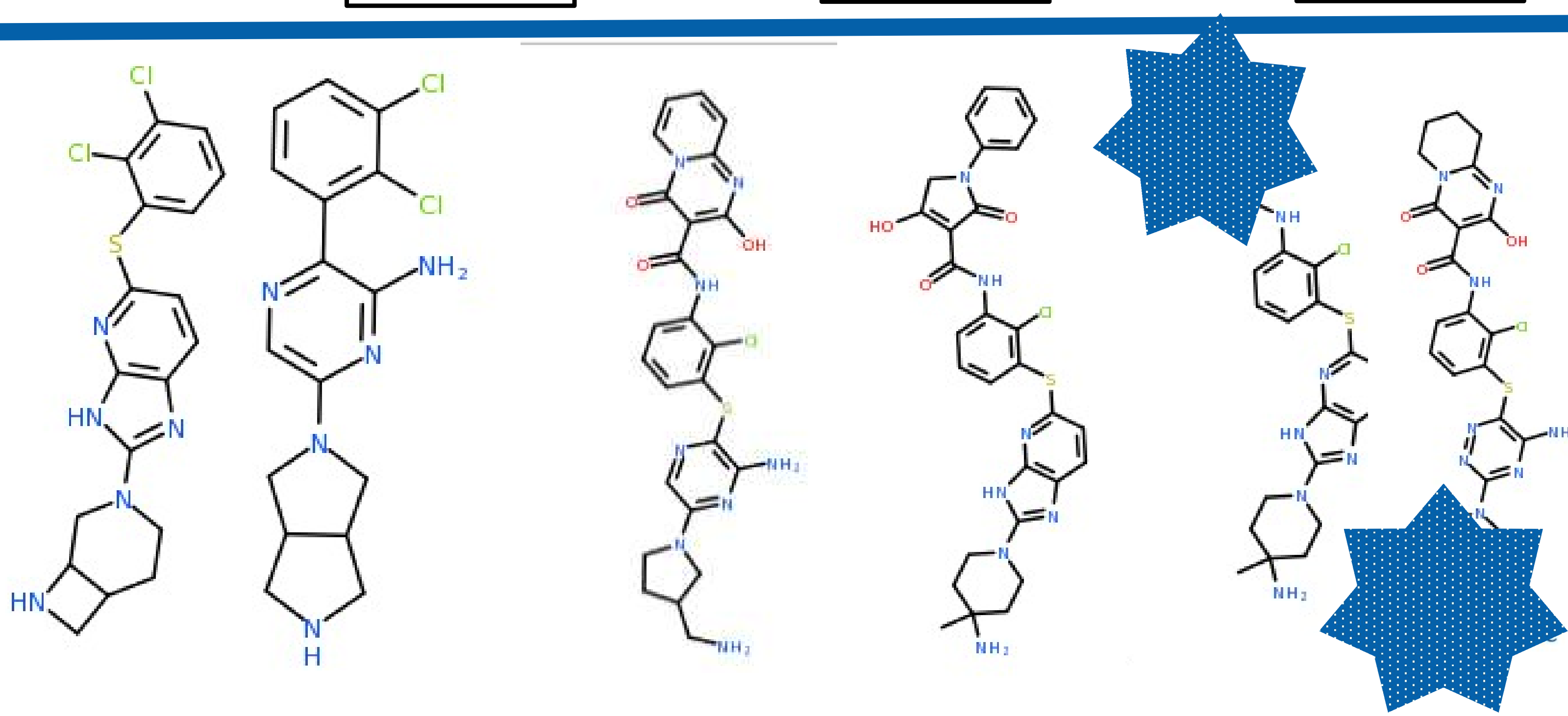
Succession of actives

Conventional

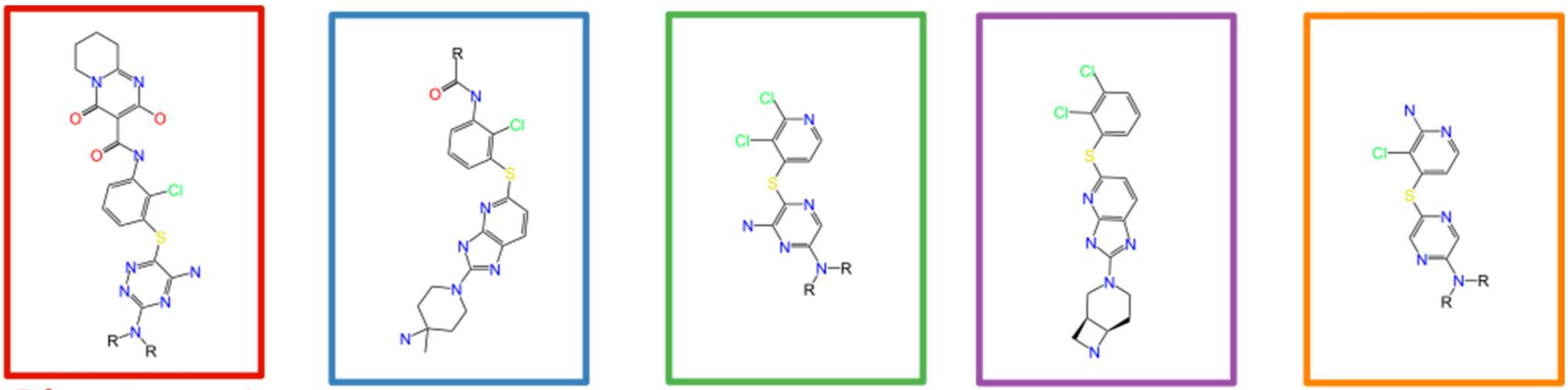


pIC50 | 7 | 8 | 9

Optimistic



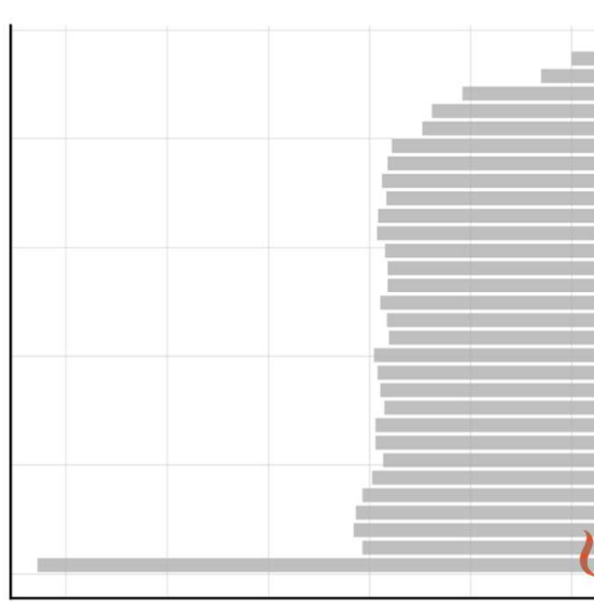
Clustering by SHP2 potency

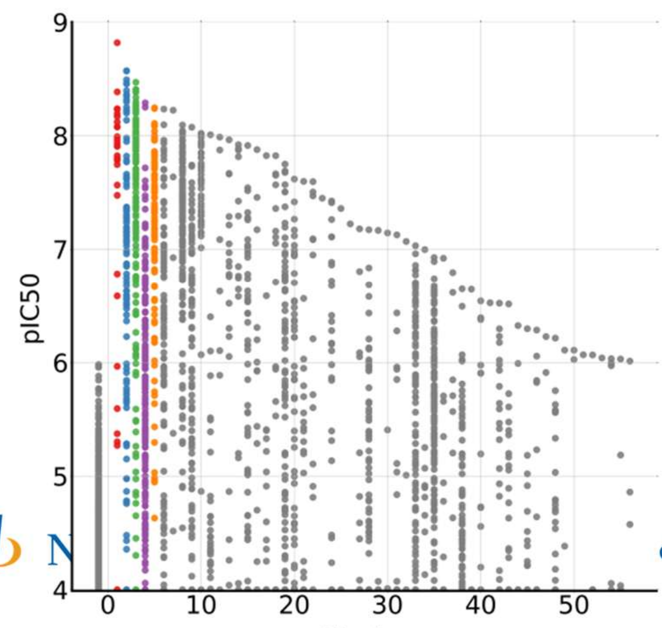


Cluster 1 Cluster 2 Cluster 3 Cluster 4 Cluster 5

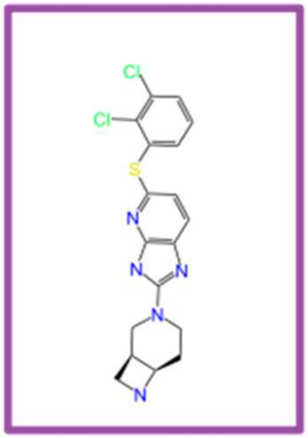
5K: compounds
with IC50s

190K inactives
from screening
treated as >100uM

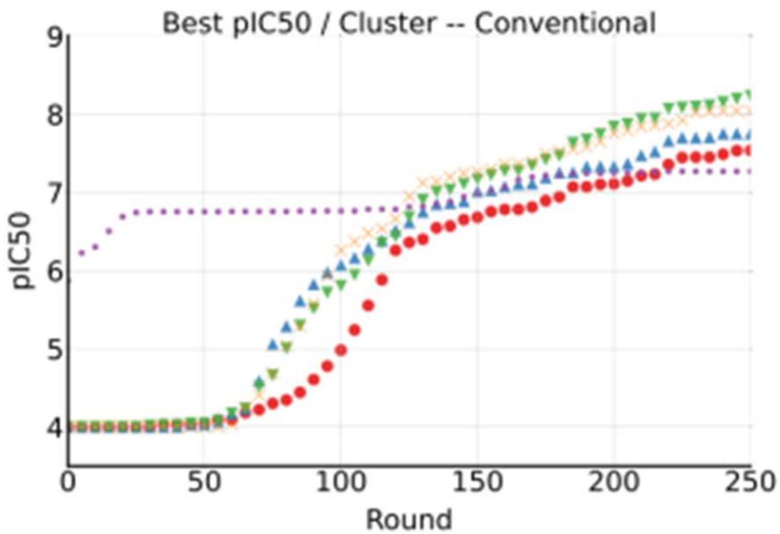


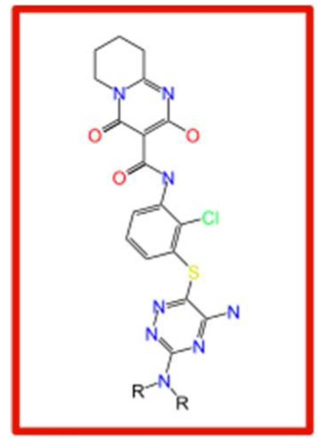


pIC50 Progress by Cluster Conventional vs Optimistic

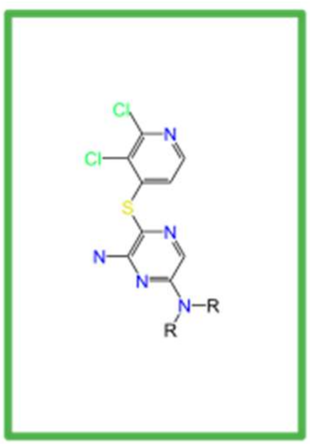


Cluster 4

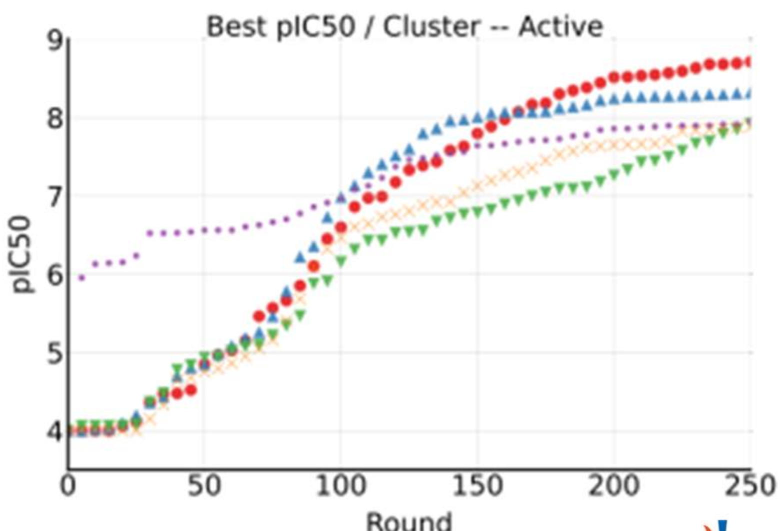


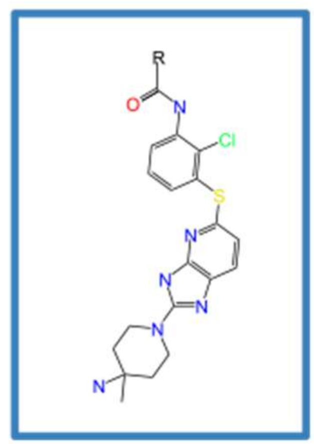


Cluster 1



Cluster 3

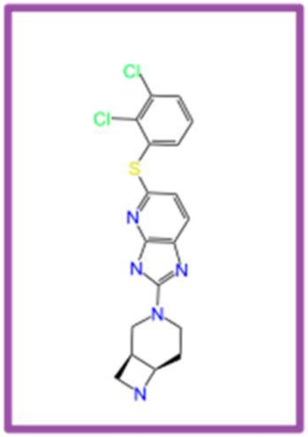




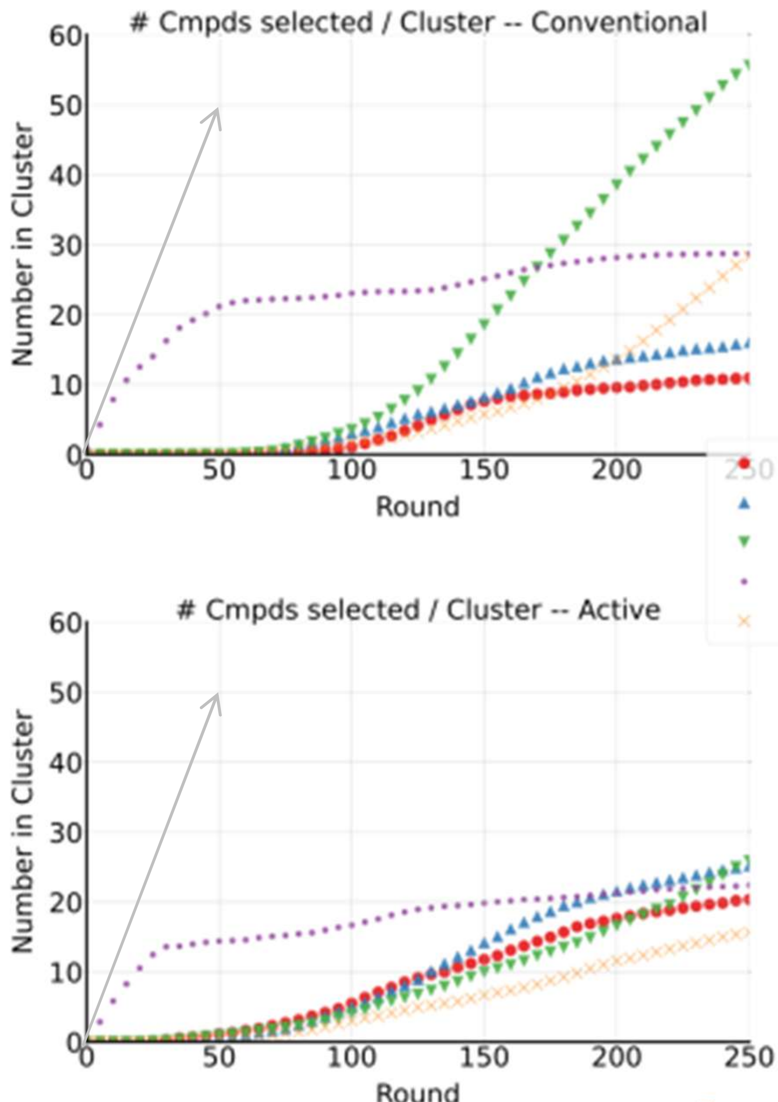
Cluster 2

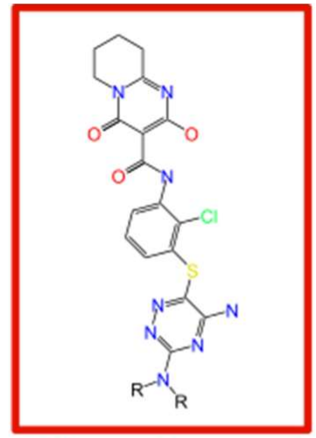
NOVARTIS | Reimagining Medicine

Clusters sampled
Conventional vs Optimistic

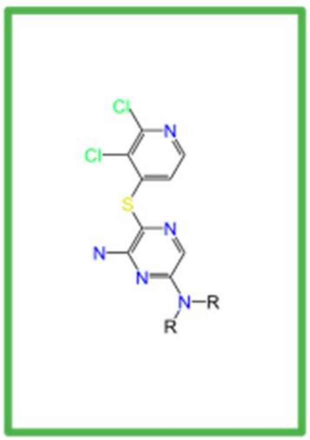


Cluster 4

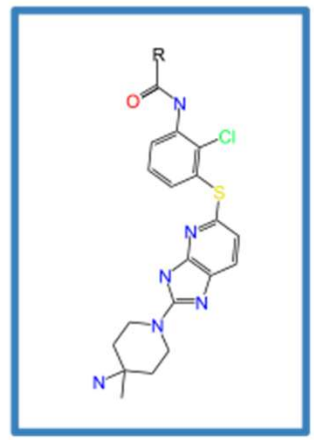




Cluster 1



Cluster 3



Cluster 2

Review of Caveats

Used vanilla QSAR

ADME

Synthetic Accessibility of compounds

Batches of one

Conclusions – retrospective analysis

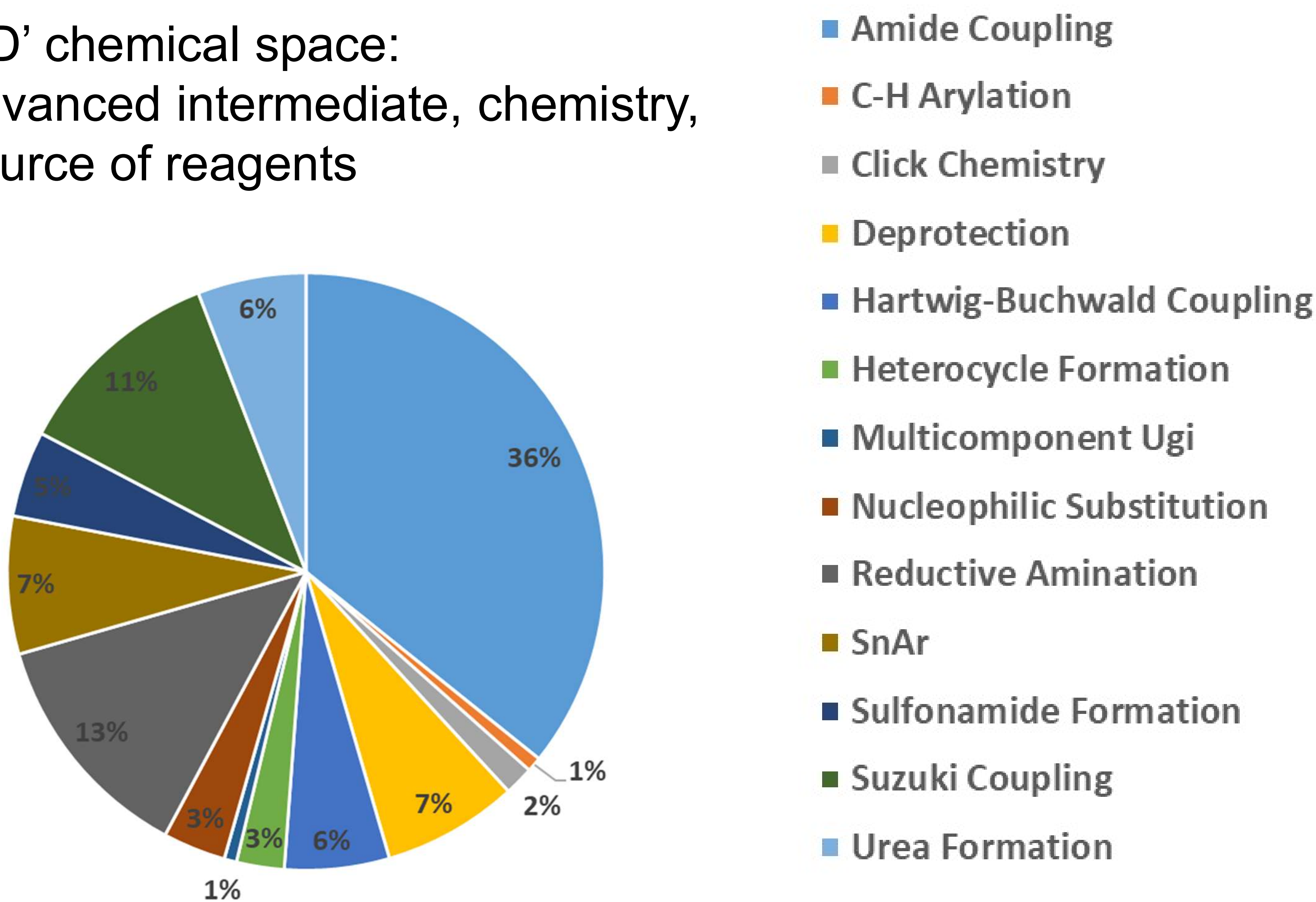
Data is expensive: optimize its construction

Active learning with optimistic searching is effective for finding actives

vanilla QSAR can transverse scaffolds in a real medicinal chemistry SAR data set

Most MicroCycle experience is with 1D libraries

‘1D’ chemical space:
advanced intermediate, chemistry, source of reagents

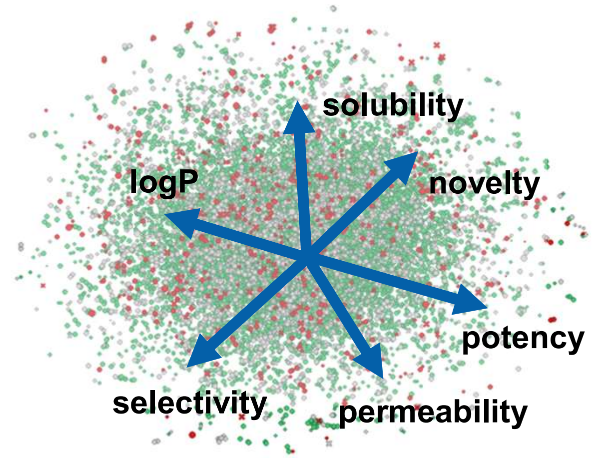


MicroCycle: Project Impact

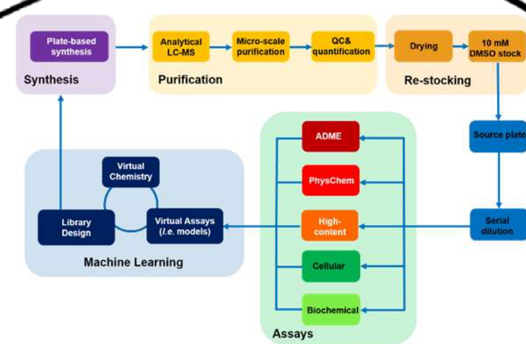
	NVP-001
Profiling assay I IC50	195.4 nM
Profiling assay II IC50	-
HT Solubility pH6.8 [mM] / in silico	0.033 / >100 uM
Direct LogP / cLogP	3.9 / 2.3
Direct LogD / cLogD	3.1 / 1.0

Starting point

Design Rapidly evaluate SAR on three hit series focusing on properties



Analyze Focused library of N-heterocycles for final iteration



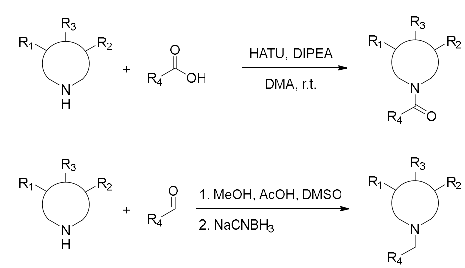
MicroCycle 1.0

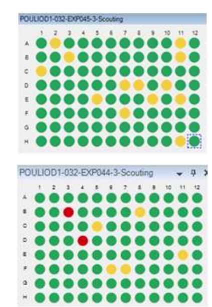
Make Microscale (1 mg) synthesis, purification

- *7 Libraries*
- *5 Reaction Types*
- *>400 NVPs*
- *3 vectors*

LC-MS heat map:

Product found
Low conversion
No product found





	NVP-002
Profiling assay I IC50	12.9 nM
Profiling assay II IC50	-
HT Solubility pH6.8 [mM] / in silico	<0.004 / >100 uM
Direct LogP / cLogP	x / 2.8
Direct LogD / cLogD	3.8 / 2.5

Test Biochemical profiling, Cl_{int}, HT-sol., logP

- 75/245 cpds with double-digit nM activity or lower
- Confirmation by re-synthesis, NMR, XRay

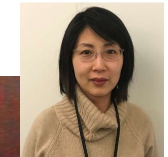
NVP-003	NVP-004	NVP-005
8.3 nM	7.6 nM	3.7 nM
180.1 nM	30.2 nM	72.0 nM
0.048 / >100 uM	0.036 / >100 uM	0.071 / >100 uM
x / 2.5	x / 2.9	x / 2.6
4.2 / 2.0	>4.7 / 2.3	4.2 / 2.1

The MicroCycle Team

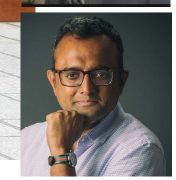










Acknowledgements

Jay Bradner
Karin Briner
Dominic Ehrismann
Tewis Bouwmeester
Steve Martin
Ian Hunt
NGL Team
Natalie Dales
Jutta Blank
Dominic Hoepfner
Guido Koch
Richard Robinson
Trixie Wagner
Mark Hellberg
René Wyler
Dominic Casalena
Mike Tarselli
Felix Thommen
Jerome Giovannoni
Angela Mackay
The CA&BS SEP teams
John Reilly
Richard Lewis

Klemens Hoegenauer
Frederic Berst
Andrea Decker
Jörg Berghausen
Gregori Gerebtzoff
Konstanze Hurth
Hansjörg Lehmann
Yves Auberson
Rick McDonald
Zinnia Kong
Luigi LaVecchia
Anna Pelliccioli
José Duca
Stefanie Flohr
Mark Nash
Berndt Oberhauser
Arndt Meyer
Alexander Smith
Paul Lavan
Paul Groot-Kormelink
Patricia Gehin
Brian Kelley
Patrick Schmutz

Inga Galuba
Matthias Hübscher
Adam Tomaszewski
Richard Janovjak
Gabe Gamber
Declan Hardy
Dyuti Majumdar
Daniela Angst
Klaus Weigand
Carole Pissot
Alan Abrams
Christian Markert
Sujal Deshmukh
Christoph Burkhart

SAB

Christoph Gaul
Jeremy Jenkins
Daniel Baeschlin
Bernard Faller
Markus Stoeckli
Mark Borowsky
Nathan Ross (ex member)

Technical Consultants

Jörg Dreesen
Ekkehard Goerlach
Jennifer Campbell
Dallas Bednarczyk
Phong Hung Nguyen
Suzanne Skolnik
Gina Geraci
Daniel Meyer
Thomas Wolf
Nicholas Flannery
Bill Ulmer
Carol Ginsburg-Moraff
Stephane Rodde

SHP099 acknowledgments

Chemistry
Matthew LaMarche
Julian Chen
Michael Dore
Rajesh Karki
Dyuti Majumdar
Mark Palermo
Lawrence Perez
Michael Shultz
Tim Ramsey
Gang Liu
Mitsunori Kato
Sarah Williams
Jay Larrow
Cary Fridrich

Oncology
Ying-Nan Chen
Brant Firestone
Pascal Fortin
Matthew Meyer
Ping Wang
Shumei Liu
Kathy Hsiao
Joanna Slisz
Kelly Slocum
Mike Acker
Suzanne Zhu
Nick Keen
Bill Sellers

CPC
Travis Stams
Michelle Fodor

TCO & OTR
Wolfgang Hackl

MAP
Laura LaBonte
Franco Lombardo

Patents
Scott Loftus-Reid
Shawn Britt

CPP
Chris Towler

PCS
Stan Spence

Acknowledgements – CADD / cheminfomatics

CADD group

Jose Duca
Viktor Hornak
John Manchester
Jacob Gora
Robert Pearlstein
Florentina Tofoleanu
Xin Chen
Andrei Golosov
Mitsunori Kato
Callum Dickson
Camilo Velez-Vega

NIBR Informatics (NX)

Joann Prescott-Roy
Brian Kelley

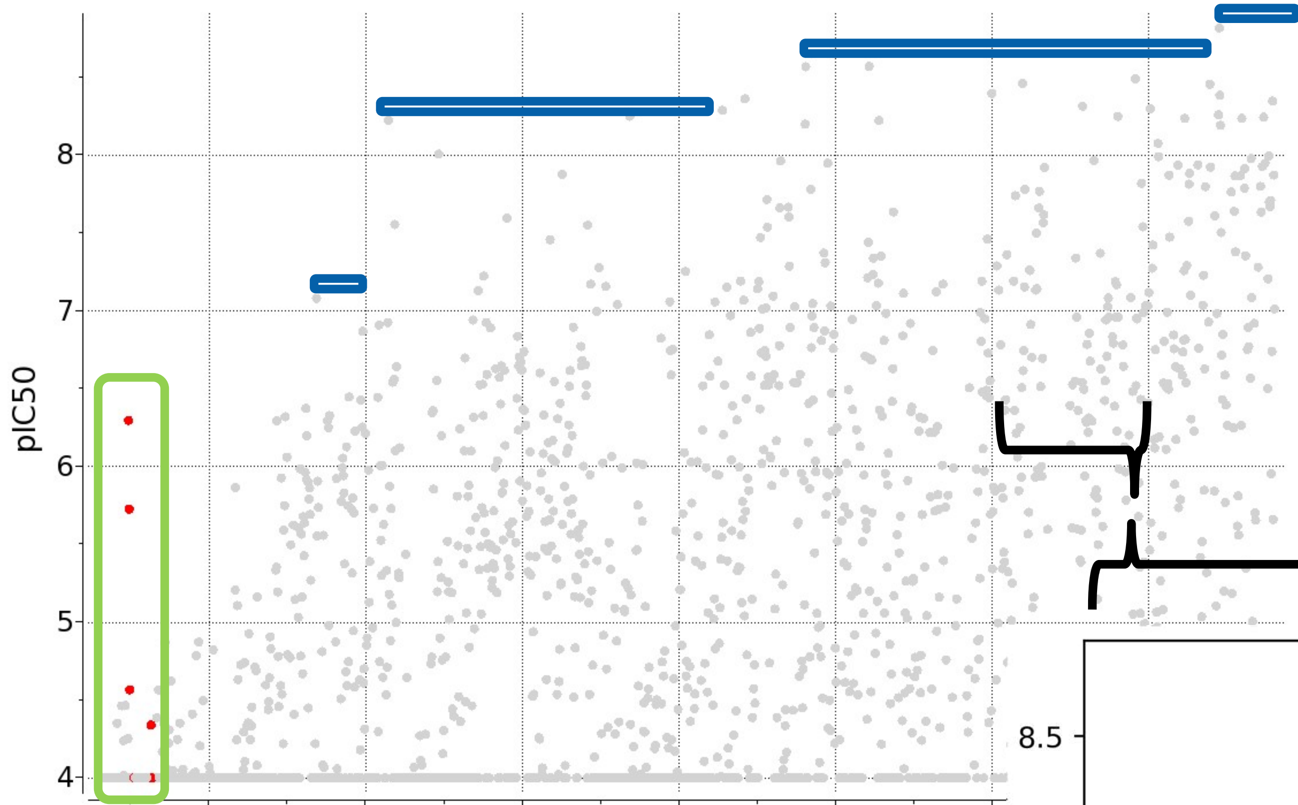
Eric Ma

Outlook

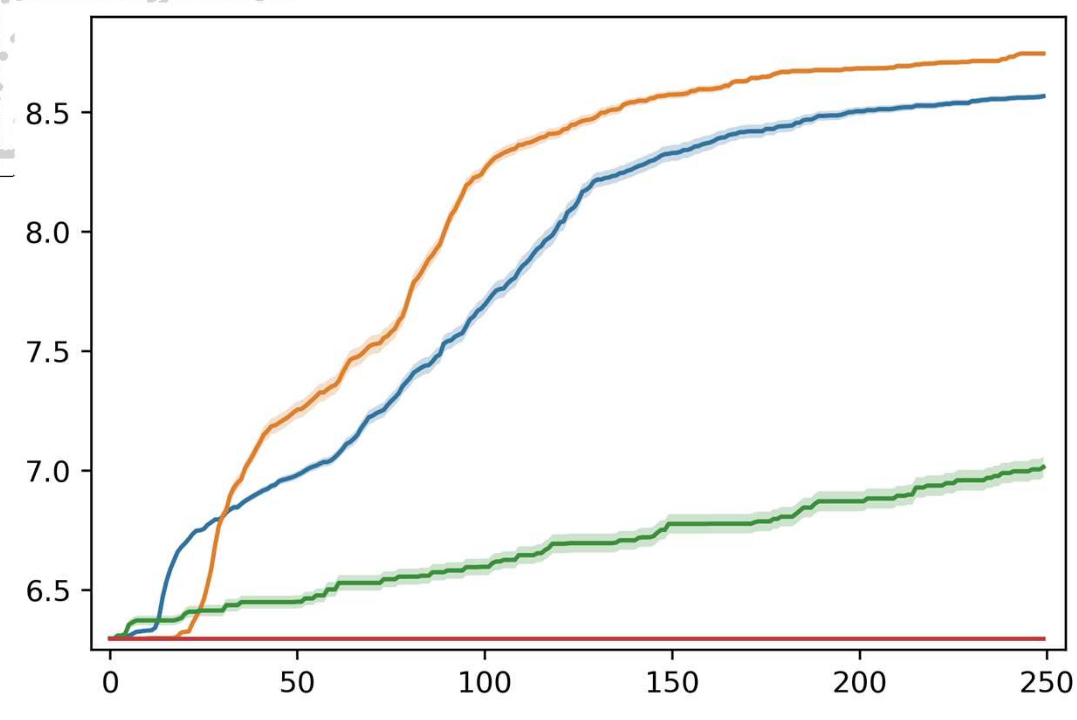
Get error bars on those models

Generate and evaluate virtual chemical space

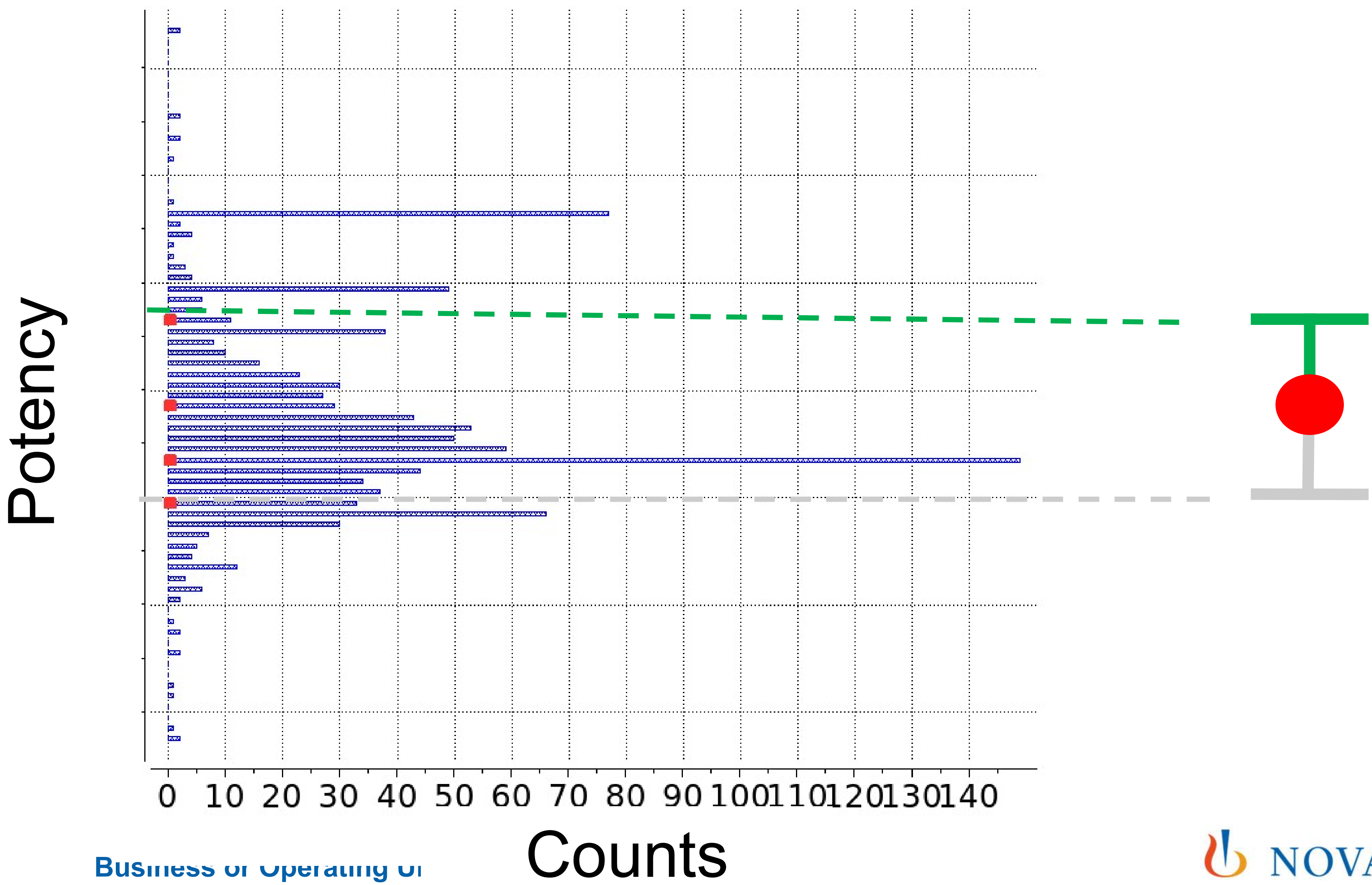
Project Team History



Time (# compounds)



Random forest confidence intervals

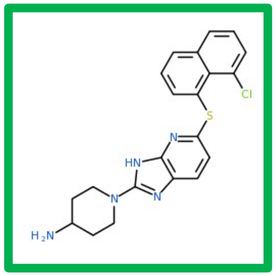


Adjust potency by ADME

$$Relative\ Desirability = (IC_{50})\left(\frac{1}{Good\ ADME}\right)$$

Rebuild model for each batch

Model #0



Training data

Optimistic selection in Round 0 ..

... while the data is pending ...

... is treated as inactive in Round 1

Model #1

Training data