7th Advanced In silico Drug Design workshop/challenge

29 January - 2 February 2024
Olomouc, Czech Republic

Univerzita Palackého v Olomouci

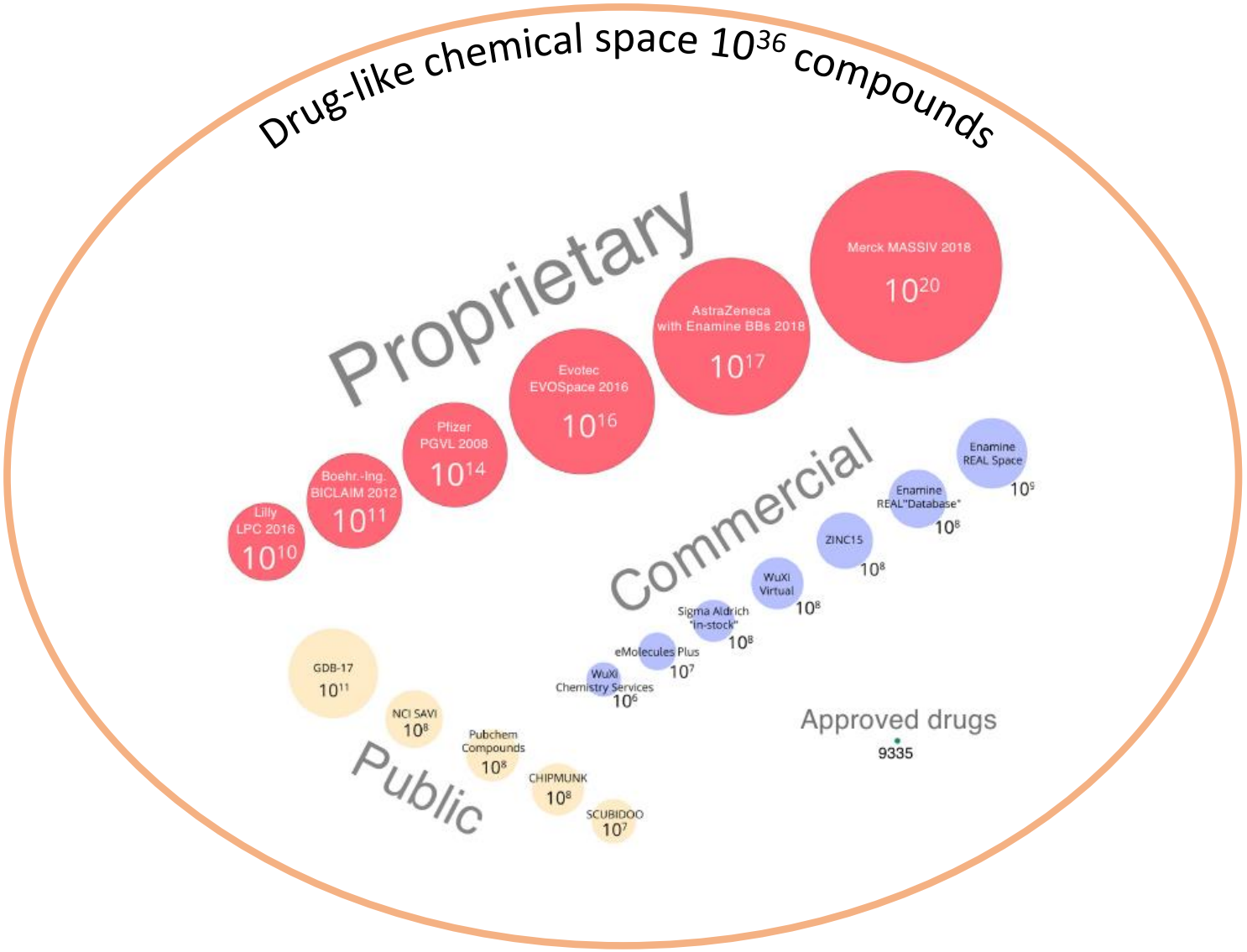
De novo drug design

Pavel Polishchuk

Institute of Molecular and Translational Medicine
Faculty of Medicine and Dentistry
Palacky University

pavlo.polishchuk@upol.cz
qsar4u.com

Drug-like chemical space

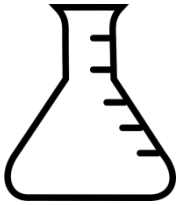


Hoffmann, T.; Gastreich, M., The next level in chemical space navigation: going far beyond enumerable compound libraries. *Drug Discovery Today* **2019**, 24, 1148-1156.
Polishchuk, P. G.; Madzhidov, T. I.; Varnek, A., Estimation of the size of drug-like chemical space based on GDB-17 data. *Journal of Computer-Aided Molecular Design* **2013**, 27, 675-679.

Virtual screening vs. de novo design

Virtual screening

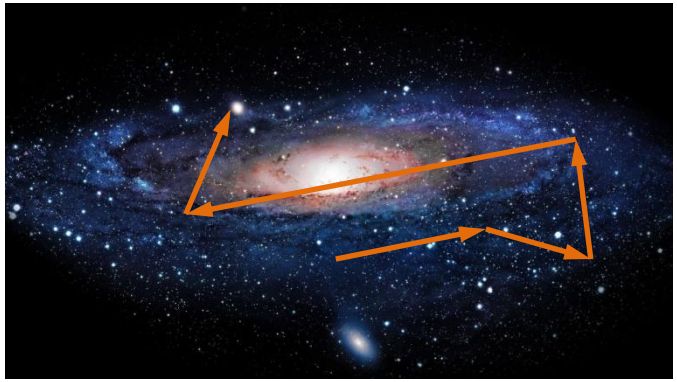
Model

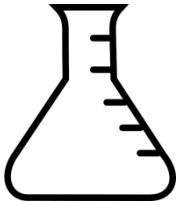


10^9-10^{12} compounds

10-100 compounds

De novo design





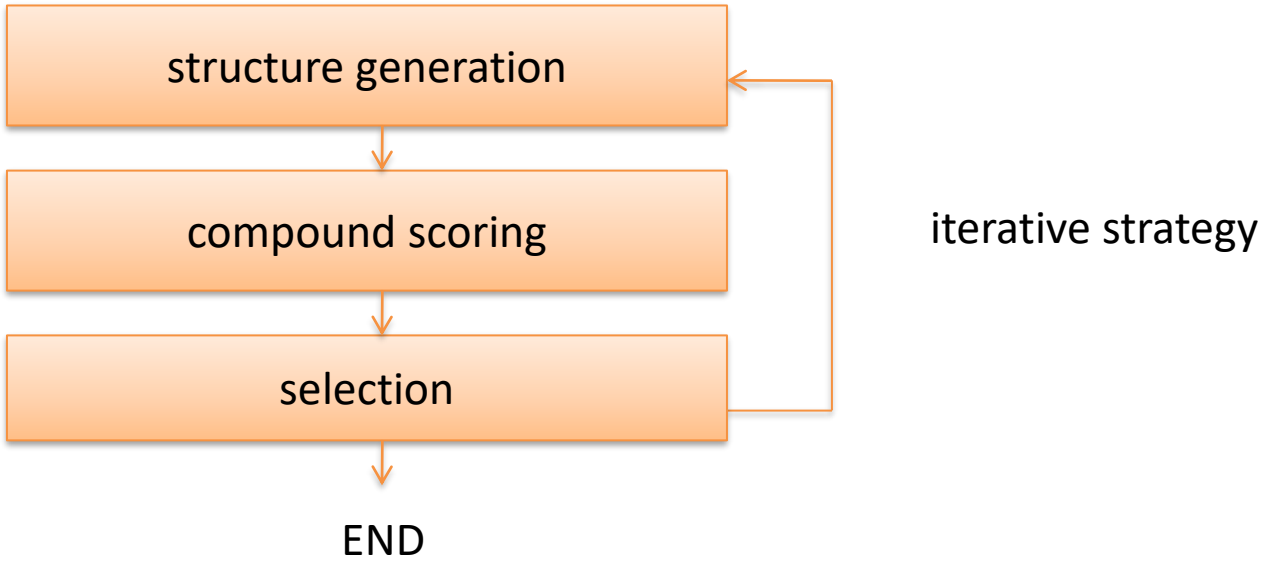
~10^{36} drug-like compounds

10-100 compounds

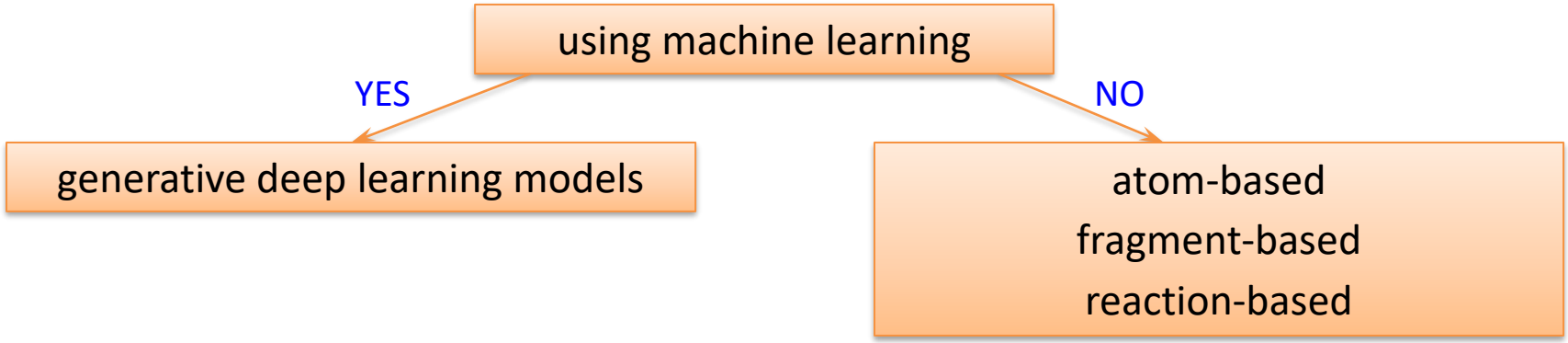
Model

Components of de novo design

1. **Structure generation** - how to create/assembly new structures
2. **Compound scoring** - how to estimate/predict a property of a compound
3. **Search strategy** - how to find compounds with optimal properties



De novo structure generation



- **atom-based** - uses simple rules like add/change/remove atom/bond to perturb structures
- **fragment-based** - uses fragment library to create structures
- **reaction-based** - uses a set of reaction rules and a library of reactants

Atom-based structure generation

Molpher

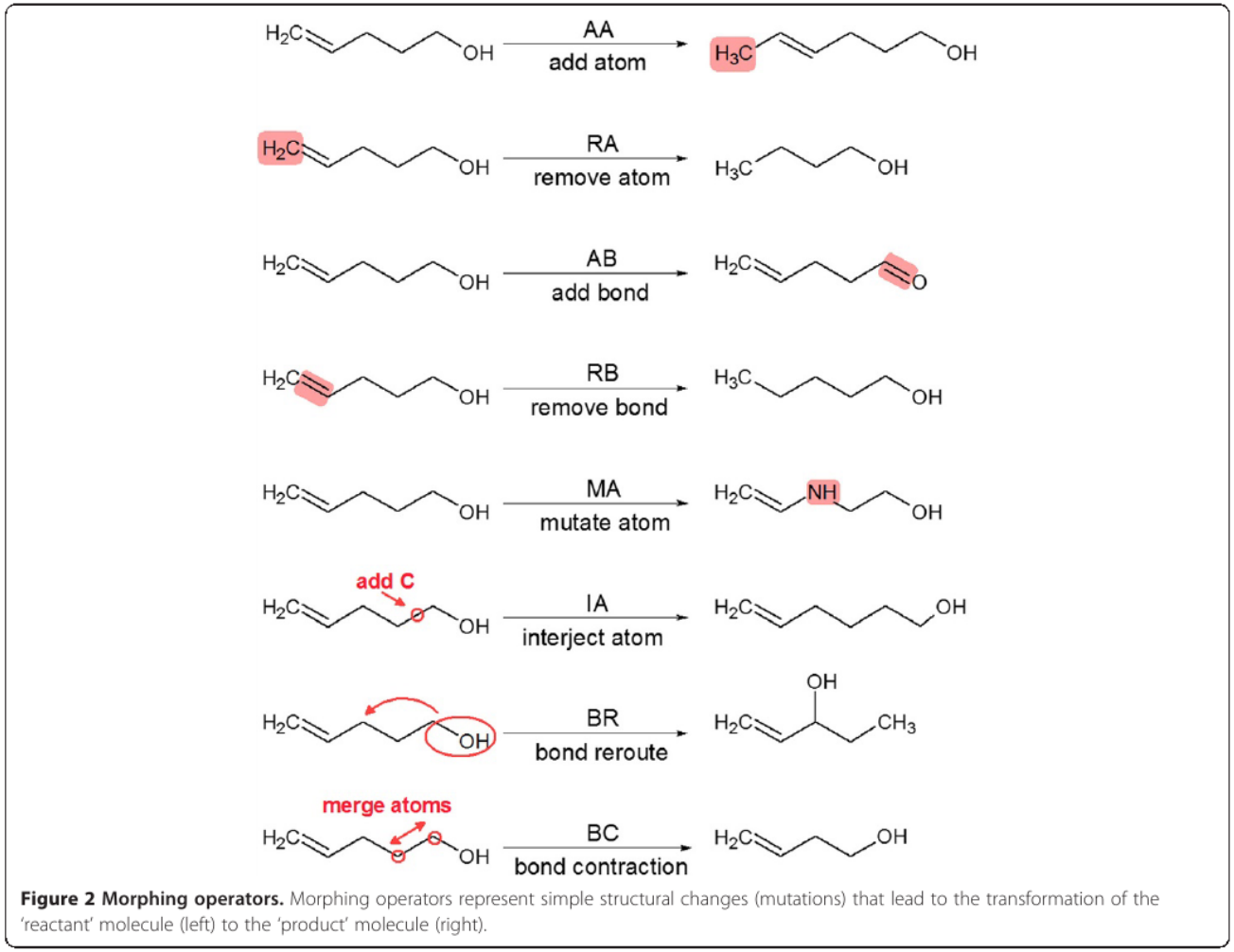


Figure 2 Morphing operators. Morphing operators represent simple structural changes (mutations) that lead to the transformation of the 'reactant' molecule (left) to the 'product' molecule (right).

Hoksza, D.; Škoda, P.; Voršilák, M.; Svozil, D., Molpher: a software framework for systematic chemical space exploration. *Journal of Cheminformatics* **2014**, 6, 7, 10.1186/1758-2946-6-7

Atom-based structure generation

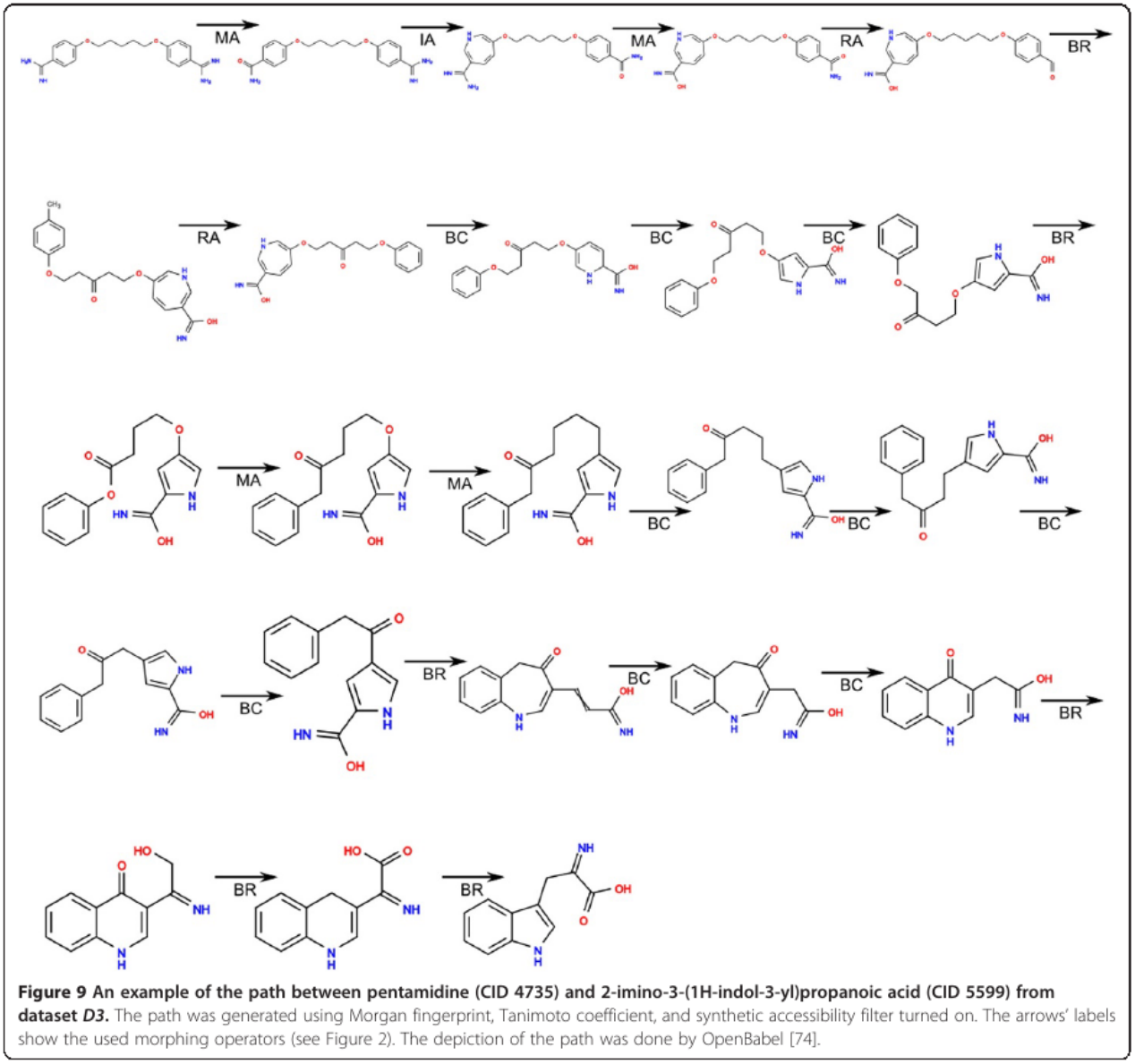
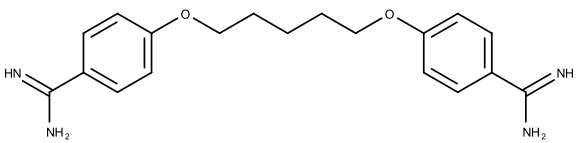
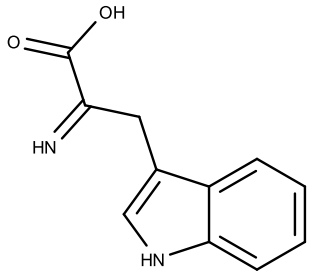


Figure 9 An example of the path between pentamidine (CID 4735) and 2-imino-3-(1H-indol-3-yl)propanoic acid (CID 5599) from dataset *D3*. The path was generated using Morgan fingerprint, Tanimoto coefficient, and synthetic accessibility filter turned on. The arrows' labels show the used morphing operators (see Figure 2). The depiction of the path was done by OpenBabel [74].



22 steps



Hoksza, D.; Škoda, P.; Voršilák, M.; Svozil, D., Molpher: a software framework for systematic chemical space exploration. *Journal of Cheminformatics* **2014**, 6, 7, 10.1186/1758-2946-6-7

IMTM

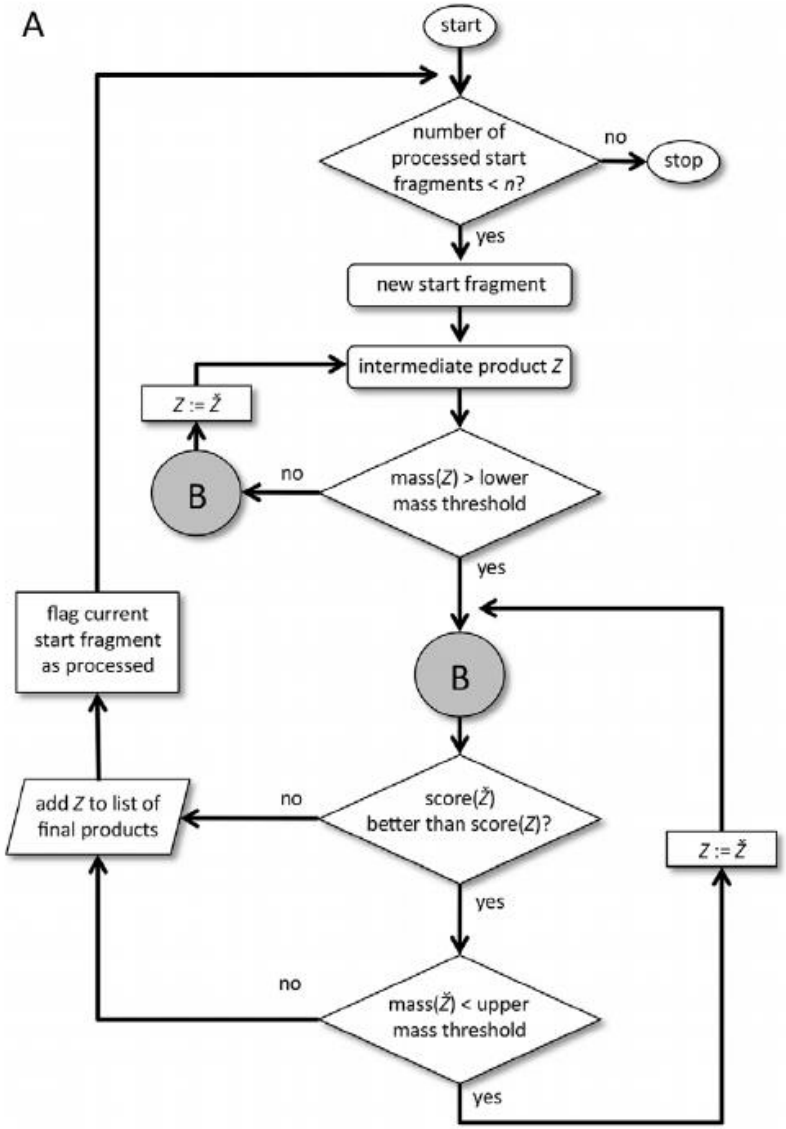
Atom-based structure generation

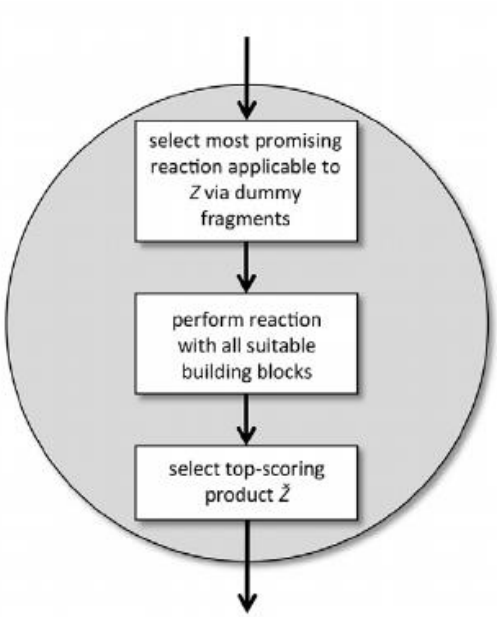
parameters	atom-based
exhaustiveness of chemical space search	++++ very small steps; more suitable for systematic exploration of local chemical space
structure novelty	+++*
structure diversity	+++*
chemically valid structures	-
synthetically feasible	---
combinatorial explosion / time consuming	---

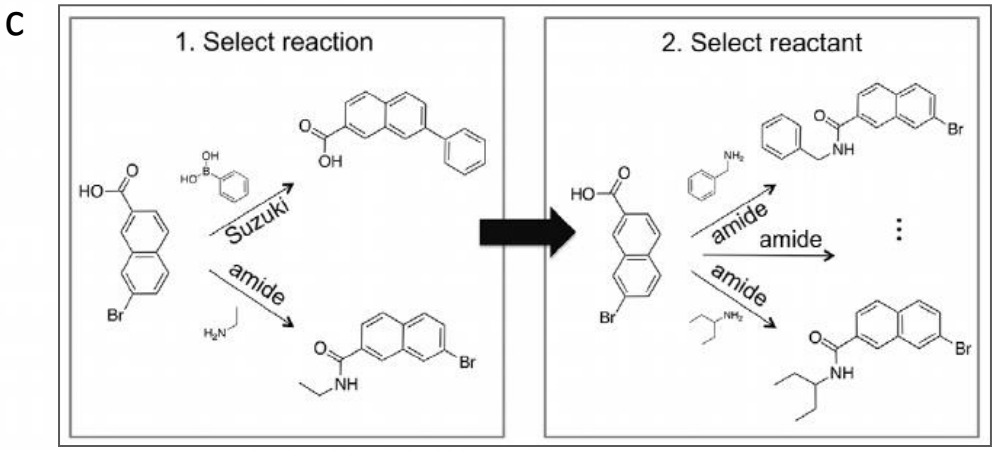
atom-based ≈ *ab initio*

Reaction-based structure generation

DOGS





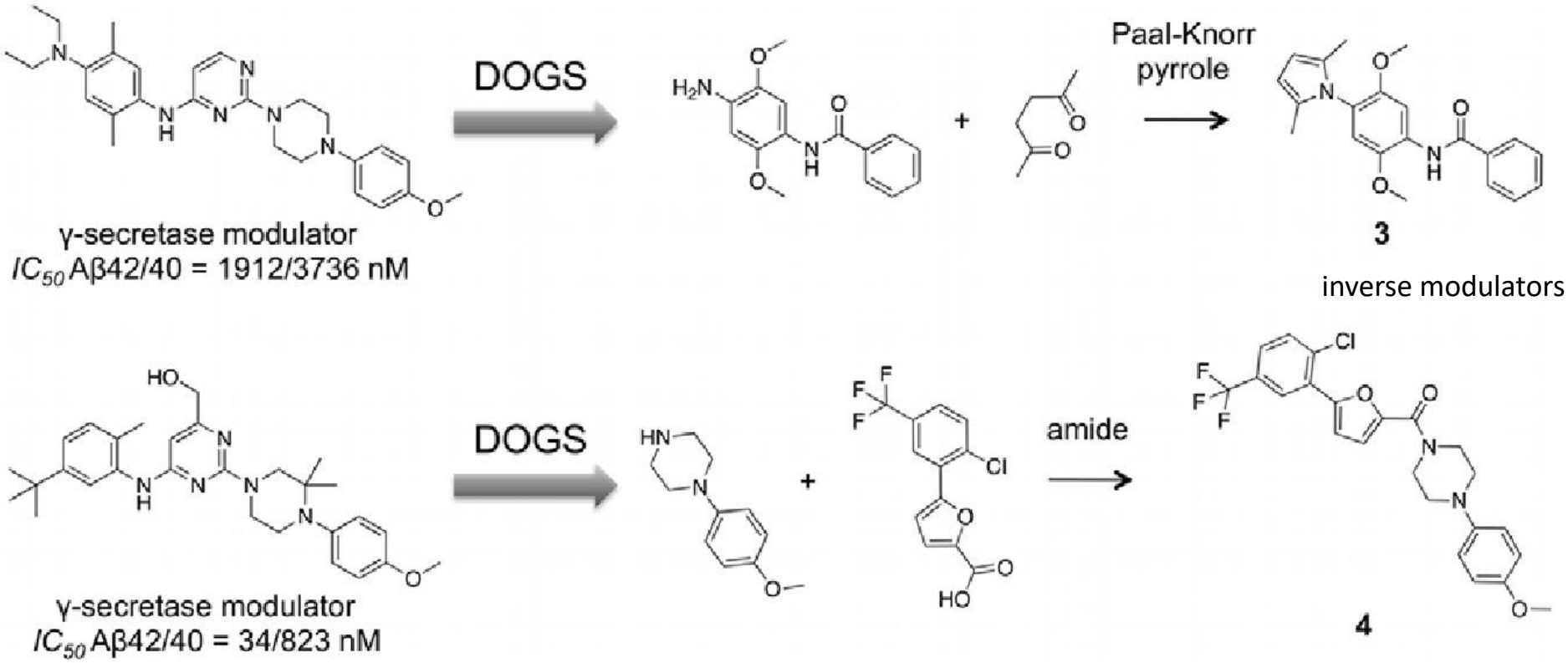


Hartenfeller, M.; Zettl, H.; Walter, M.; Rupp, M.; Reisen, F.; Proschak, E.; Weggen, S.; Stark, H.; Schneider, G., DOGS: Reaction-Driven de novo Design of Bioactive Compounds. *PLOS Computational Biology* **2012**, 8, e1002380.

Reaction-based structure generation

DOGS

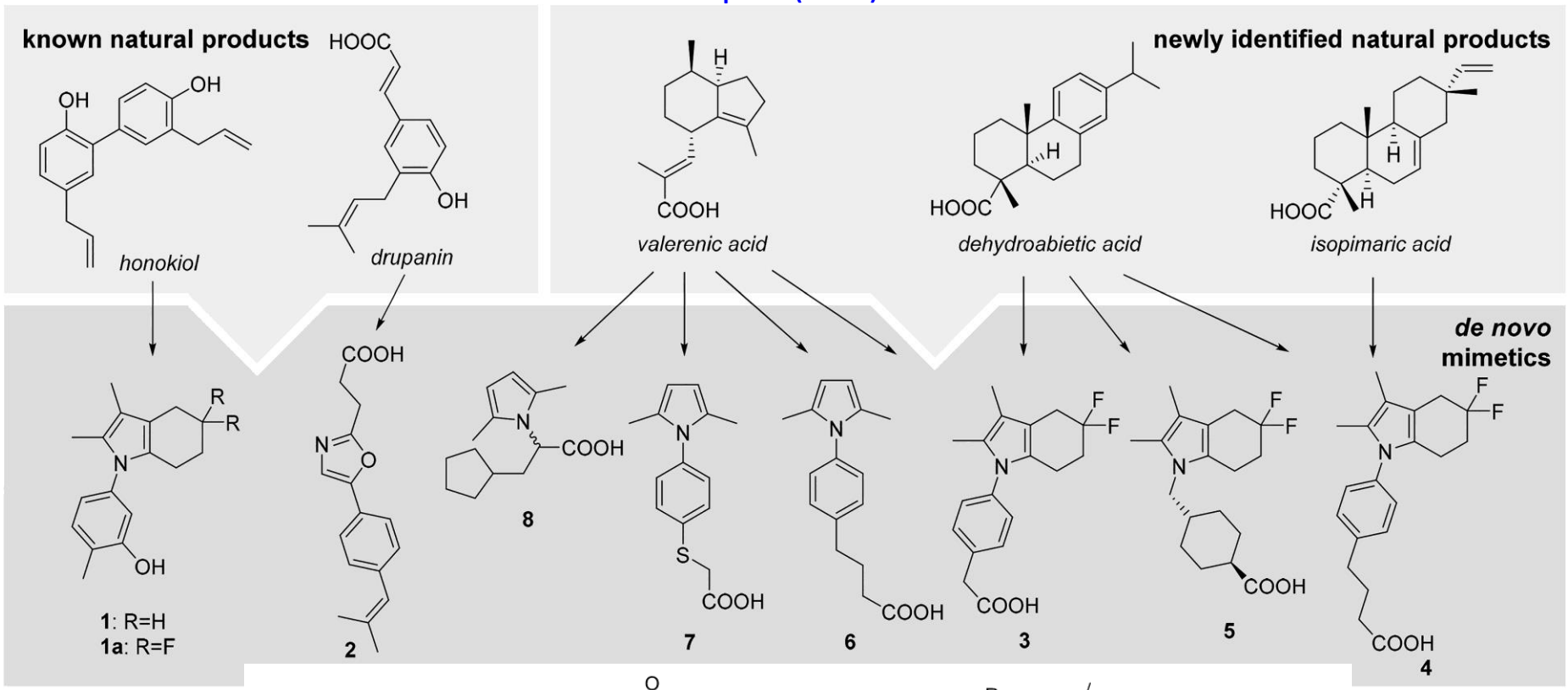
γ-secretase modulators

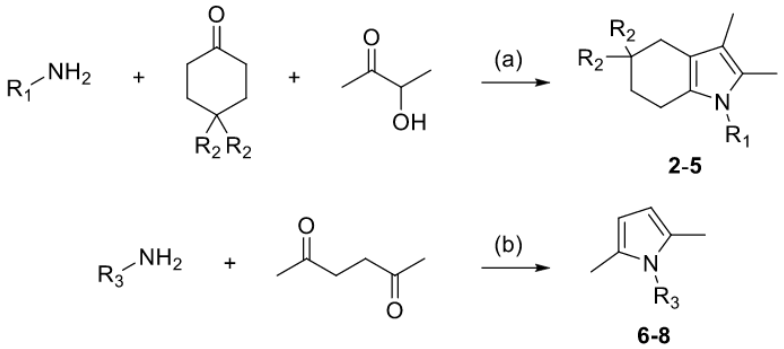


Hartenfeller, M.; Zettl, H.; Walter, M.; Rupp, M.; Reisen, F.; Proschak, E.; Weggen, S.; Stark, H.; Schneider, G., DOGS: Reaction-Driven de novo Design of Bioactive Compounds. *PLOS Computational Biology* **2012**, 8, e1002380.

Reaction-based structure generation

Retinoid X Receptor(RXR) Modulators





isopimaric
dehydroabi
valerenic a
sclareol
conocarpar

Supporting figure 5: Synthesis of de novo mimetics **1a** and **3-8**. Reagents and conditions: (a) EtOH, HOAc, μw, 100°C, 3-6 h, 43-78%; (b) montmorillonite K10, μw, 90°C, 30 min, 41-85%.

Merk D., et al. *J. Med. Chem.*, **2018**, 61 (12), pp 5442–5447

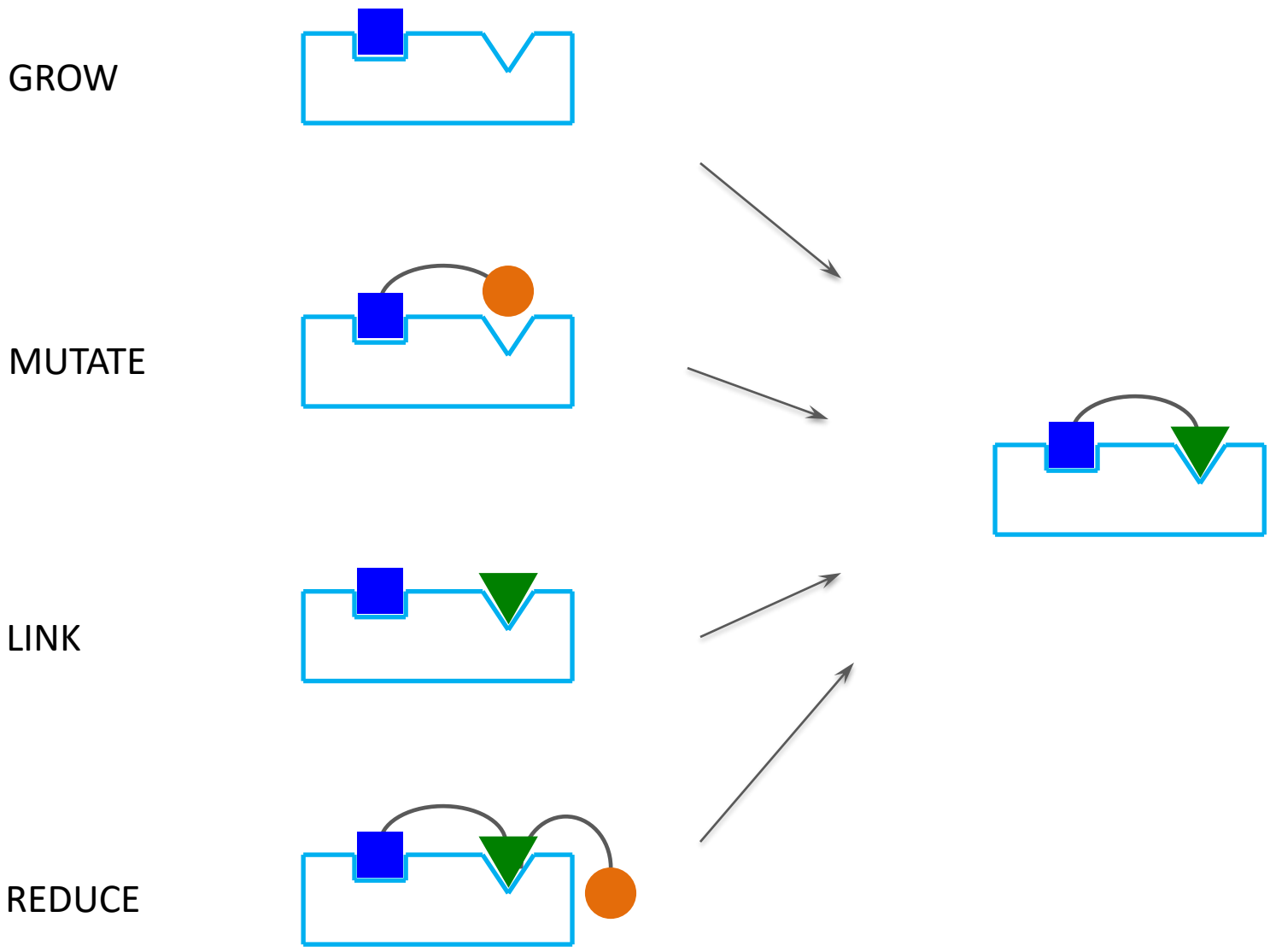
IMTM

Reaction-based structure generation

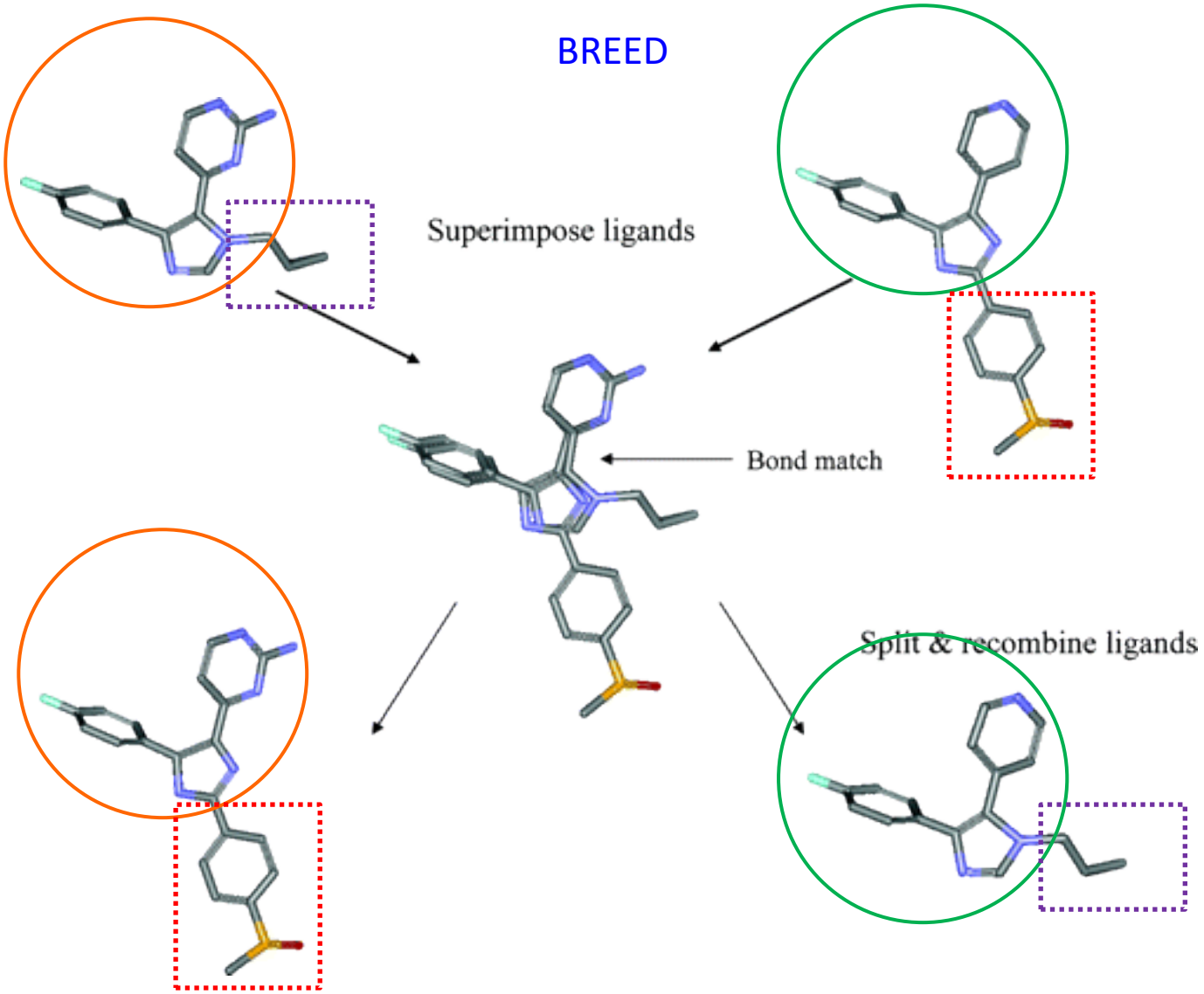
	reaction-based
exhaustiveness of chemical space search	+ depends on reactant library and reaction rules; only grow molecules
structure novelty	++
structure diversity	++
chemically valid structures	+++
synthetically feasible	+++
combinatorial explosion / time consuming	+++

reaction-based ≈ empirical

Fragment-based structure generation



Fragment-based structure generation



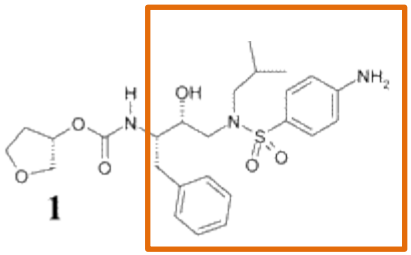
Pierce A.C., Rao G., Bemis G.W. *J. Med. Chem.*, **2004**, 47 (11), pp 2768–2775

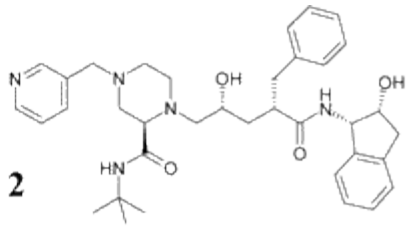
Fragment-based structure generation

BREED: HIV-1 protease inhibitors

known

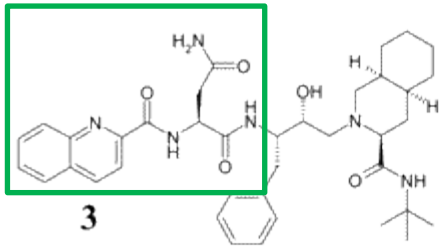
K_i = 0.4-0.6 nM

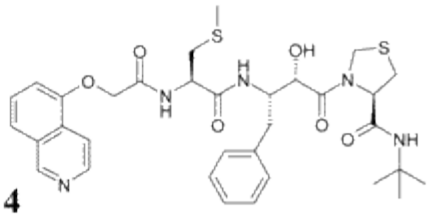




K_d = 1.1 nM

K_i = 1.7 nM

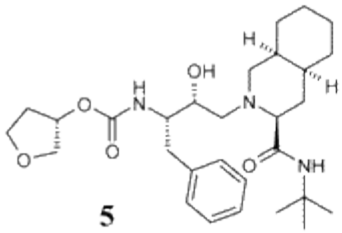


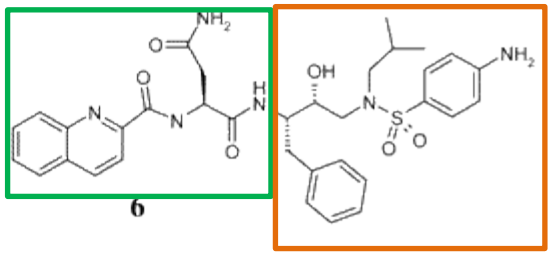


K_d = 0.3 nM

designed

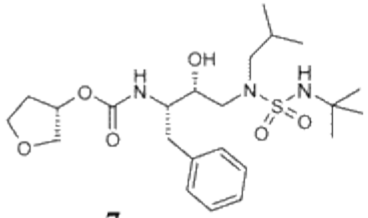
IC_{50} = 160 nM

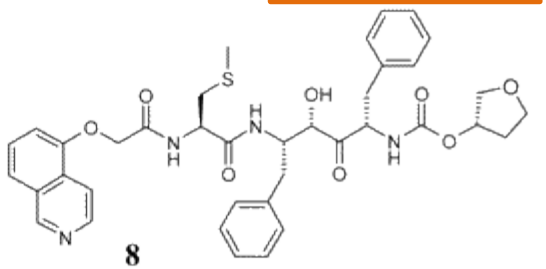




K_i = 0.1 nM

K_i = 42 nM

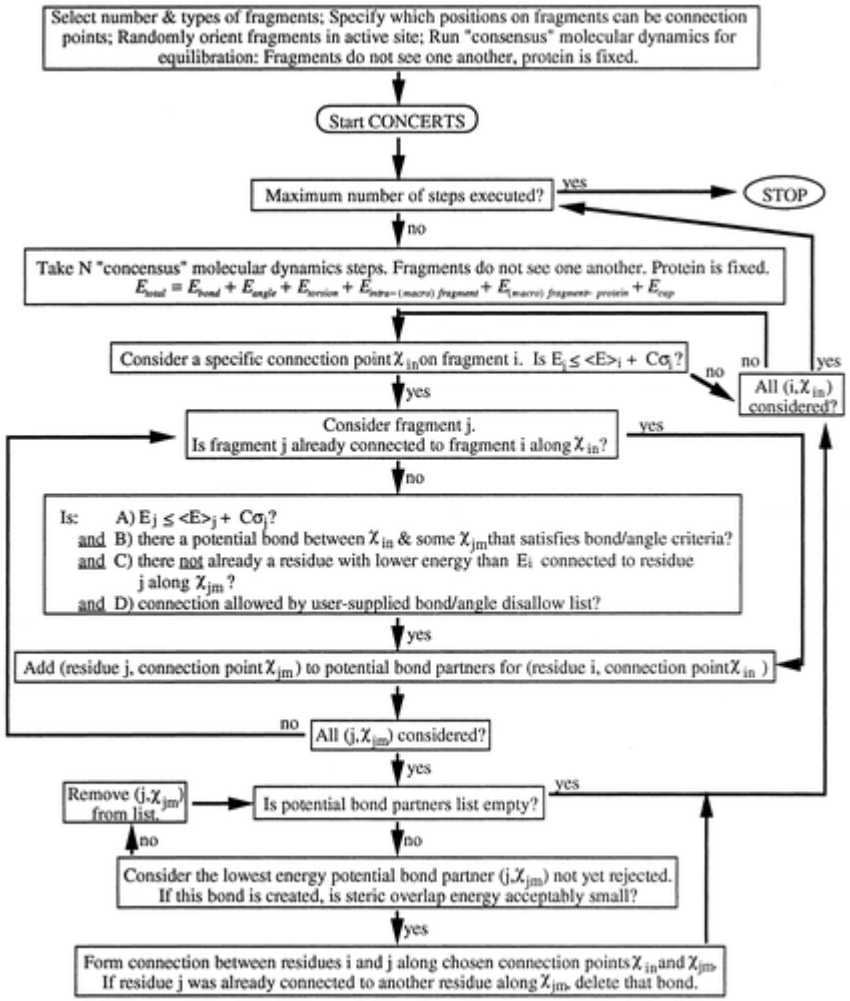




Pierce A.C., Rao G., Bemis G.W. *J. Med. Chem.*, **2004**, 47 (11), pp 2768–2775

Fragment-based structure generation

CONCEPTS



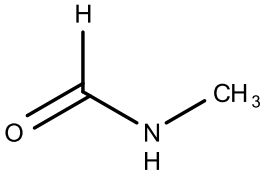
MD of fragments which are linking or breaking during the simulation in order to create more favorable structures

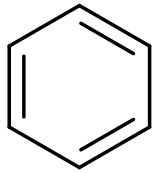
formation of certain bonds was forbidden:
O–O, N–N, N–O, S–O, O–C–O, O–N–O, N–C–N, C_α–C_α, C–C_α–C

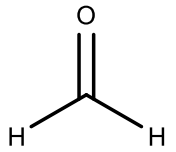
Pearlman D.A., Murcko M.A. *J. Med. Chem.*, **1996**, 39 (8), pp 1651–1663

Fragment-based structure generation

CONCEPTS: HIV-1 protease inhibitors

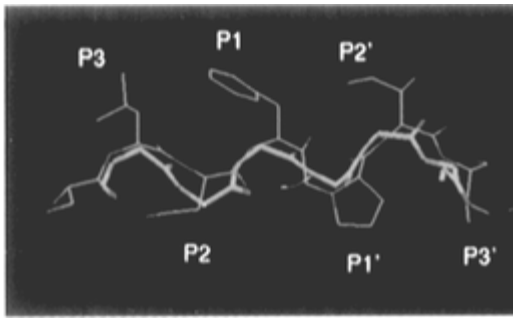


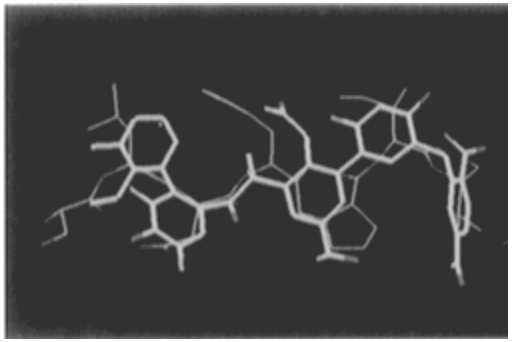




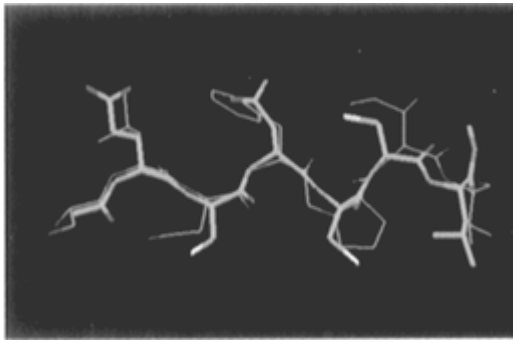
CH_4 H_2O

NH_3





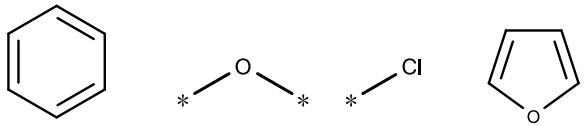
+ 19 side chains

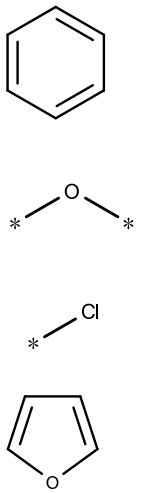


Pearlman D.A., Murcko M.A. *J. Med. Chem.*, **1996**, 39 (8), pp 1651–1663

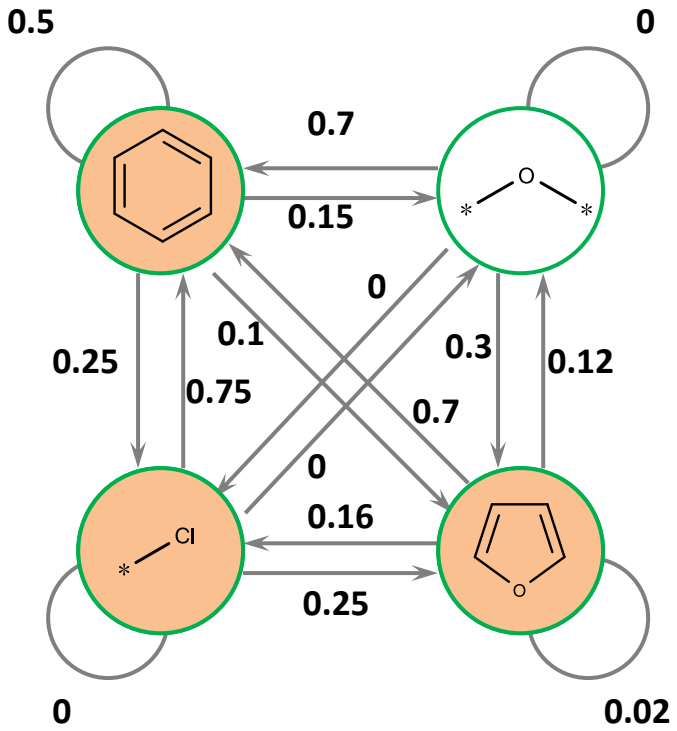
Fragment-based structure generation

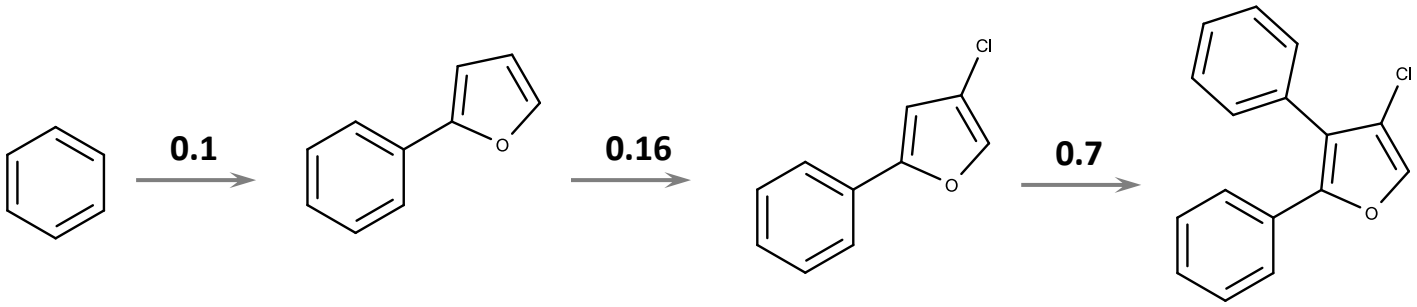
FOG





0.5	0.15	0.25	0.1
0.7	0	0	0.3
0.75	0	0	0.25
0.7	0.12	0.16	0.02

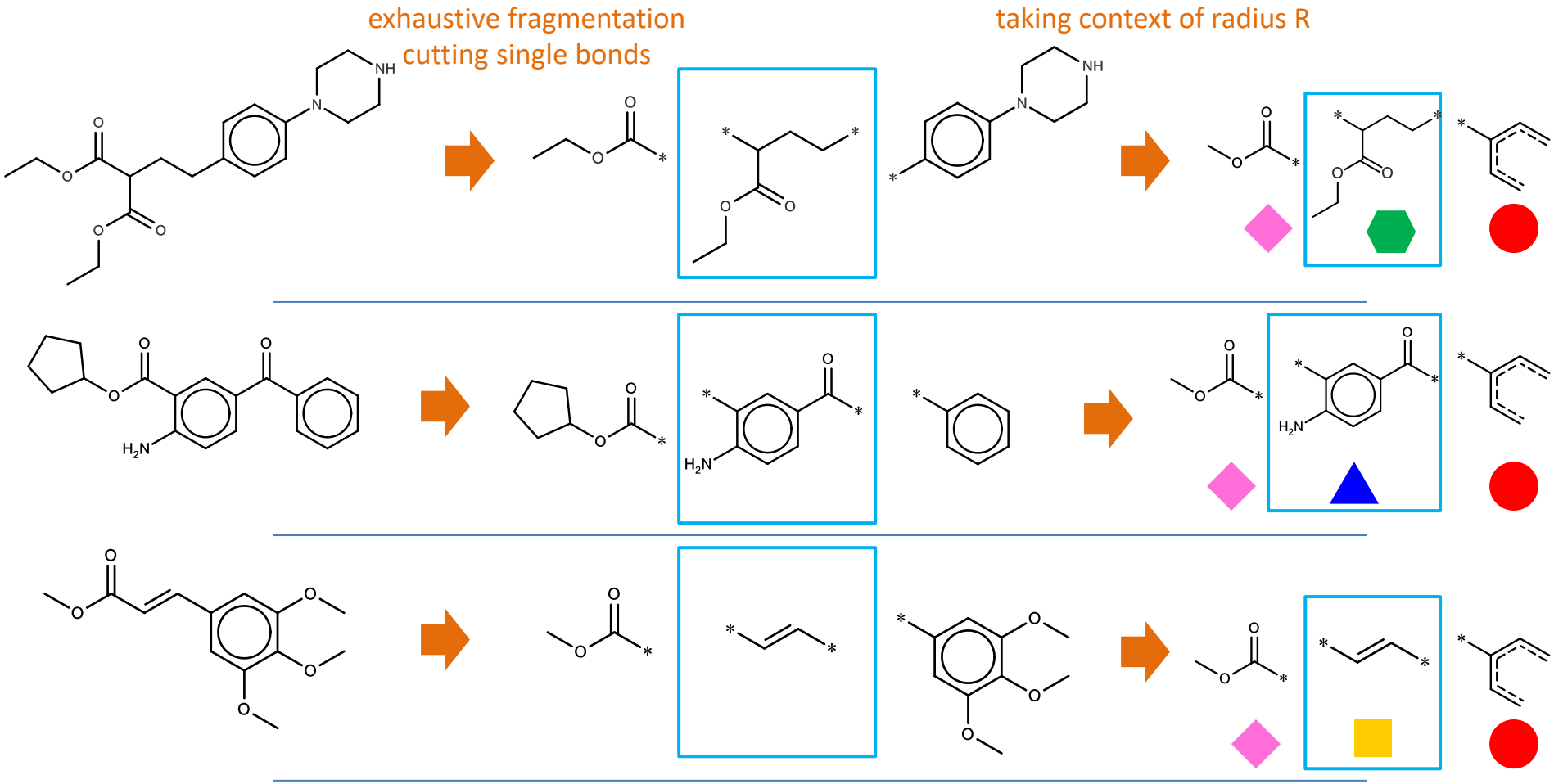




Kutchukian, P. S.; Lou, D.; Shakhnovich, E. I., FOG: Fragment Optimized Growth Algorithm for the de Novo Generation of Molecules Occupying Druglike Chemical Space. *Journal of Chemical Information and Modeling* **2009**, 49, 1630-1642.

Fragment-based structure generation

CReM: chemically reasonable mutations



DB of replacements

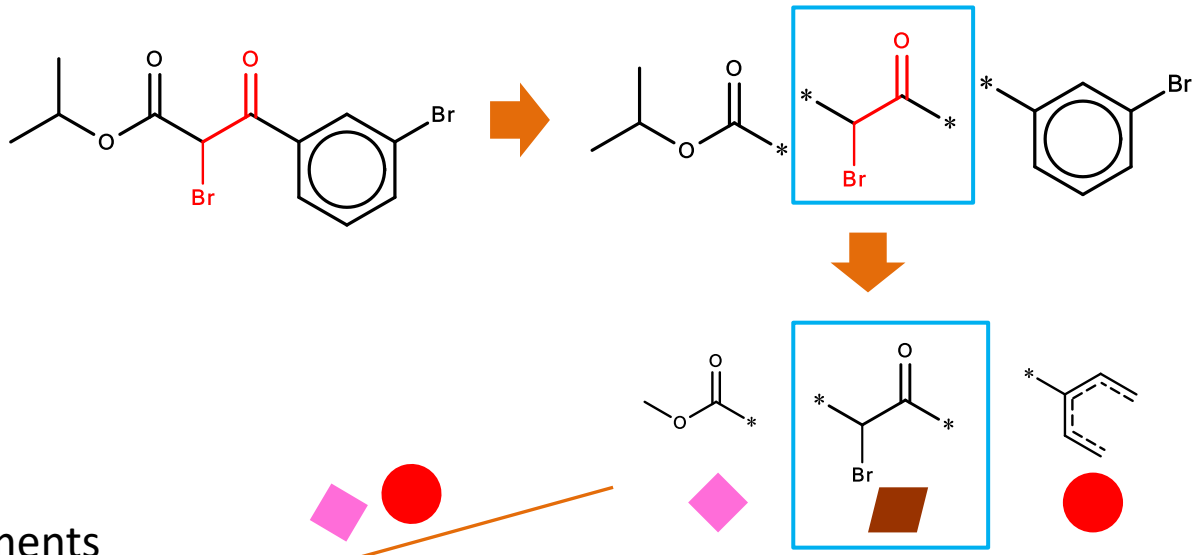
environment (radius = 3)	fragments
◆ ●	⬢ ▲ ■ ...
...	...

mutually exchangeable fragments

Polishchuk, P. CReM: chemically reasonable mutations framework for structure generation. *J. Cheminf.* **2020**, 12 (1), 28.

Fragment-based structure generation

CReM: chemically reasonable mutations



Generated structures are always chemically valid!

Polishchuk, P. CReM: chemically reasonable mutations framework for structure generation. *J. Cheminf.* **2020**, 12 (1), 28.

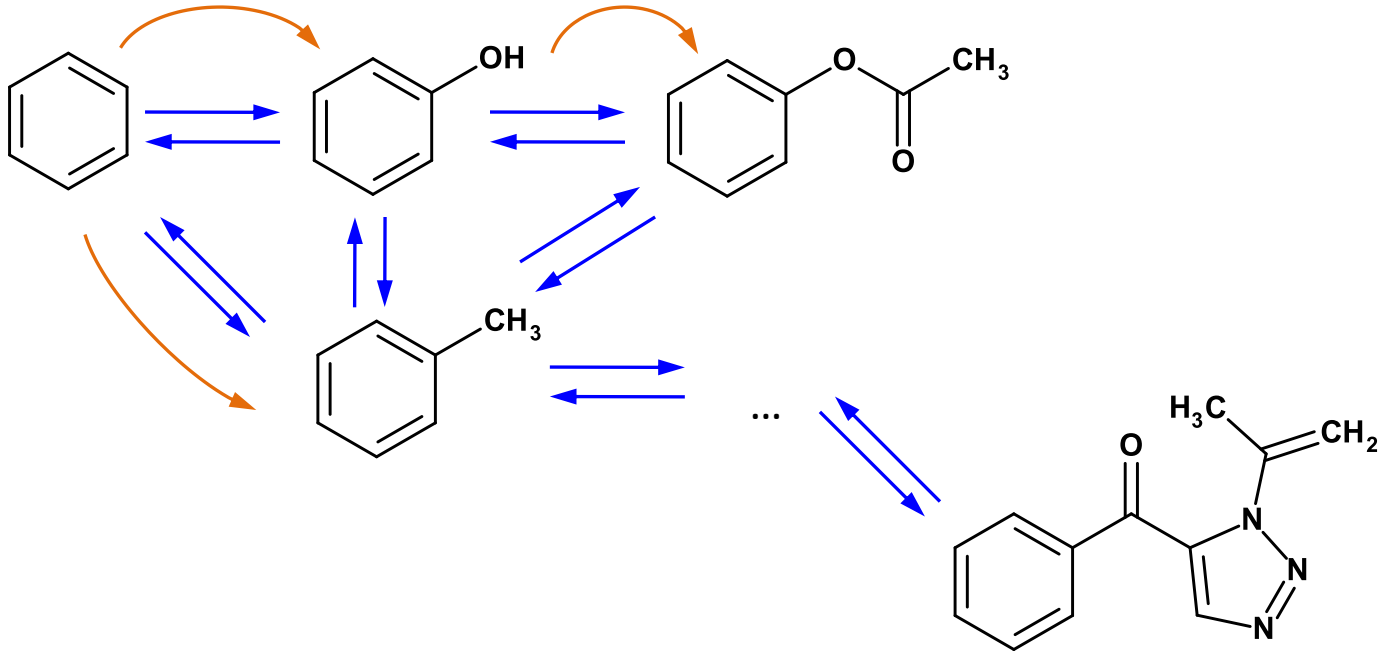
Fragment-based structure generation

	fragment-based
exhaustiveness of chemical space search	++ variable, controlled by the size of fragments to replace
structure novelty	++
structure diversity	++
chemically valid structures	(+++)
synthetically feasible	(++)
combinatorial explosion / time consuming	++

fragment-based ≈ semi-empirical

Reaction-based vs. fragment-based

	Reaction-based	Fragment-based
Prerequisites:	reaction rules set database of building blocks	database of fragments
Abilities & issues:	• molecules are more likely to be feasible • not all moves are allowed • usually only increase complexity • some molecules can be unreachable	• do not control synthetic feasibility • many moves are allowed • arbitrary direction of exploration • cover larger chemical space



IMTM

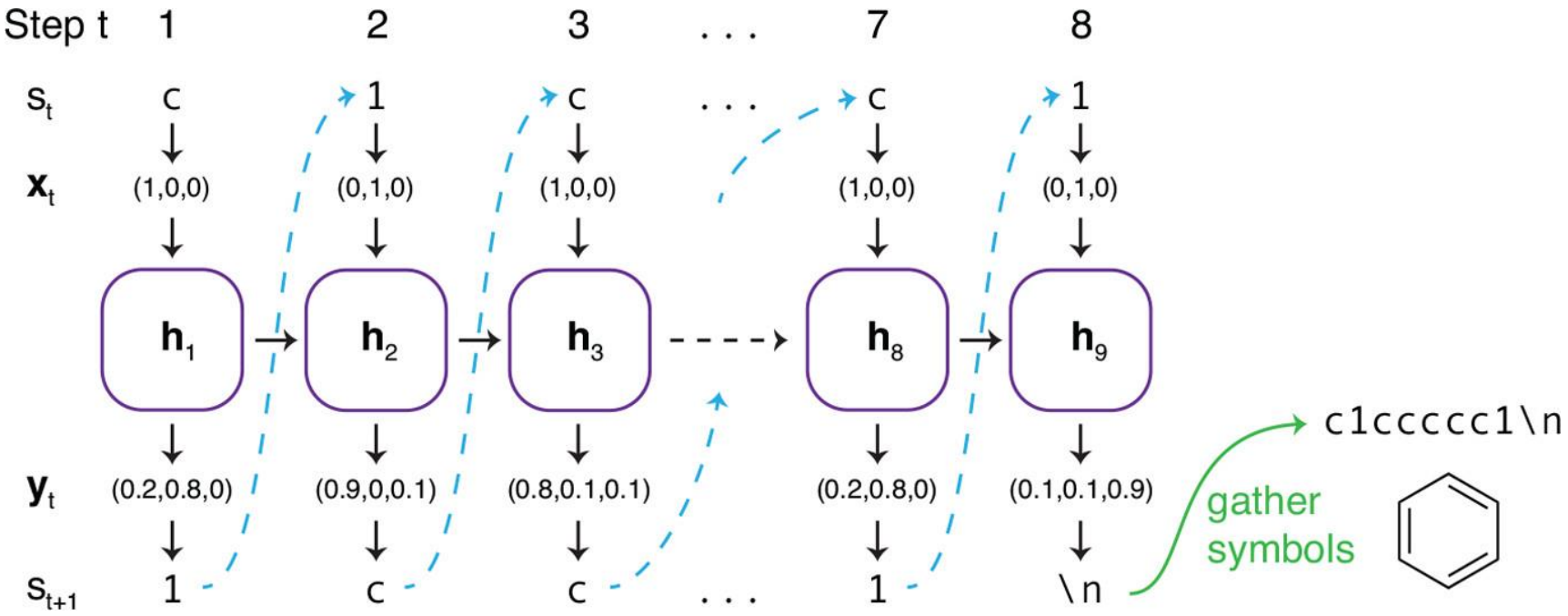
De novo structure generation

Summary

	atom-based	fragment-based	reaction-based
exhaustiveness of chemical space search	++++ very small steps; more suitable for systematic exploration of local chemical space	++ variable, controlled by the size of fragments to replace	+ depends on reactant library and reaction rules; only grow molecules
structure novelty	+++*	++	++
structure diversity	+++*	++	++
chemically valid structures	-	(+++)	+++
synthetically feasible	---	(++)	+++
combinatorial explosion / time consuming	---	++	+++

Deep learning model for structure generation

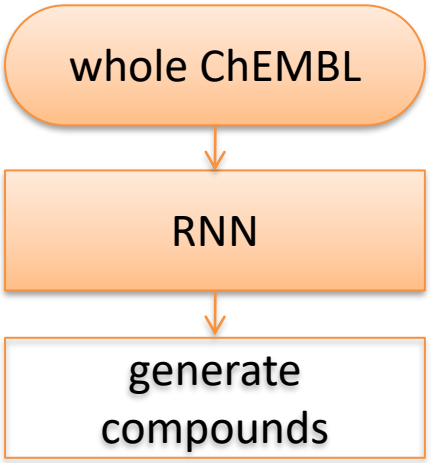
Recurrent neural network (RNN)



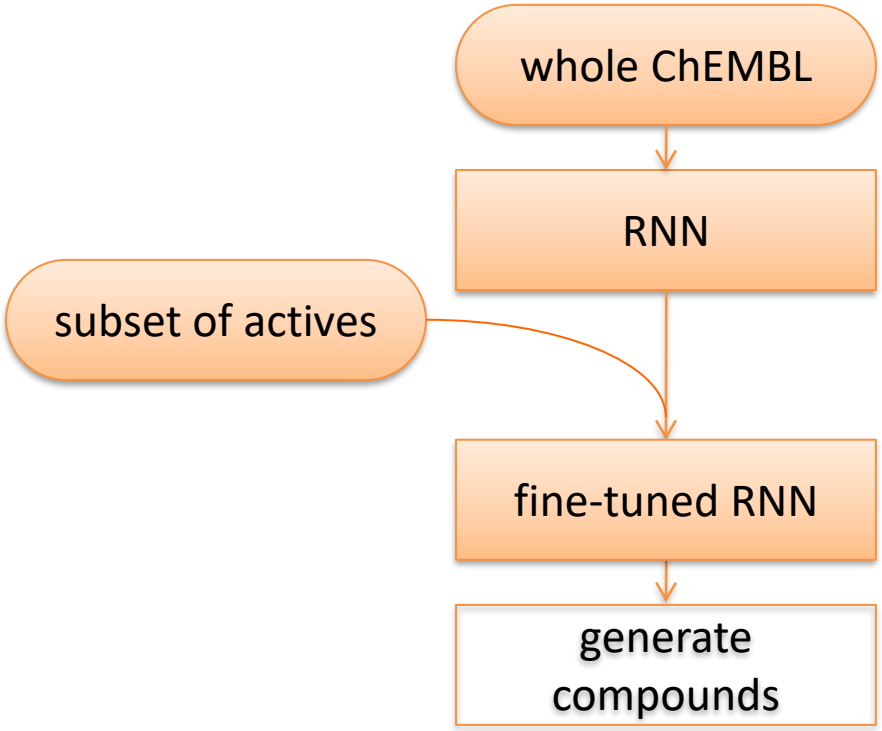
Segler, M. H. S.; Kogej, T.; Tyrchan, C.; Waller, M. P., Generating Focused Molecule Libraries for Drug Discovery with Recurrent Neural Networks. *ACS Central Science* **2018**, 4, 120-131.

Deep learning model for structure generation

unsupervised generation



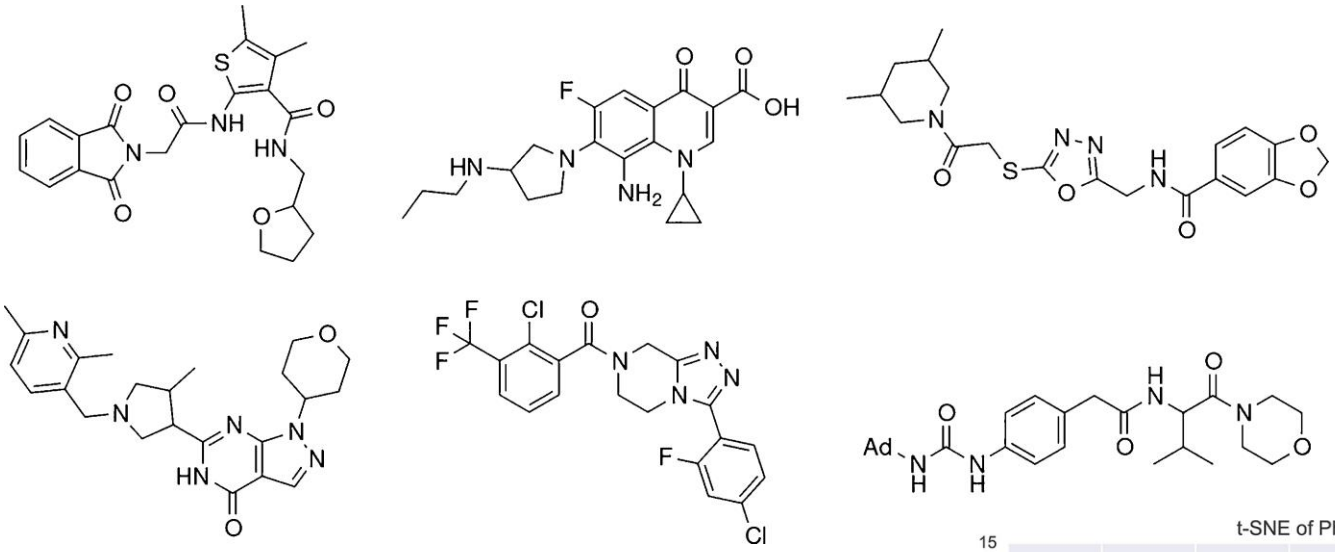
transfer learning



Segler, M. H. S.; Kogej, T.; Tyrchan, C.; Waller, M. P., Generating Focused Molecule Libraries for Drug Discovery with Recurrent Neural Networks. *ACS Central Science* **2018**, 4, 120-131.

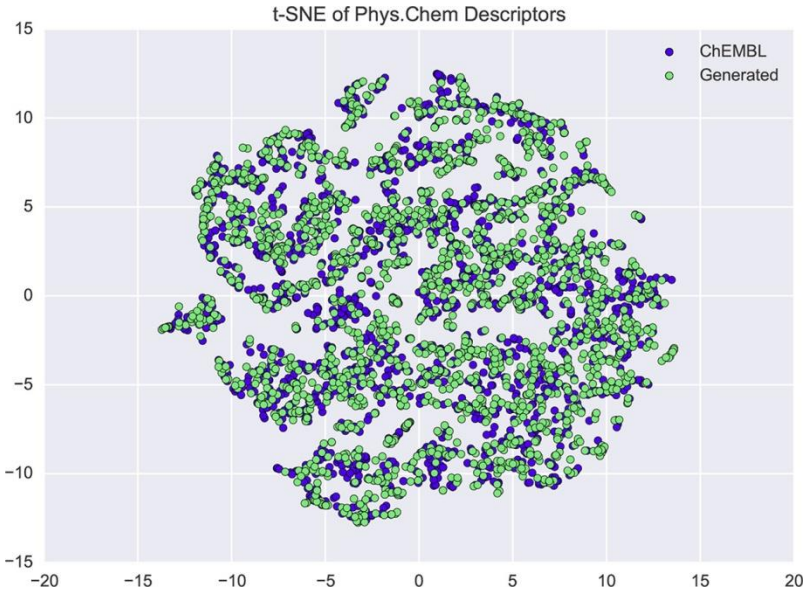
Deep learning model for structure generation

unsupervised generation



976 327 compounds
97.7% chemically valid
11.5% were duplicated with ChEMBL
1.7% of duplicates

75% passed AZ filters (similar to ChEMBL)
12% of scaffolds were common with ChEMBL



Segler, M. H. S.; Kogej, T.; Tyrchan, C.; Waller, M. P., Generating Focused Molecule Libraries for Drug Discovery with Recurrent Neural Networks. *ACS Central Science* **2018**, 4, 120-131.

Deep learning model for structure generation

	deep learning
exhaustiveness of chemical space search	++
structure novelty	++
structure diversity	++
chemically valid structures	++
synthetically feasible	?
combinatorial explosion / time consuming	+++

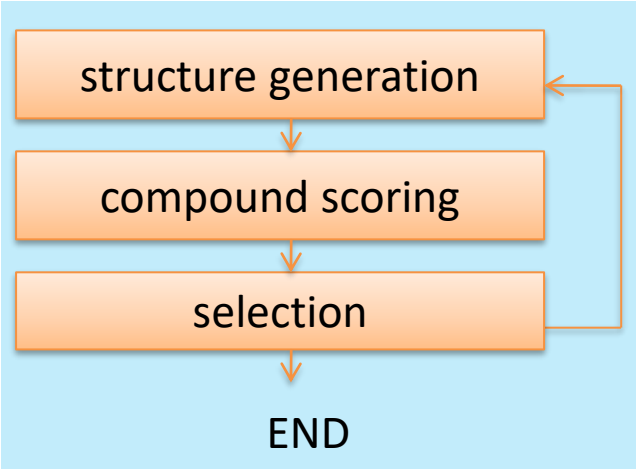
Scoring and objective functions

Can be any but preferably smooth to follow the chemical similarity principle:

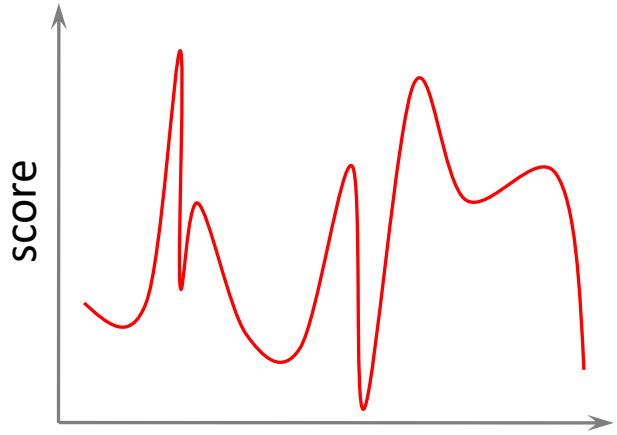
- physicochemical properties
- similarity measures
- QSAR model prediction
- pharmacophore fit
- docking scoring
- molecular dynamics

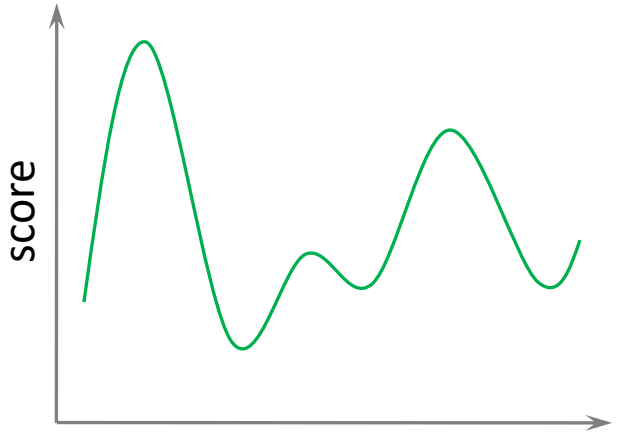
ligand-based scoring functions

structure-based scoring functions



…



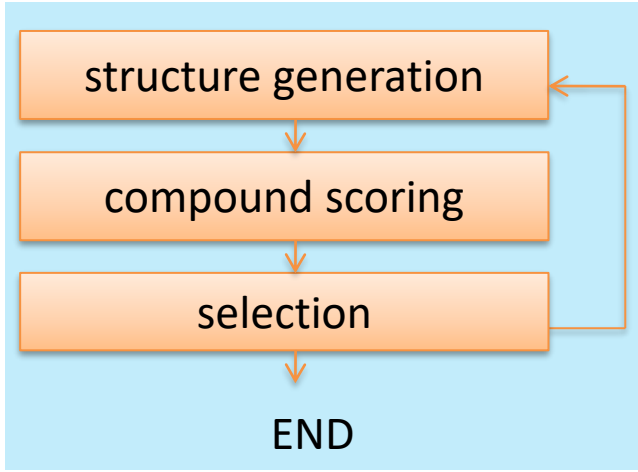


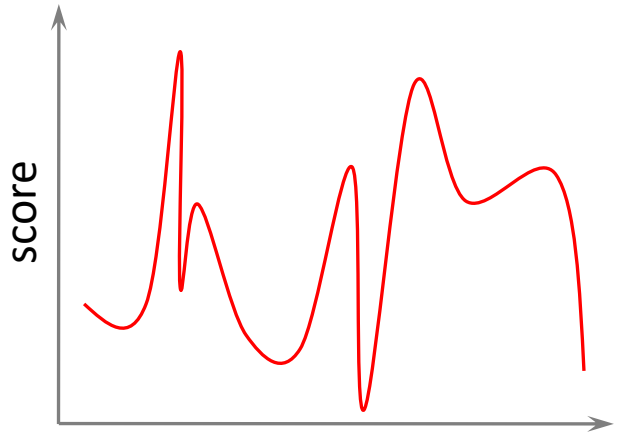
Search algorithms

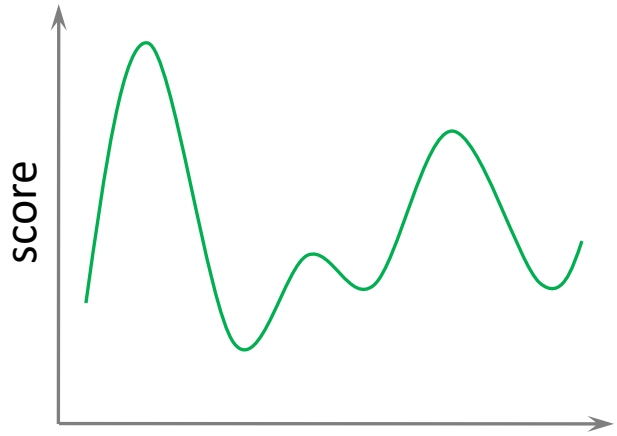
Can be any, for example:

- greedy search
- Monte Carlo
- evolutionary algorithms, e.g.:
  - genetic algorithm
- simulated annealing
- reinforcement learning

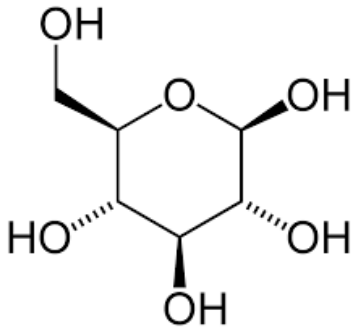
…



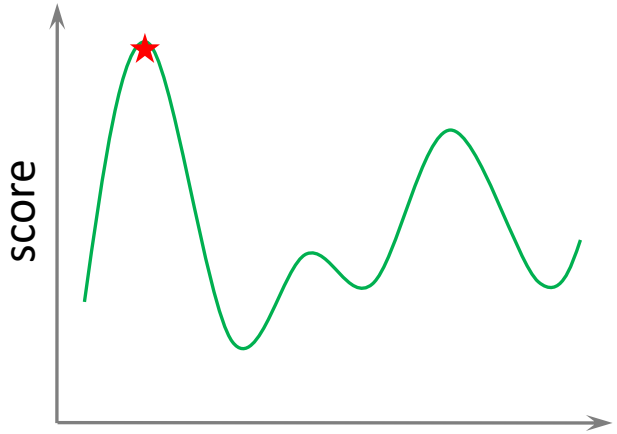




Inverse QSAR



D_1	D_2	D_3	...	D_N
1	0	9	...	1
4	0	1	...	1
0	2	3	...	3
...	...	...	...	...
4	0	0	...	1

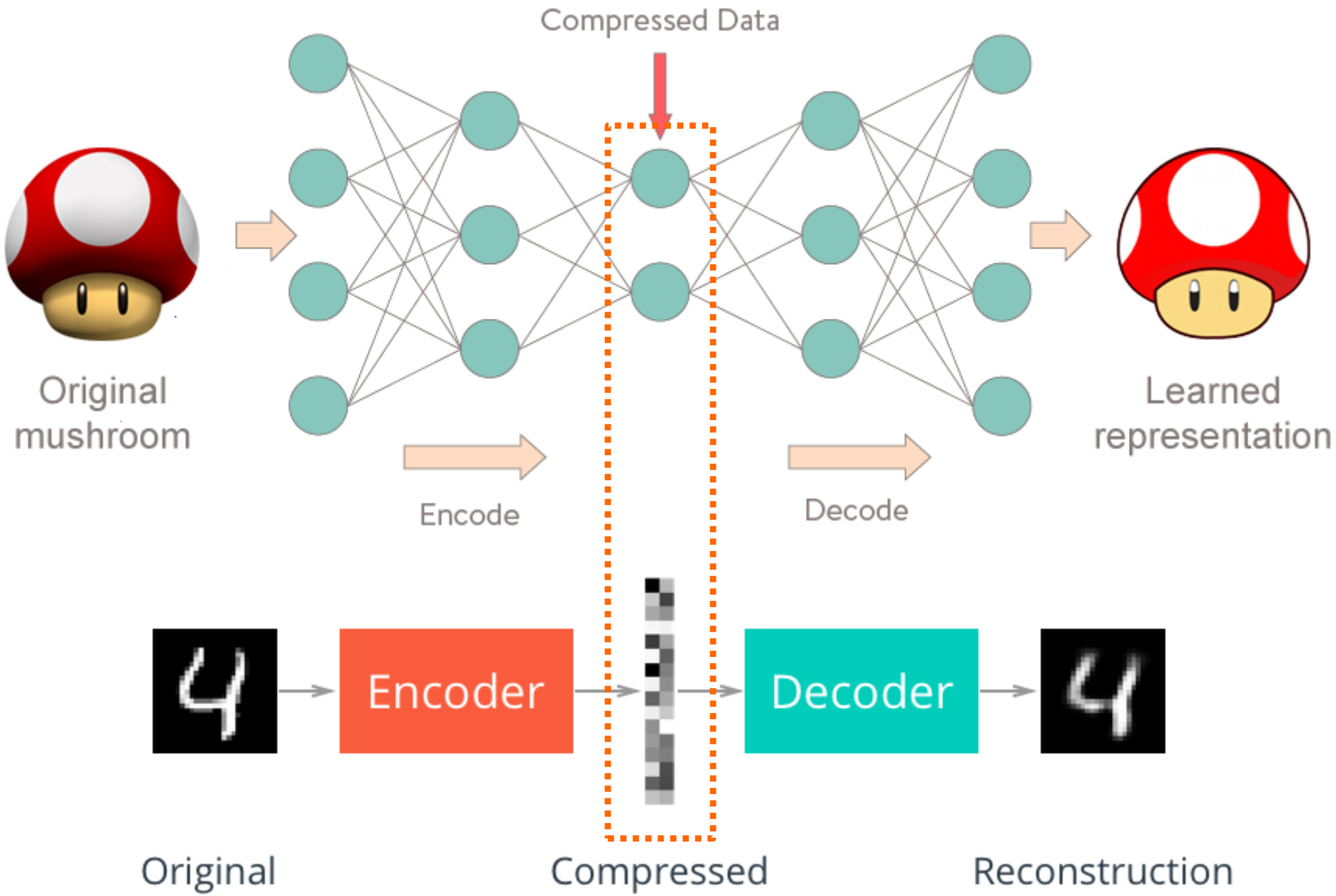


D_1	D_2	D_3	...	D_N
11	3	1	...	15

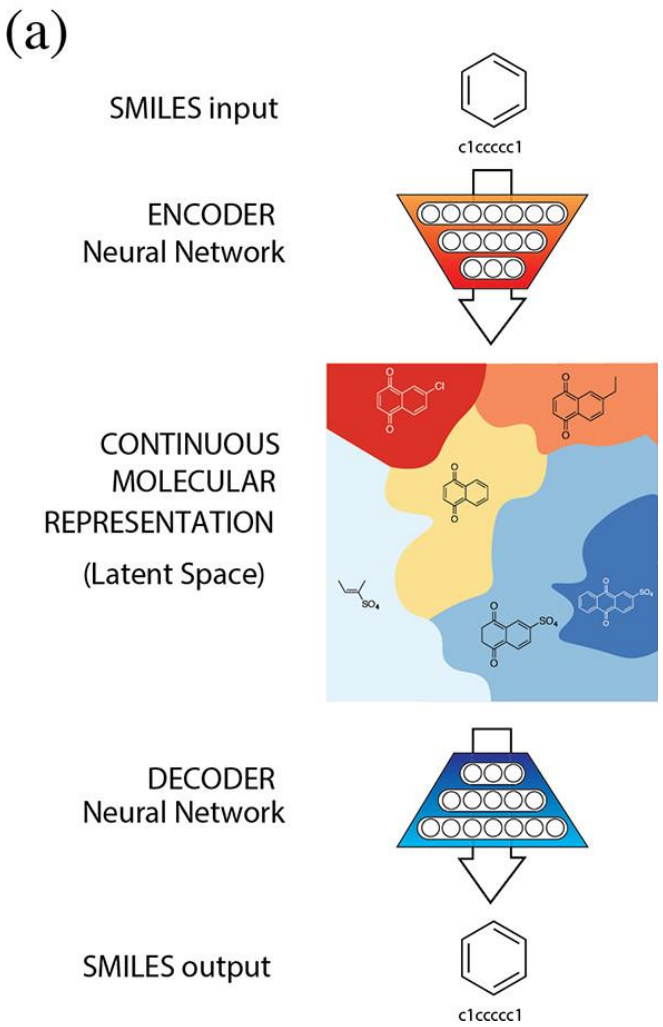
STRUCTURE ?

Inverse QSAR: deep learning

Autoencoder



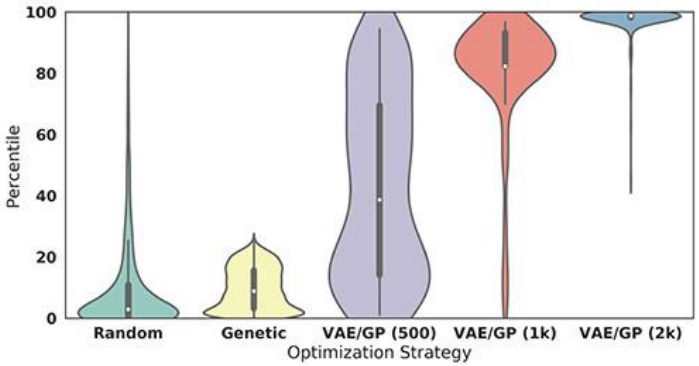
Inverse QSAR: deep learning



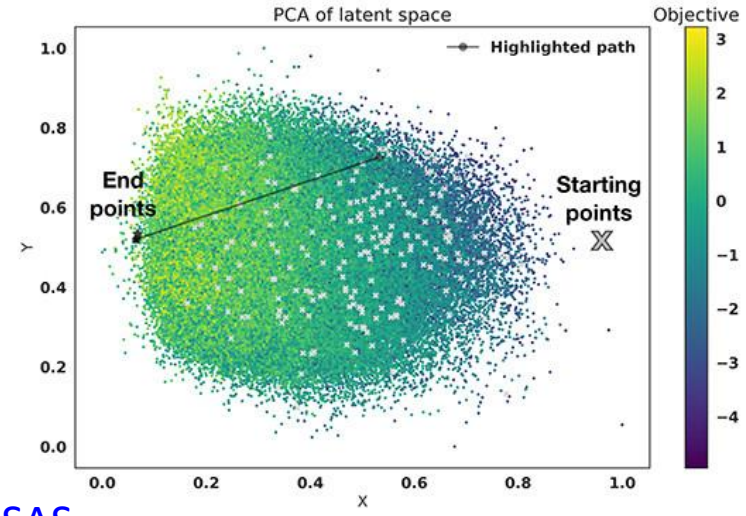
Gómez-Bombarelli, R.; Wei, J. N.; Duvenaud, D.; Hernández-Lobato, J. M.; Sánchez-Lengeling, B.; Sheberla, D.; Aguilera-Iparraguirre, J.; Hirzel, T. D.; Adams, R. P.; Aspuru-Guzik, A., Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules. *ACS Central Science* **2018**, 4, 268-276.

Inverse QSAR: deep learning

(a)

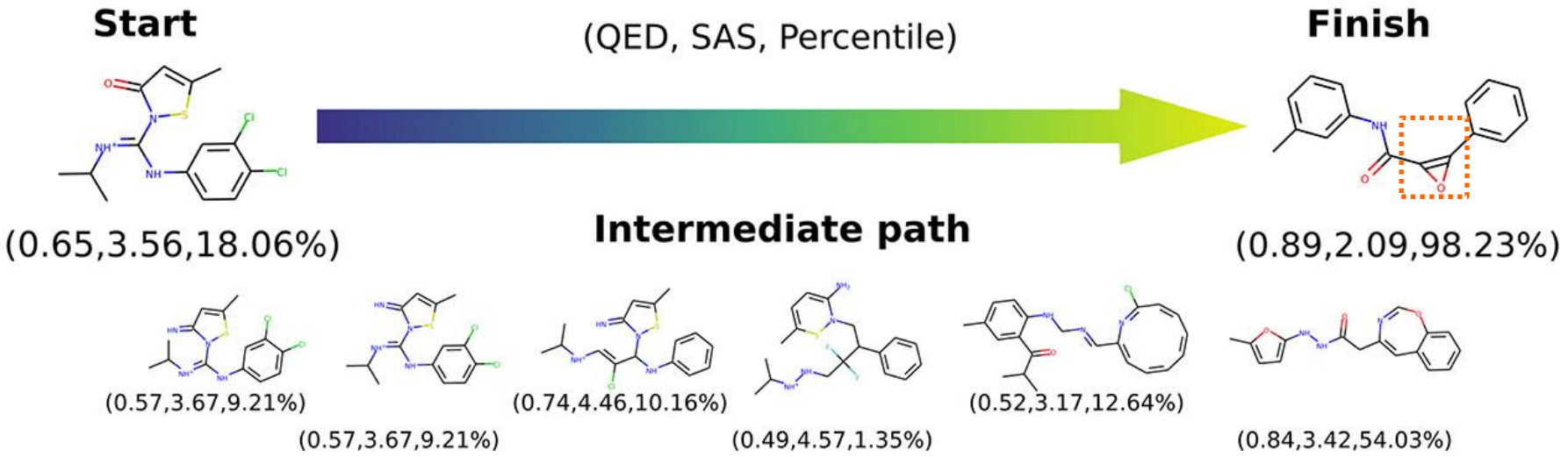


(b)



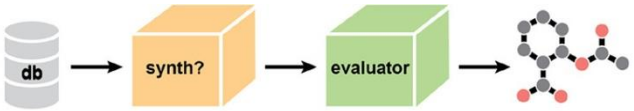
5xQED - SAS

(c)

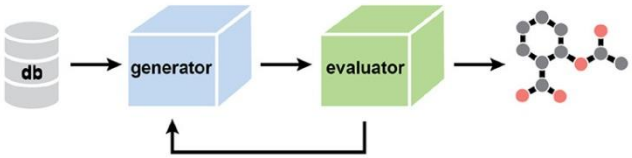


Gómez-Bombarelli, R.; Wei, J. N.; Duvenaud, D.; Hernández-Lobato, J. M.; Sánchez-Lengeling, B.; Sheberla, D.; Aguilera-Iparraguirre, J.; Hirzel, T. D.; Adams, R. P.; Aspuru-Guzik, A., Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules. *ACS Central Science* **2018**, 4, 268-276.

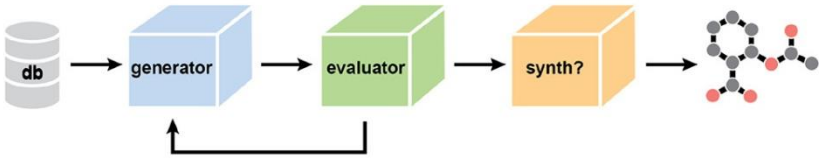
Control of synthetic accessibility



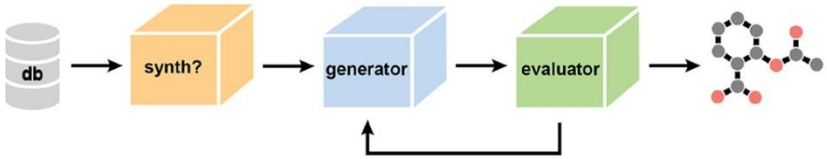
(a) Virtual screening



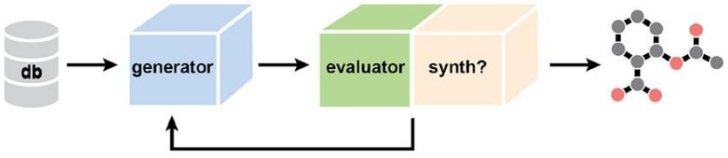
(b) Standard molecular generation



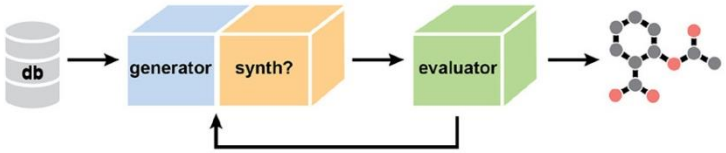
(c) Post hoc filter



(d) Bias by training set



(e) Bias by heuristics



(g) Explicit constraints

Gao, W.; Coley, C. W., The Synthesizability of Molecules Proposed by Generative Models. *Journal of Chemical Information and Modeling* **2020**

Assessment of synthetic accessibility

Journal of Cheminformatics
ChemistryCentral

Research article
Open Access

Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions

Peter Ertl* and Ansgar Schuffenhauer

Genheden *et al. J Cheminform* (2020) 12:70
https://doi.org/10.1186/s13321-020-00472-1

Journal of Chem

SOFTWARE
Open Access

Check for updates

AiZynthFinder: a fast, robust and flexible open-source software for retrosynthetic planning

Samuel Genheden[1*], Amol Thakkar[1,2], Veronika Chadimová[1], Jean-Louis Reymond[2], Ola Engkvist[1] and Esben Bjerrum[1*]

Chemical Science
ROYAL SOCIETY OF CHEMISTRY

EDGE ARTICLE
View Article Online
View Journal | View Issue

Check for updates

Cite this: *Chem. Sci.*, 2021, **12**, 3339

All publication charges for this article have been paid for by the Royal Society of Chemistry

Retrosynthetic accessibility score (RAscore) – rapid machine learned synthesizability classification from AI driven retrosynthetic planning†

Amol Thakkar, *[ab] Veronika Chadimová, [a] Esben Jannik Bjerrum, [a] Ola Engkvist [a] and Jean-Louis Reymond *[b]

Voršilák *et al. J Cheminform* (2020) 12:35
https://doi.org/10.1186/s13321-020-00439-2

Journal of Cheminformatics

RESEARCH ARTICLE
Open Access

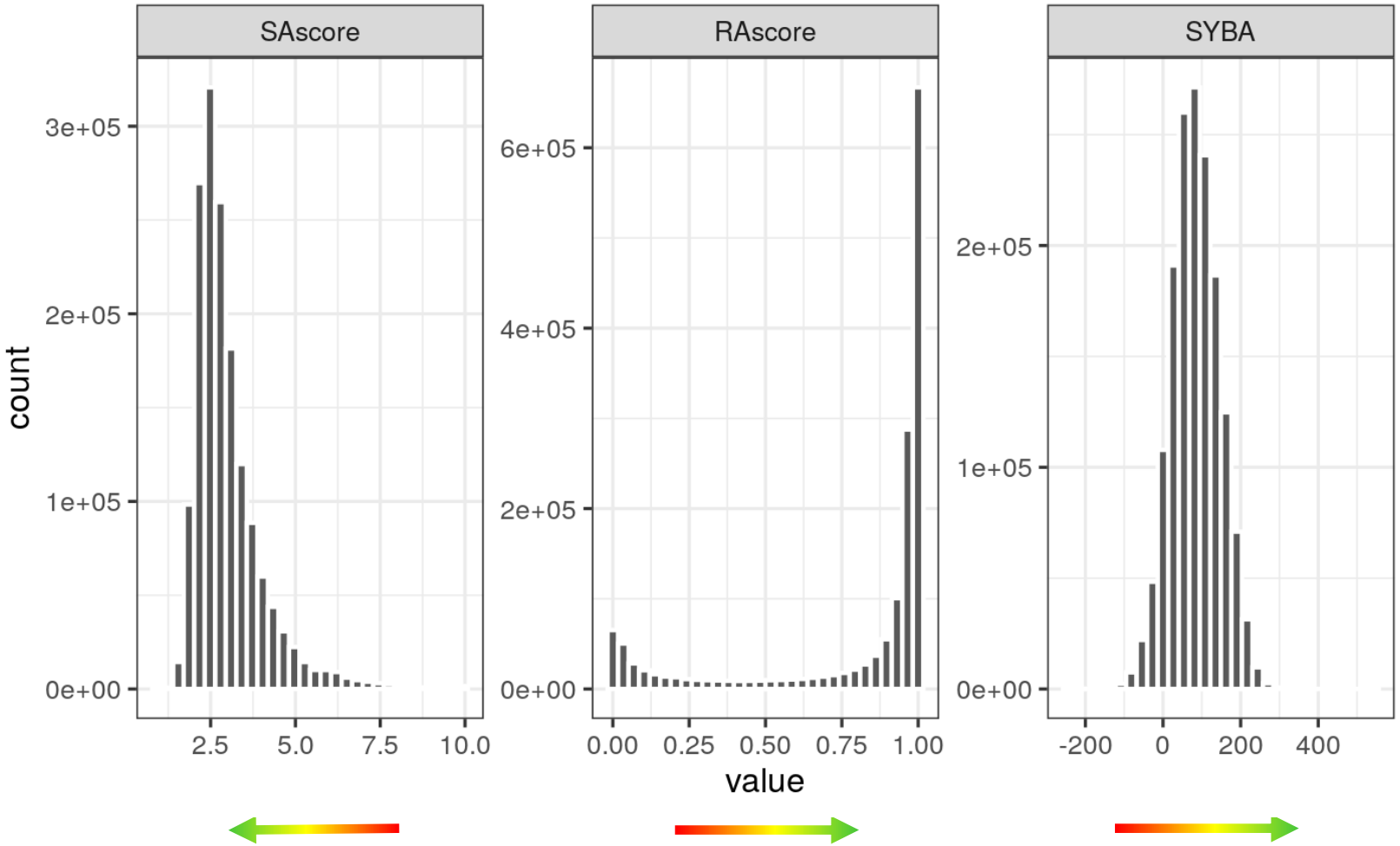
Check for updates

SYBA: Bayesian estimation of synthetic accessibility of organic compounds

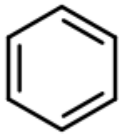
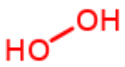
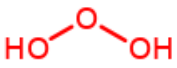
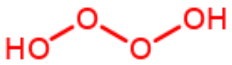
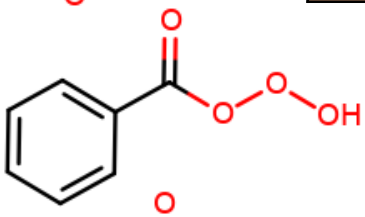
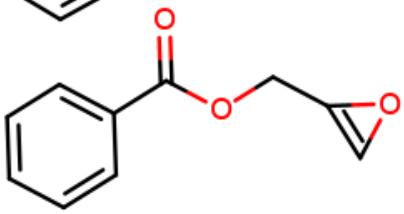
Milan Voršilák[1,2], Michal Kolář[3,4], Ivan Čmelo[1] and Daniel Svozil[1,2*]

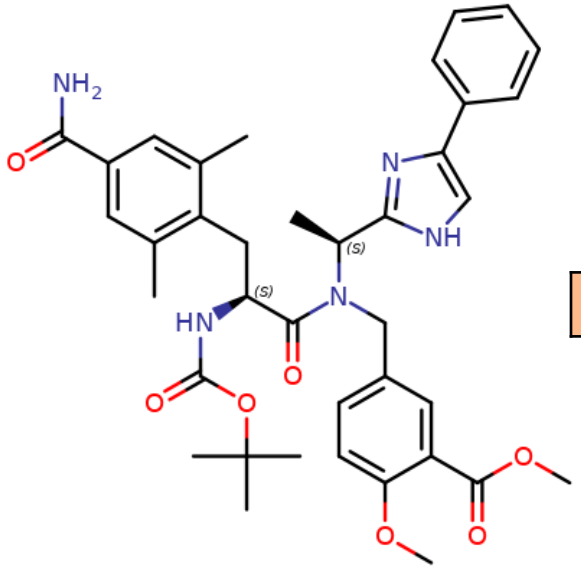
Assessment of synthetic accessibility

ChEMBL22 (1.58 M compounds)

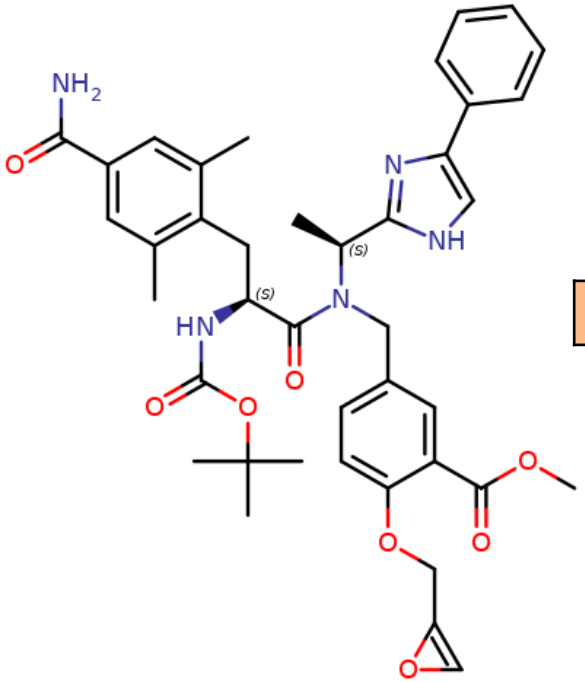


Assessment of synthetic accessibility

	SA	RA	SYBA
	1	0.98	11.6
	3.07	0.97	7.99
	3.22	0.97	1.63
	3.46	0.97	-2.94
	1.81	0.97	7.34
	3.44	0.98	-9.66
	2.03	0.99	15.0
	2.1	0.99	33.0



3.87	0.80	174



4.16	0.56	148

Control of synthetic accessibility

Content of fragmented library

all ChEMBL compounds (1 554 160)

compounds with SA score ≤ 2.5 (572 527)

compounds with SA score ≤ 2 (107 806)

Context radius

1 less conservative replacements

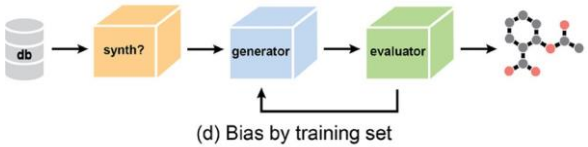
2

3

4

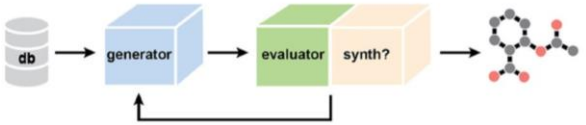
5 more conservative replacements

CReM

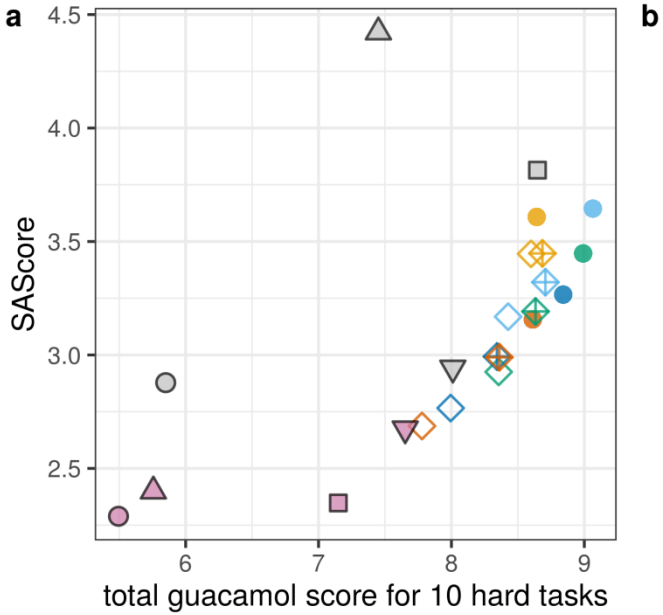


(d) Bias by training set

state-of-the-art approaches

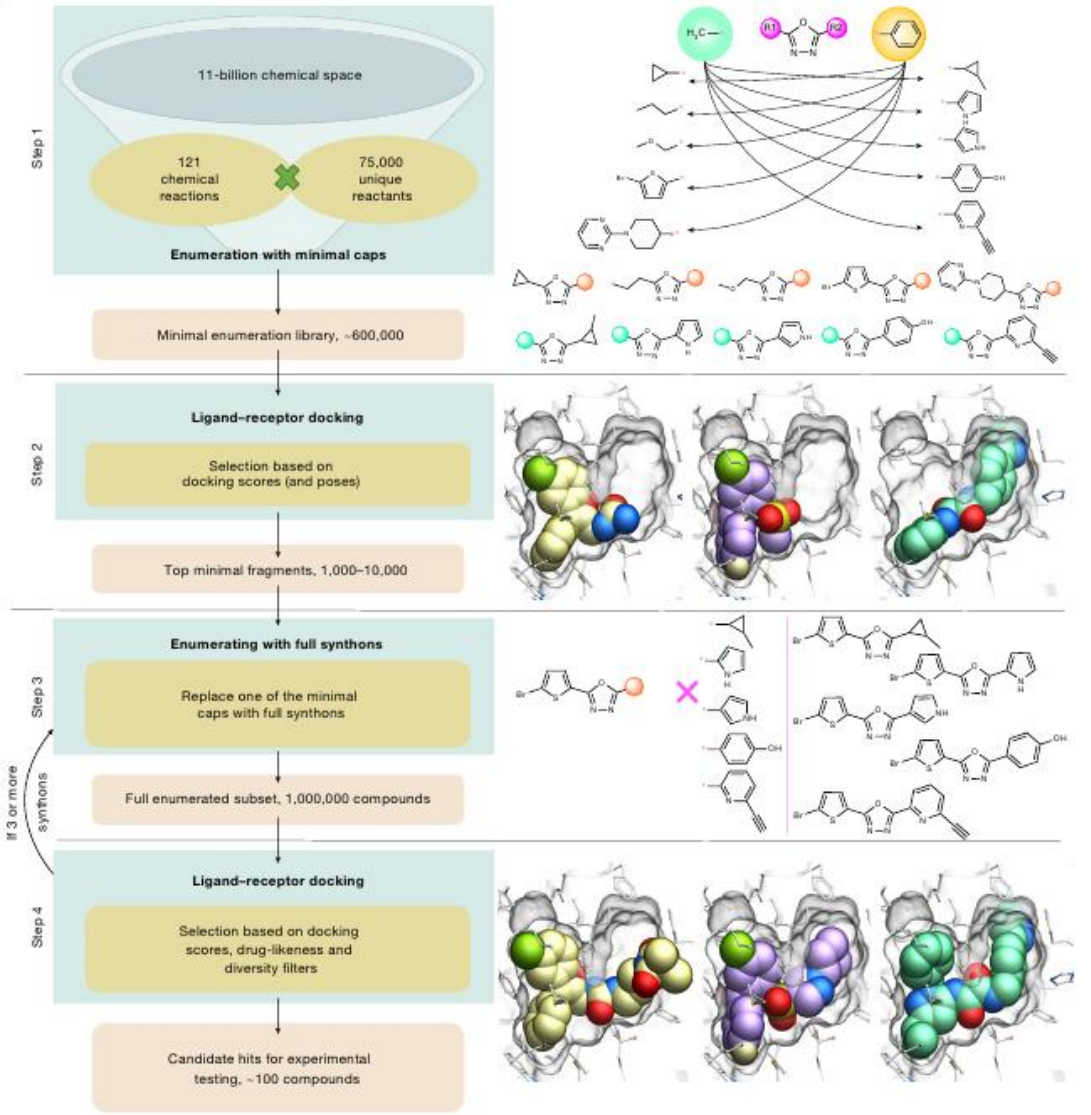


(e) Bias by heuristics



Polishchuk, P., Control of Synthetic Feasibility of Compounds Generated with CReM. *Journal of Chemical Information and Modeling* **2020**, 60, 6074-6080. (10.1021/acs.jcim.0c00792)

V-SYNTHES



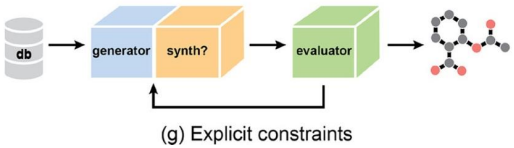
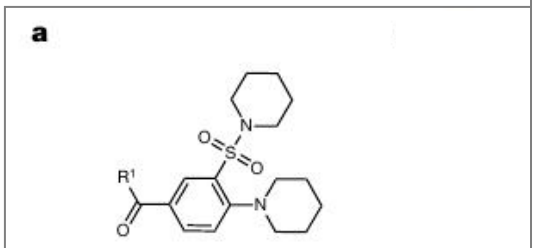


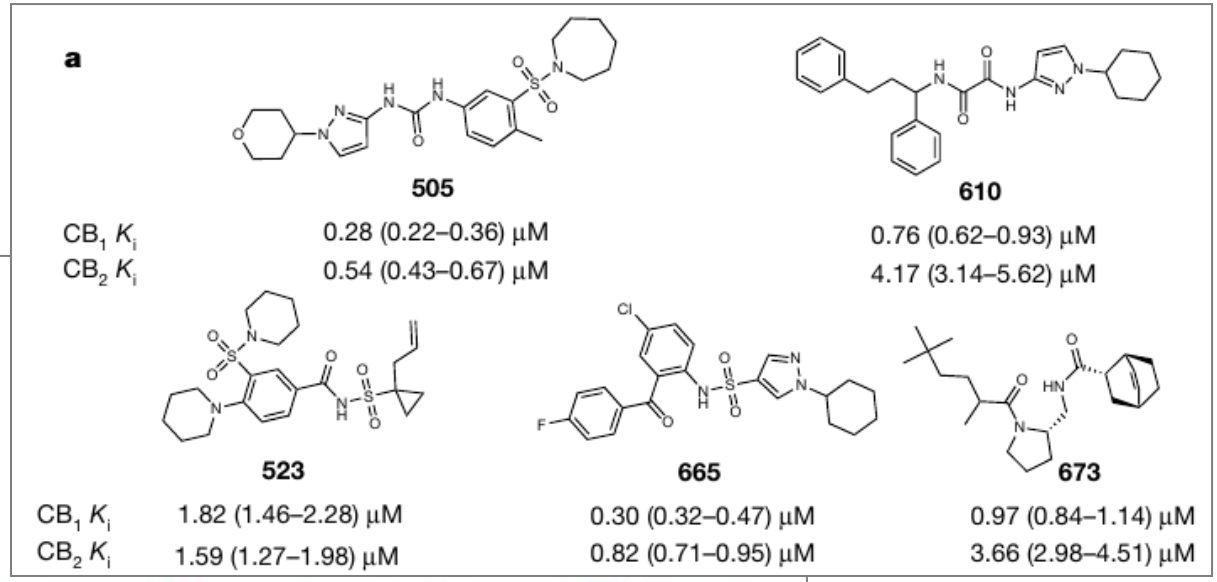
Fig. 1 | V-SYNTHES approach to modular screening of Enamine REAL Space. A general overview of the four-step algorithm (left) and examples for each step (right). Asterisks in step one show the attachment points of synthons; arrows show possible pairing of minimal synthons with real synthons.

Sadybekov, A. A. et al, Synthon-based ligand discovery in virtual libraries of over 11 billion compounds. *Nature* **2021**. (10.1038/s41586-021-04220-9)

V-SYNTHES



523 scaffold

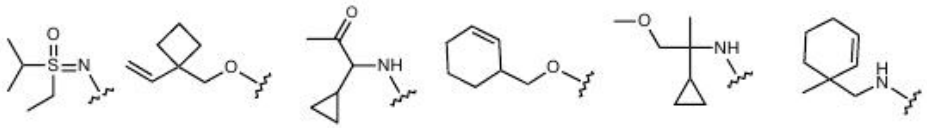


	505	**610**
CB_1 K_i	0.28 (0.22–0.36) μM	0.76 (0.62–0.93) μM
CB_2 K_i	0.54 (0.43–0.67) μM	4.17 (3.14–5.62) μM

	523	**665**	**673**
CB_1 K_i	1.82 (1.46–2.28) μM	0.30 (0.32–0.47) μM	0.97 (0.84–1.14) μM
CB_2 K_i	1.59 (1.27–1.98) μM	0.82 (0.71–0.95) μM	3.66 (2.98–4.51) μM

c

R1



	Compound	**733**	**736**	**738**	**742**	**747**	**749**
CB_1 functional potency	K_i (nM)	871	1,185	856	2,340	455	209
	CI 95% (nM)	(720–1,051)	(868–1,603)	(725–1,009)	(1,878–2,919)	(373–558)	(177–248)
CB_2 functional potency	K_i (nM)	*10.9*	48.5	125	120	*9.6*	49.2
	CI 95% (nM)	9.3–12.9	38.6–61.0	105–148	101–144	8.58–10.8	42.1–57.6
CB_1 binding affinity	K_i (nM)	43.2	140	23.1	394	228	689
	CI 95% (nM)	28.2–66.1	105–186	13.9–38.6	281–551	172–303	472–1,004
CB_2 binding affinity	K_i (nM)	*1.2*	2.8	13.0	6.4	*0.9*	4.0
	CI 95% (nM)	0.9–1.6	2.0–3.7	10.2–16.6	5.2–7.8	0.6–1.2	2.5–6.5

Sadybekov, A. A. et al, Synthon-based ligand discovery in virtual libraries of over 11 billion compounds. *Nature* **2021**. (10.1038/s41586-021-04220-9)

IMTM

Take home message

- De novo design can efficiently explore much larger chemical space than virtual screening
- There are multiple approaches to generate chemically valid structures, all of them have their pros and cons
- The main issue of de novo design is synthetic feasibility of generated compounds
- There are several ways how to control synthetic feasibility