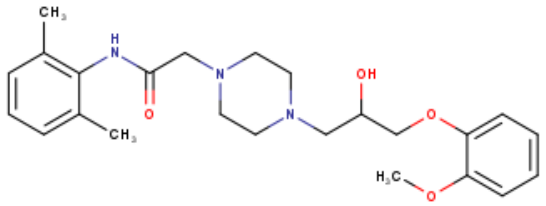
Integrating synthetic accessibility with AI-based generative drug design

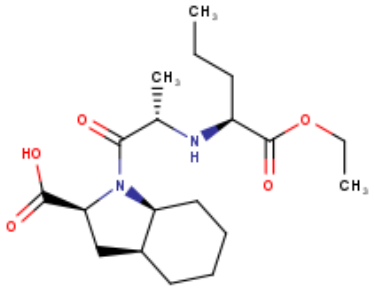
UKQSAR Spring 2021 Meeting – 27th April 2021 Virtual Meeting

IKTOS

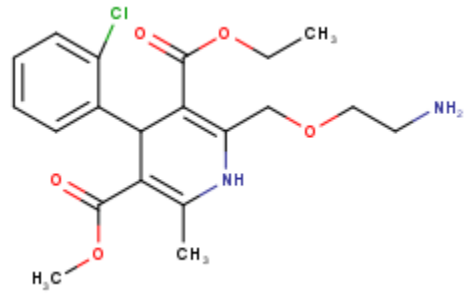
Issues with synthetic accessibility of generated compounds



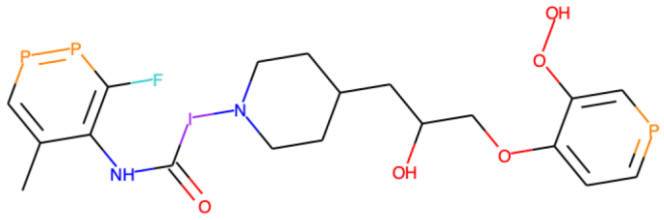
Ranolazine



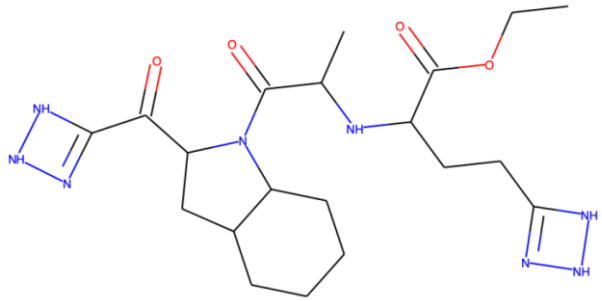
Perindopril



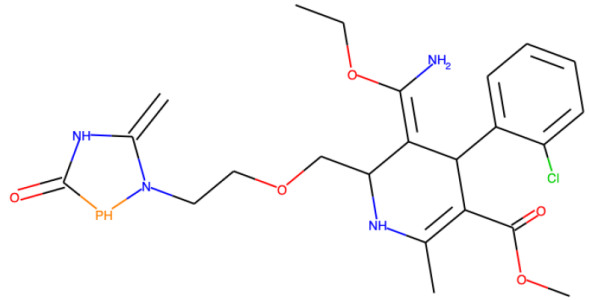
Amlodipine



Generated Ranolazine



Generated Perindopril



Generated Amlodipine

Example of generative model output **without** synthetic assessment heuristics[1]

1 - The Synthesizability of Molecules Proposed by Generative Models; Wenhao Gao and Connor W. Coley; *Journal of Chemical Information and Modeling* Article ASAP April 2020

Known scores and heuristics

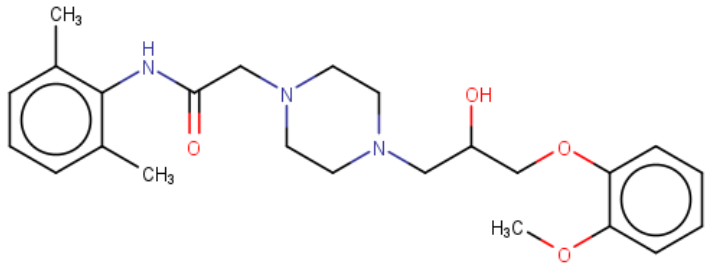
- ✓ Smiles: a simple heuristic which considers the length of the smiles
- ✓ SA score[1]: A score which use the complexity and frequency of known fragments (from 1 to 10 the lower the better)
- ✓ SCScore[2]: A model which encode the increasing complexity of reactions' sequences (from 1 to 5 the lower the better)
- ✓ RAscore[3]: A classification model which predict the feasibility of a molecule according to AiZynthFinder (from 0 to 1 the higher the better)

Known heuristics are fast but fail to distinguish the synthesizability of compounds perfectly.

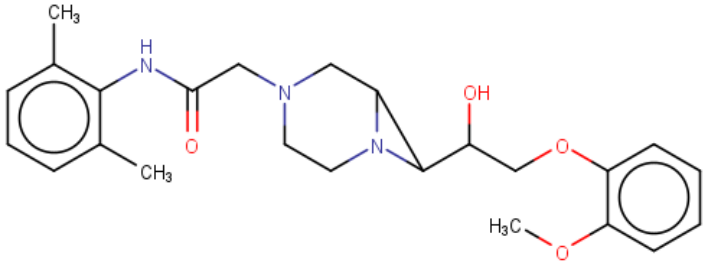
1 - Ertl, P.; Schuffenhauer, A., Estimation of Synthetic Accessibility Score of Drug-Like Molecules Based on Molecular Complexity and Fragment Contributions. Journal of Cheminformatics 2009, 1 (1), 8
2 - Coley, C. W.; Rogers, L.; Green, W. H.; Jensen, K. F., Scscore: Synthetic Complexity Learned from a Reaction Corpus. Journal of Chemical Information and Modeling 2018, 58 (2), 252-261
3 – Thakkar A.; Chadimova V.; Bjerrum E. J.; Engkvist O.; Reymond J.L., Retrosynthetic Accessibility Score (RAscore) - Rapid Machine Learned Synthesizability Classification from AI Driven Retrosynthetic Planning, Chem. Sci., 2021,12, 3339-3349

Challenge to assess synthetic feasibility

Synthetic accessibility prediction is a highly nonlinear task, with a lot of “Feasibility cliff”



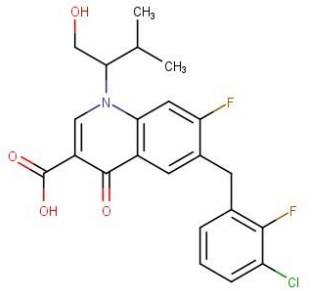
Easy to make



Very hard to make

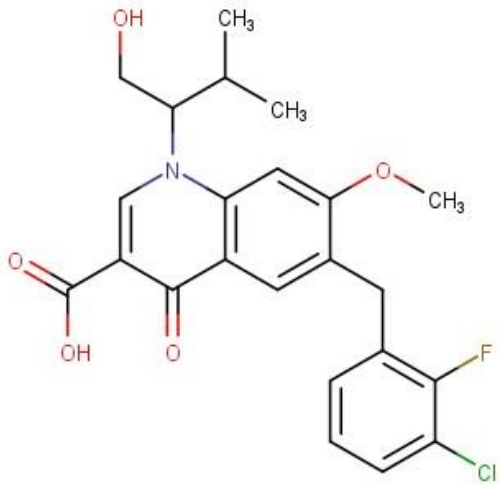
Definition of synthetic accessibility

Easy from



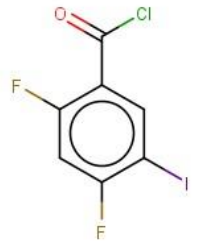
Commercially available SM

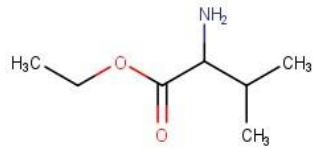
1 reaction step

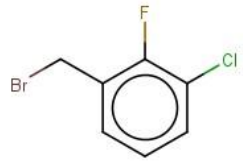


Elvitegravir

Complex from







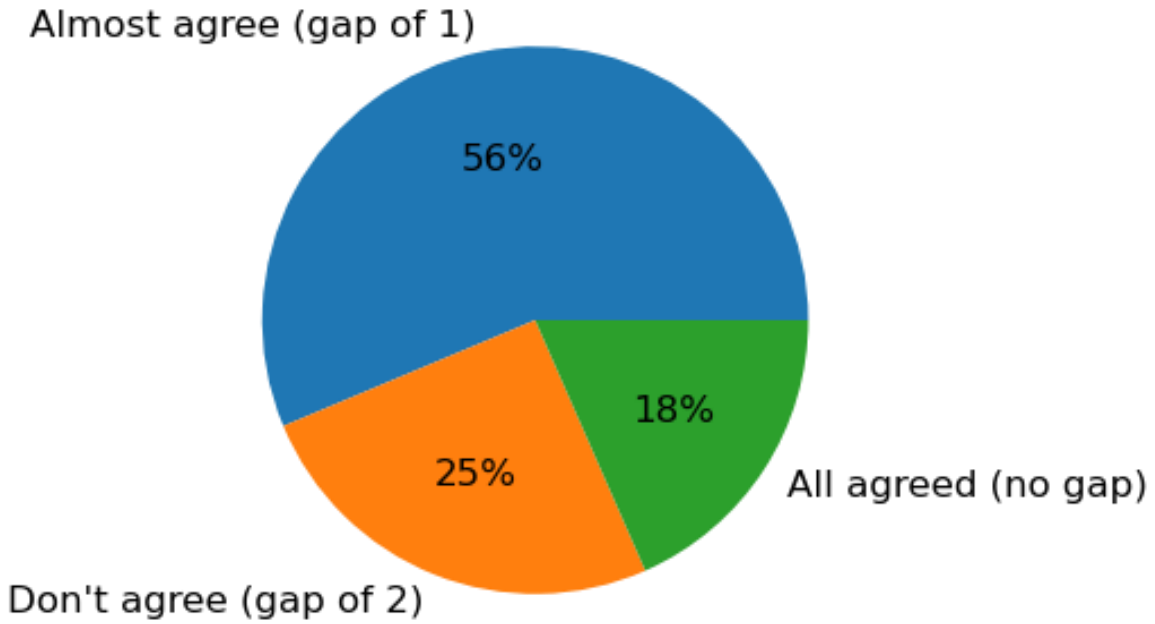
Commercially available SM

10 reactions steps

Depending on the database of SM we go from synthetic accessibility to synthetic complexity. We need more than a heuristic

Chemist bench

We asked 4 chemists to score 100 molecules from a generative AI output (from 1 hard to 4 easy)



Synthetic feasibility is also difficult to define even from expert chemists

Spaya: Iktos' AI driven retrosynthesis

Databases

Customizable

Reactions database

Customizable

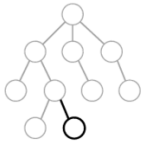
Starting materials database

Data driven approach

Algorithm

Disconnection prediction

Search strategy



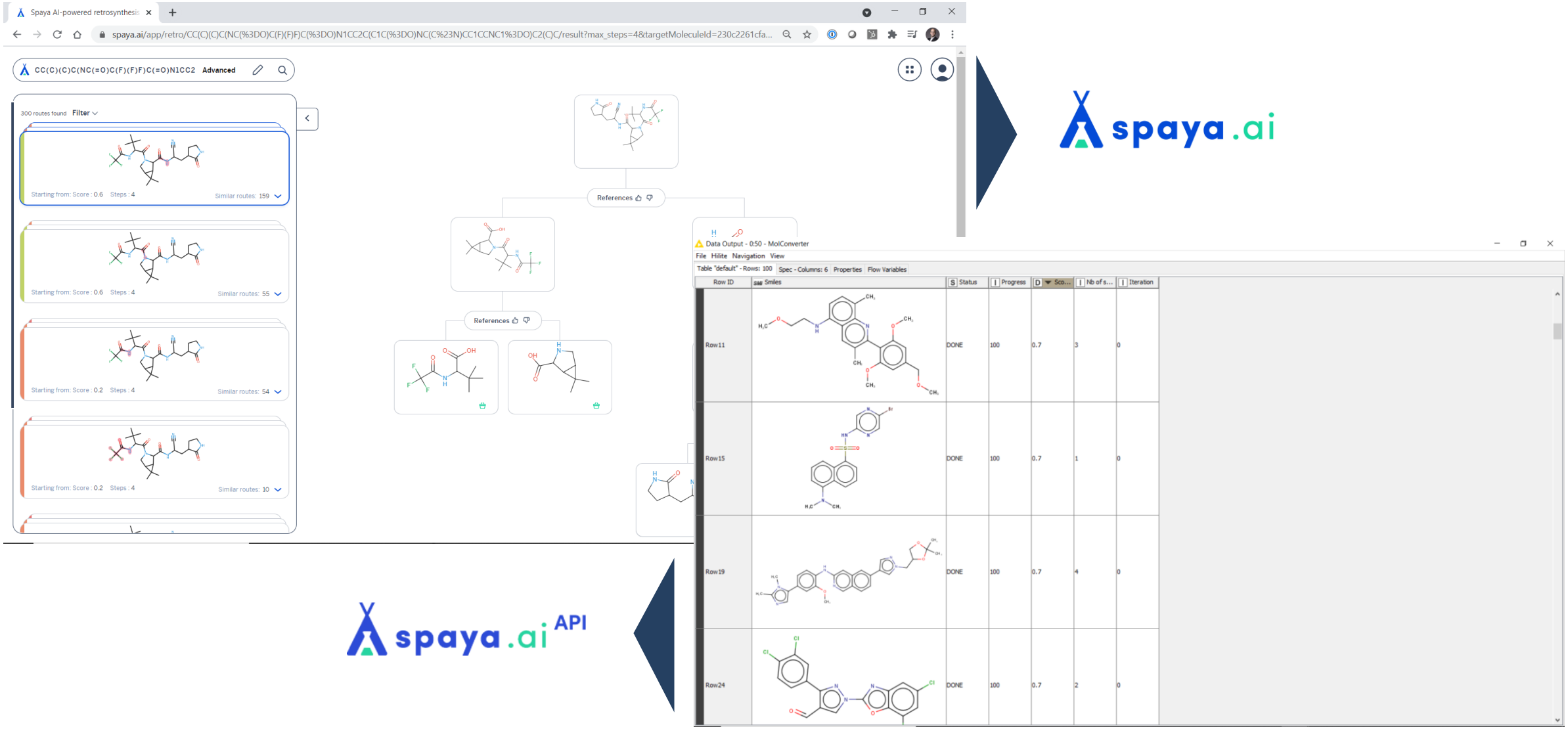
Reaction feasibility

Interface

UX/UI

For a detailed description of the technology see Spaya on Youtube: https://www.youtube.com/watch?v=M7vH9xMdLgw&t=780s
Or test our software freely available at spaya.ai

Spaya: Iktos' AI driven retrosynthesis

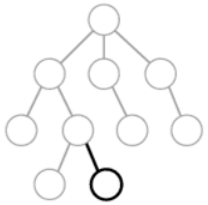


Retro Score or RScore or RS from Spaya

Molecule

⟹

Spaya API

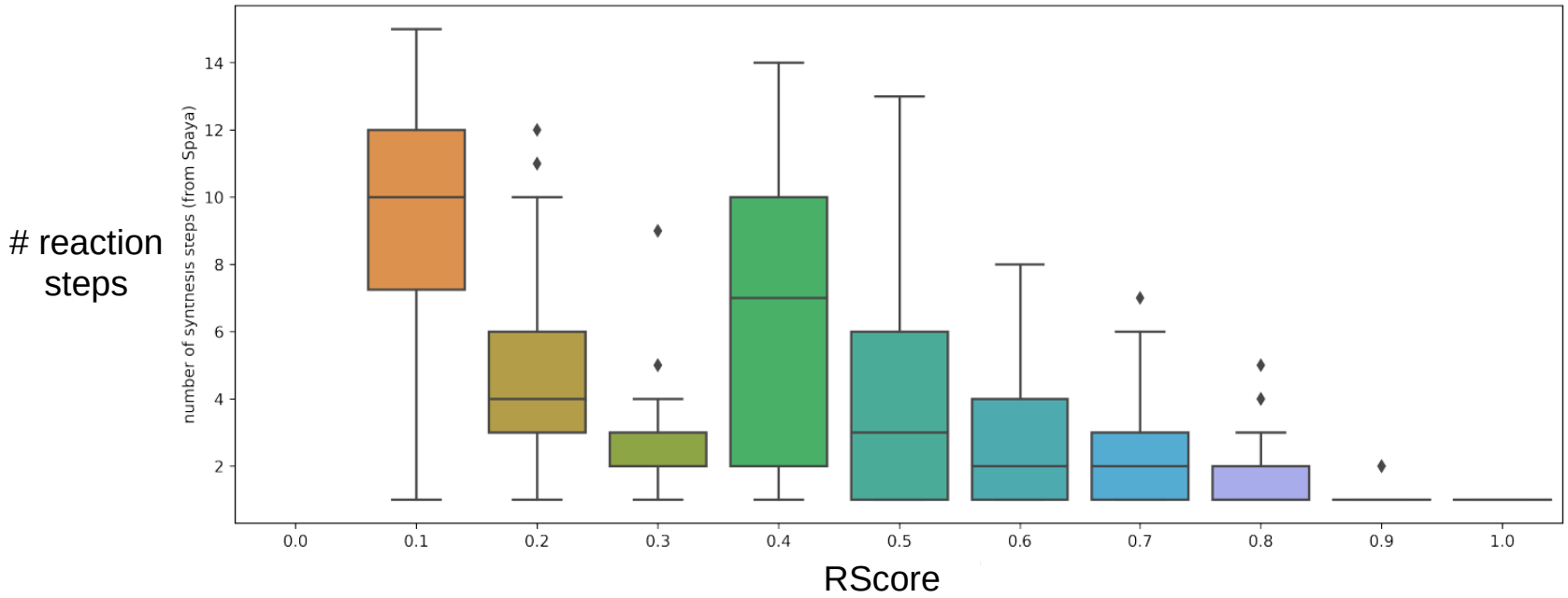


⟹

RScore	#steps
0.6	3

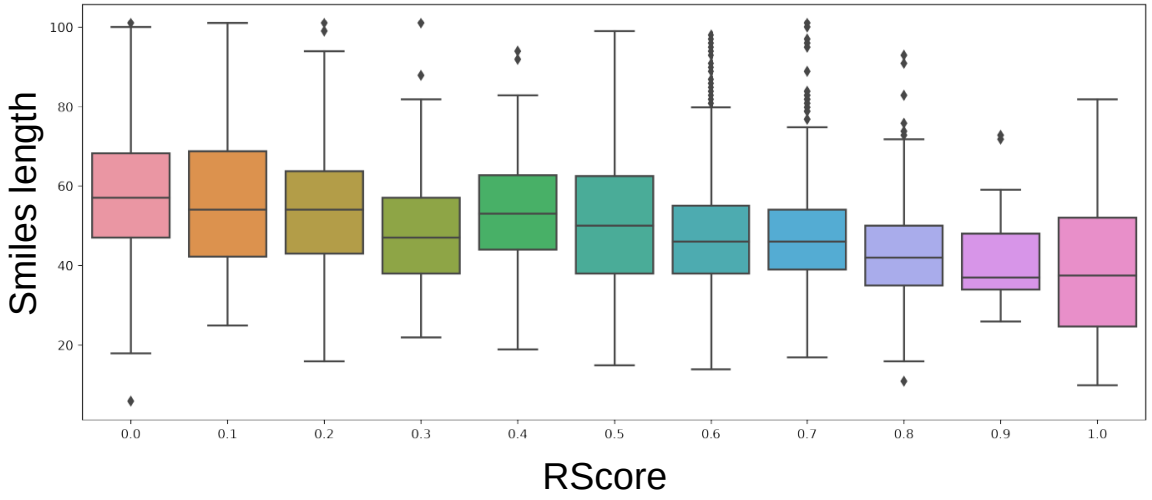
RScore (the higher the better from 0 to 1) is defined for a given route and is dependent on the **probability of the models,** its **applicability domain**, the **number of steps** in the route and the **convergence** of the route.
(except for Score = 1 which means exact literature match)

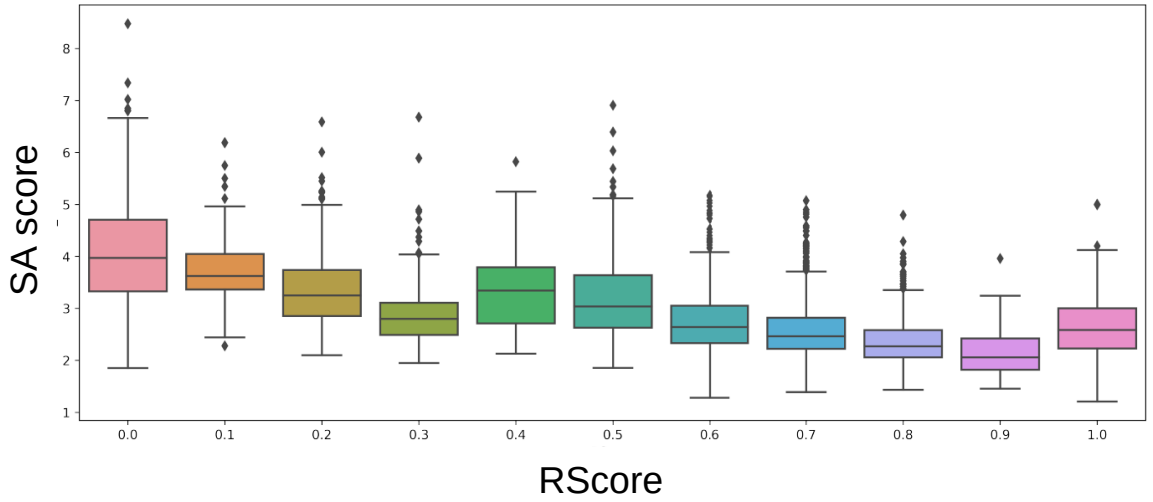
aws From seconds to one minute per molecule per CPU

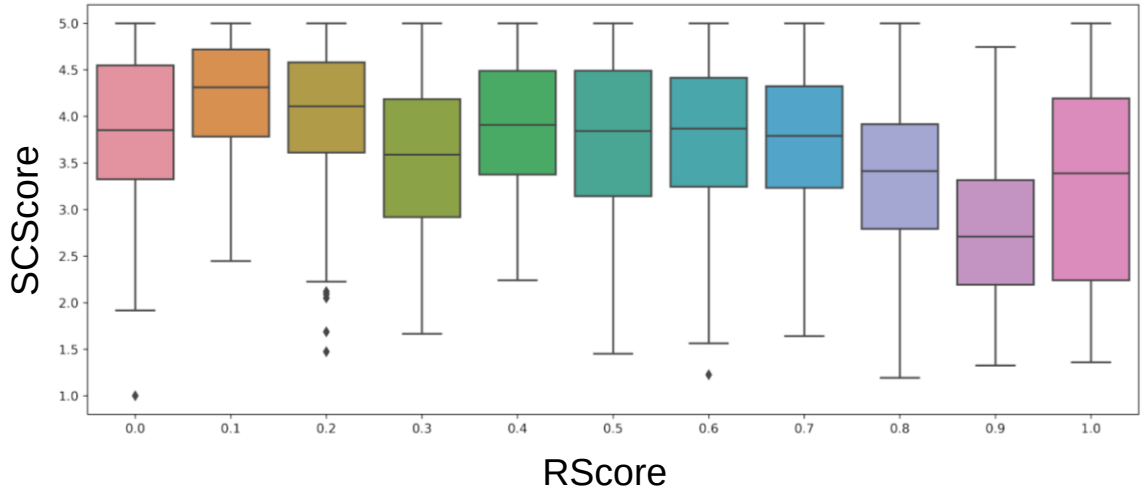


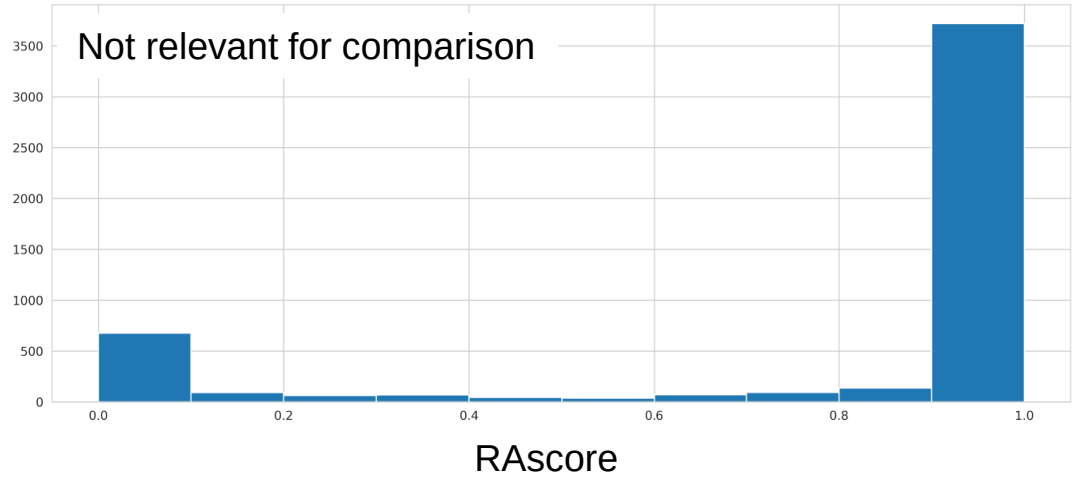
Correlation between the RScore and the number of reaction steps on a sample of molecules from chembl

RScore vs known metrics





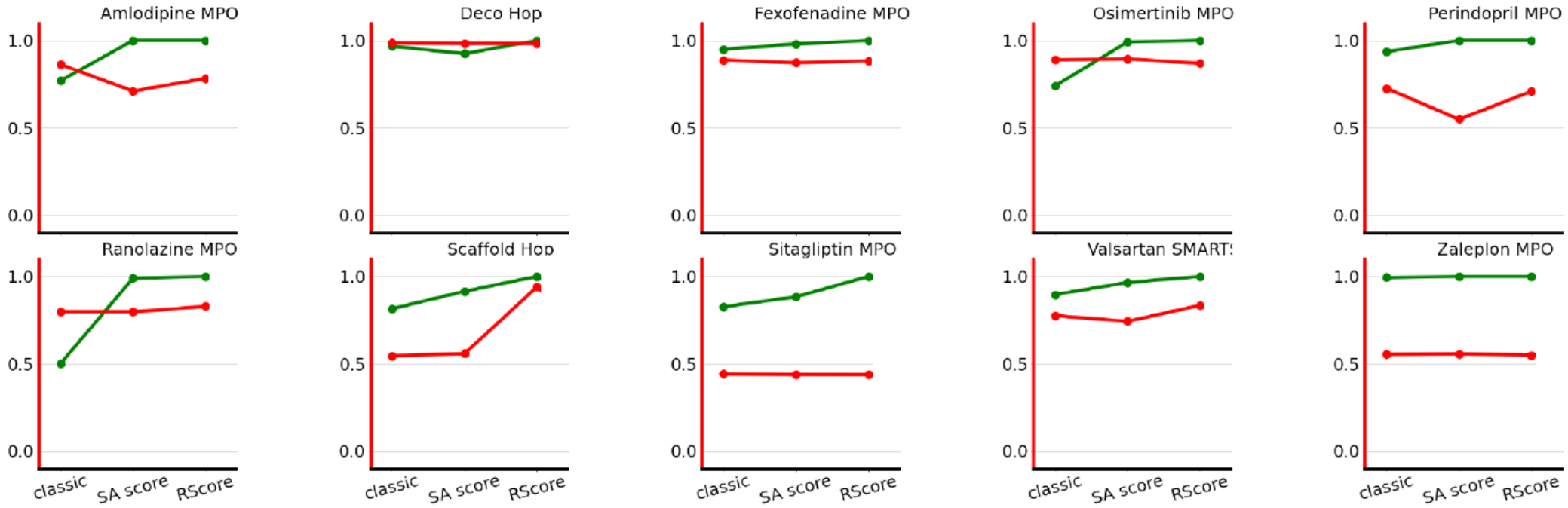




If we consider RScore as a ground truth we can see that SA score is the best known metric to assess synthetic feasibility compared to SCScore or Smiles length

RScore vs SA score during the generation

SA score is the metric that correlate the best with RScore, so we decided to compare on different tasks from guacamole benckmark



% of molecule with RScore > 0.5

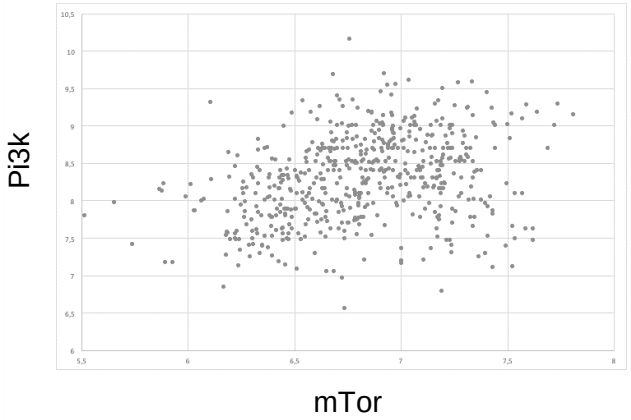
Average reward on the top 100 molecules

We do not observe a major difference on the output between SA score and RScore on those simple tasks. In our experience generative AI use to produce ugly molecules when solutions are difficult to find.

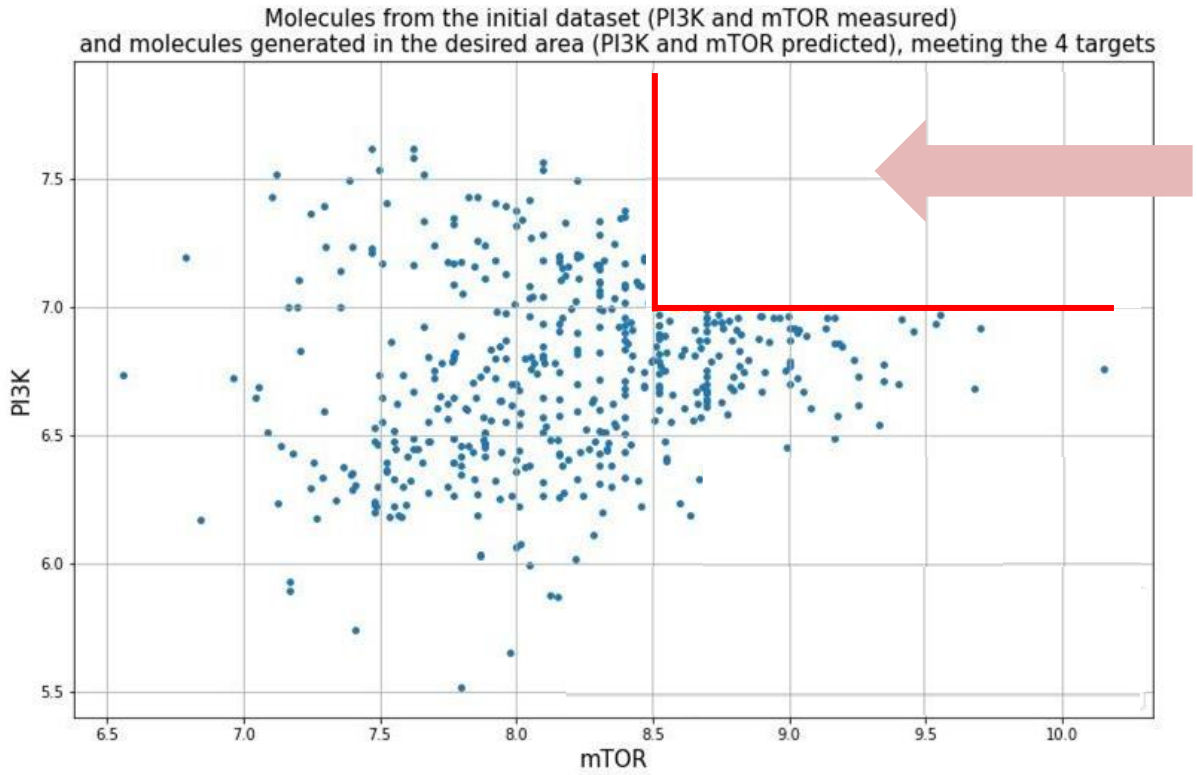
RScore vs known scores during the generation on more complex tasks

Objectives to optimize

> 2 QSAR models
> 2 metrics



Generation against 2 kinase activity (QSAR)
Including similarity and QED

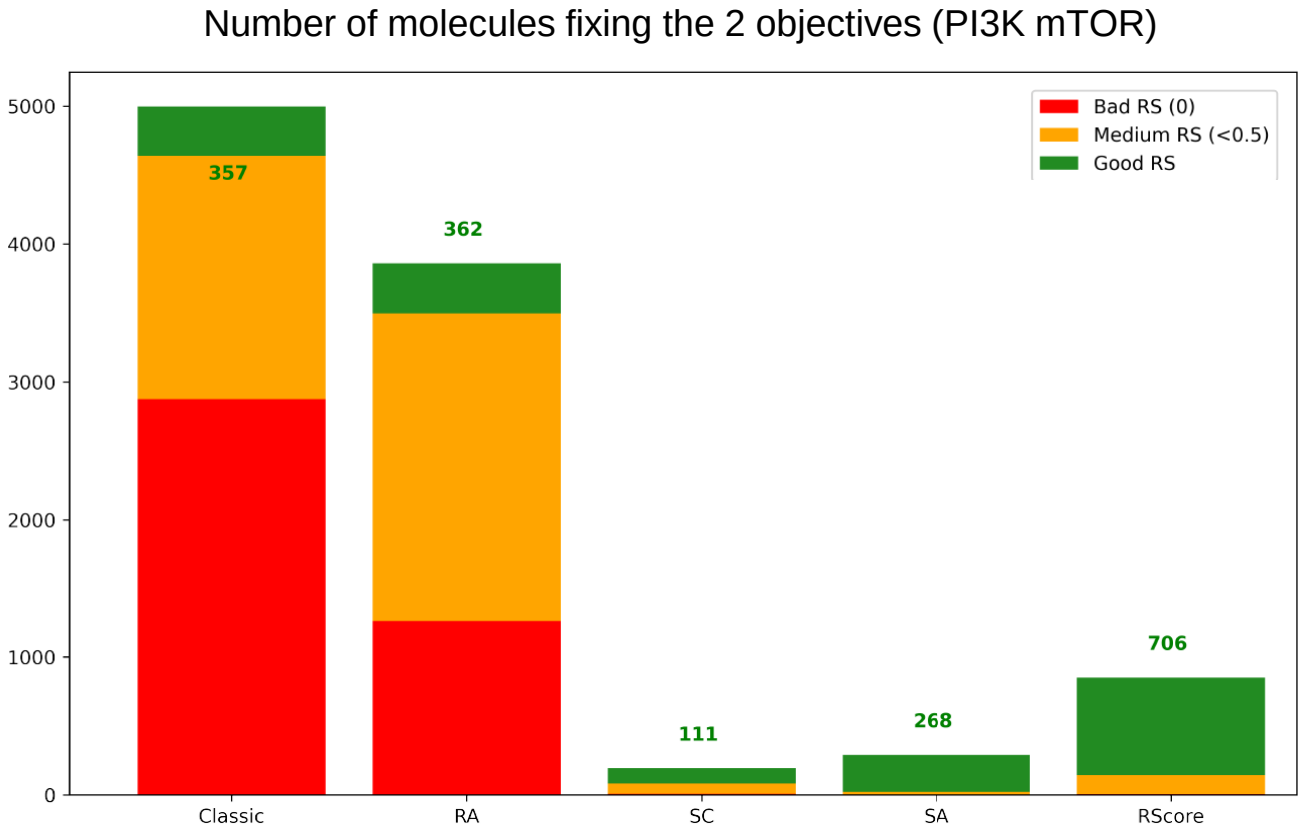


Goal:
Generate new molecules in that chemical space

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.

RScore vs known scores during the generation on more complex tasks

Impact of the different scores during the generation, RScore is used in post processing as an oracle.



Classic generation and RA score produce lot of molecule but most of them are difficult or unfeasible contrary to other scores.

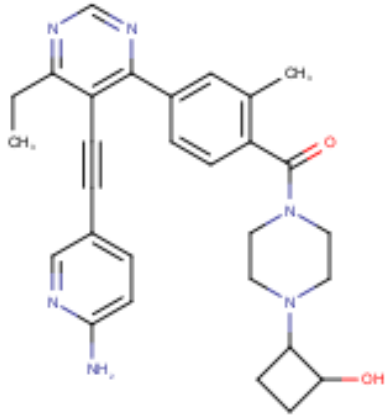
RScore vs known scores during the generation on more complex tasks

Impact of the different scores during the generation, RScore is used in post processing as an oracle.

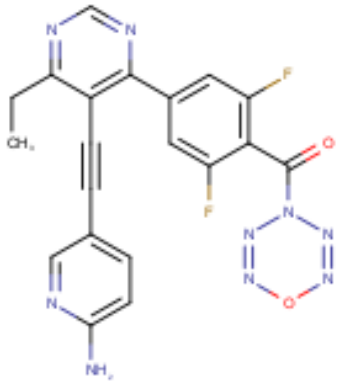
	blueprint	in GT	Coverage	Simi	Standard murcko	Generic murcko	average RScore	feasible	Good RScore
classic	5005	2	56	0.67	1230	97	0.08	1959	282
RA	3574	2	54	0.69	775	112	0.11	2660	360
SC	211	1	23	0.68	33	19	0.35	202	127
SA	311	2	35	0.76	48	36	0.56	311	286
RS	850	2	51	0.71	102	60	0.49	843	706

RScore produces more molecules in the objectives, with a better diversity and with a very high proportion of easy to make molecules compared to other scores
It is interesting to observe that a simple metric like SA score produces nice results in this case study

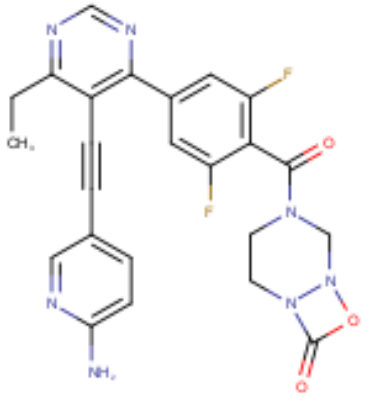
Some example of generative model output



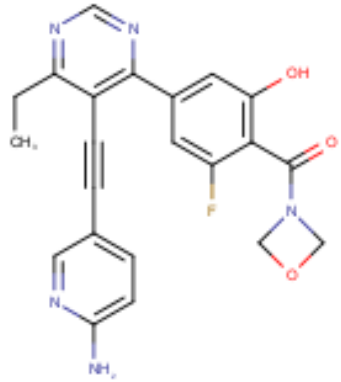
Spaya RScore = **0.72**
SA score = **3.7**
SC score = **5**
RAscore = **0.99**



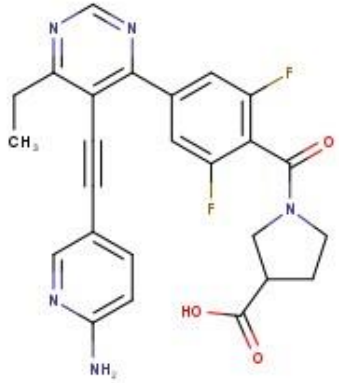
Spaya RScore = **0**
SA score = **3.8**
SC score = **4.3**
RAscore = **0.02**



Spaya RScore = **0**
SA score = **3.6**
SC score = **4.8**
RAscore = **0.25**



Spaya RScore = **0**
SA score = **3.2**
SC score = **4.7**
RAscore = **0.99**

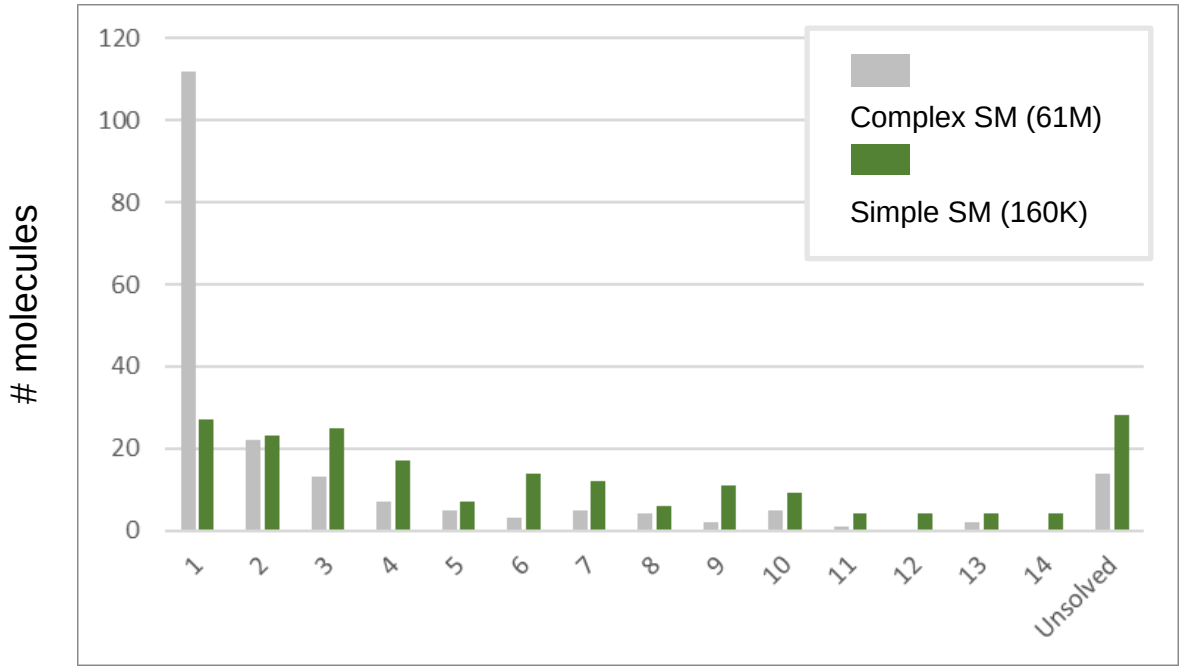


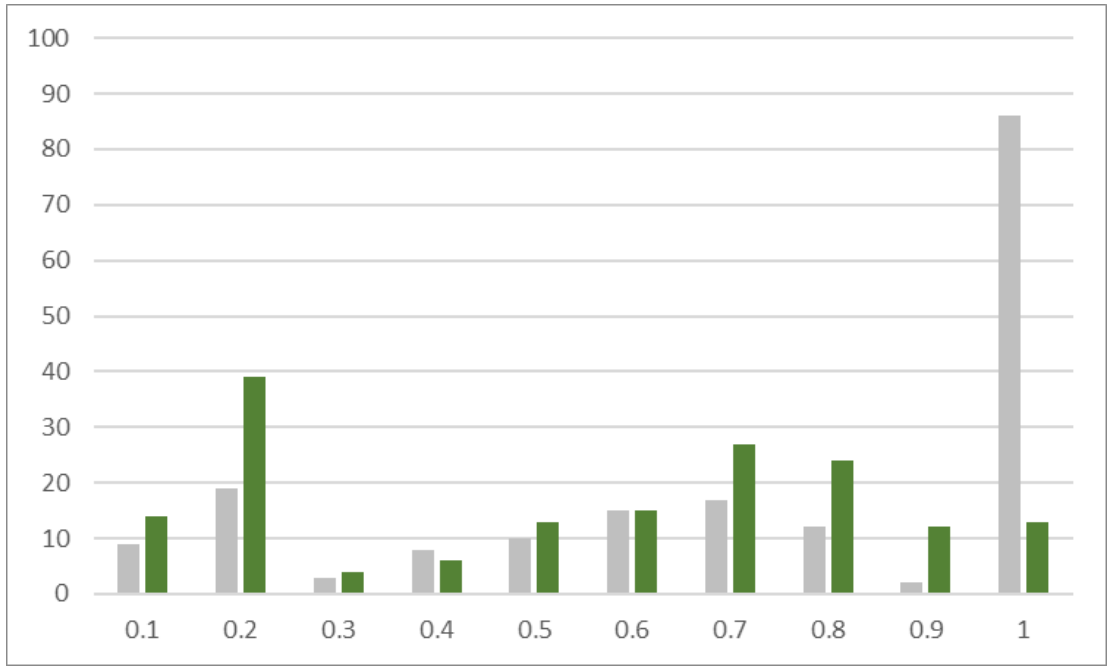
Spaya RScore = **0.6**
SA score = **3.4**
SC score = **4.8**
RAscore = **0.15**

Contrary to other scores, Spaya RScore is able to distinguish between complex molecules which are still easy to make and molecules containing irrelevant motives: obviously not synthesizable and sometime unstable.

Advantage of the RScore

Contrary to heuristics or predictive model, RScore can be easily customizable





Here we can see the impact of the database of commercially available starting material on a bench of known molecules. RScore can be tune for synthetic feasibility or for synthetic complexity assessment. Chemistry can also be changed, focusing on classic reactions: Amide coupling, Cross coupling…

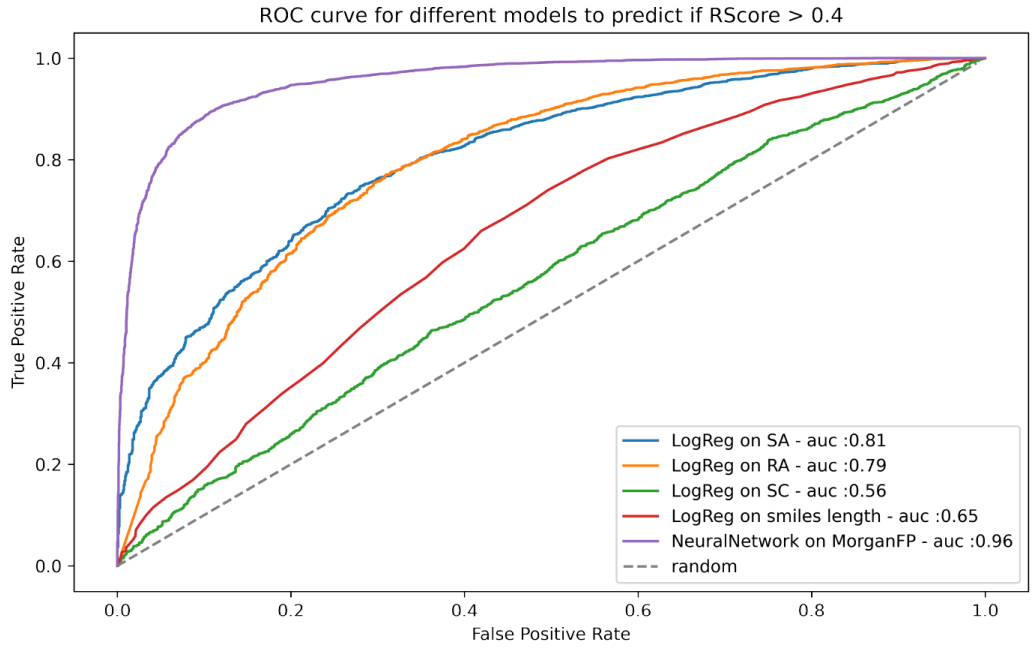
RScore main wickness

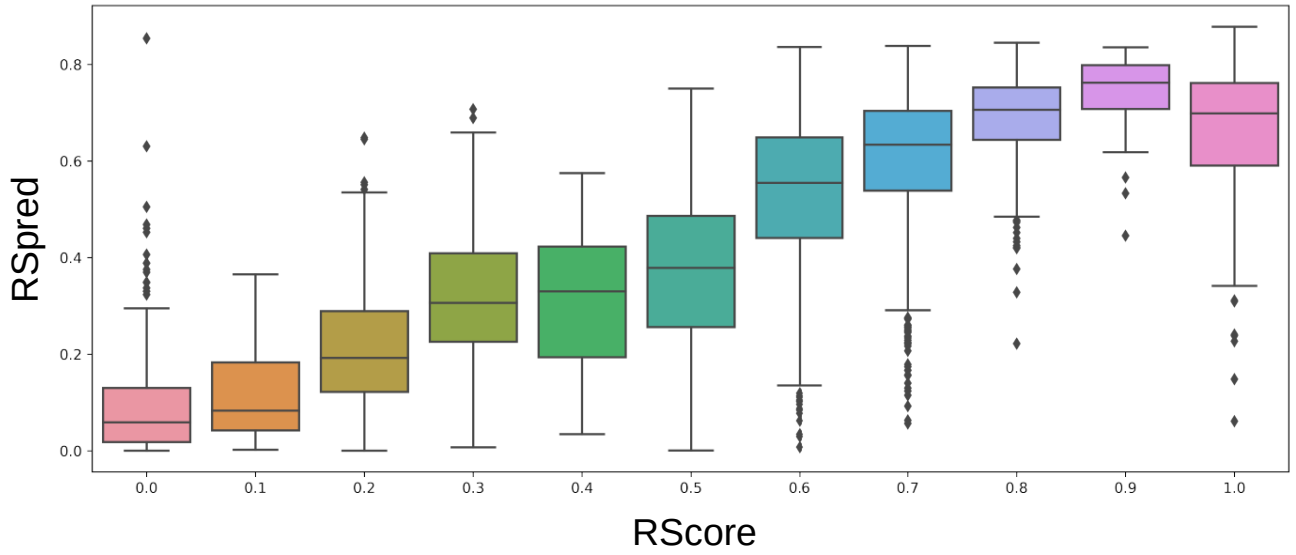
RScore is highly accurate and produce very nice results,
but it take time to compute (second to minute)

From RScore to RSpred

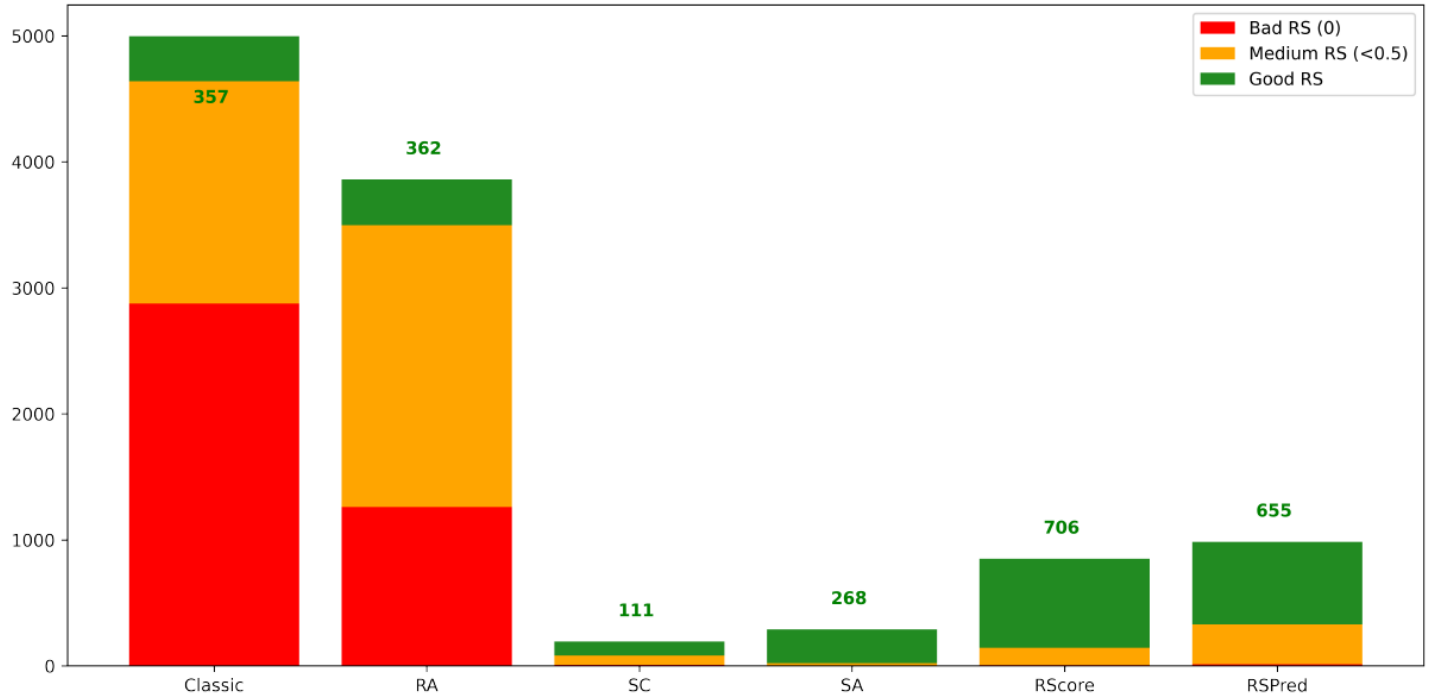
	Correlation
LinReg on smiles lenght	0.32
LinReg on RA score	0.49
LinReg on SC score	0.22
LinReg on SA score	0.60
Neural Network on MorganFp	0.75

Neural Network trained on a 230K molecules retrosynthesized by Spaya API





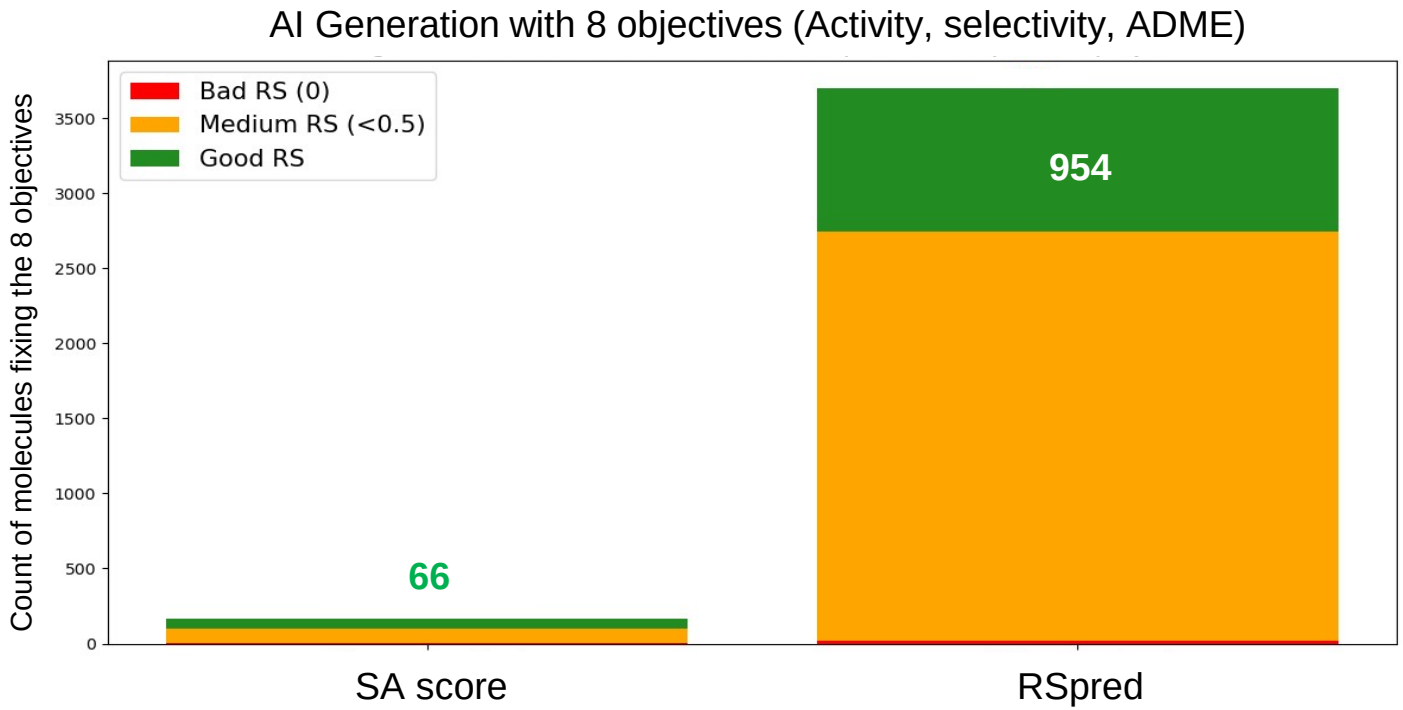
From RScore to RSpred



RScore and RSpred produce comparable results:
From seconds to 10 milliseconds!

	blueprint	in GT	Coverage	Simi	Standard murcko	Generic murcko	average RScore	feasible	Good RScore
classic	5005	2	56	0.67	1230	97	0.08	1959	282
RA	3574	2	54	0.69	775	112	0.11	2660	360
SC	211	1	23	0.68	33	19	0.35	202	127
SA	311	2	35	0.76	48	36	0.56	311	286
RS	850	2	51	0.71	102	60	0.49	843	706
RSPred	985	2	54	0.73	167	90	0.46	971	655

RScore on a complex client project



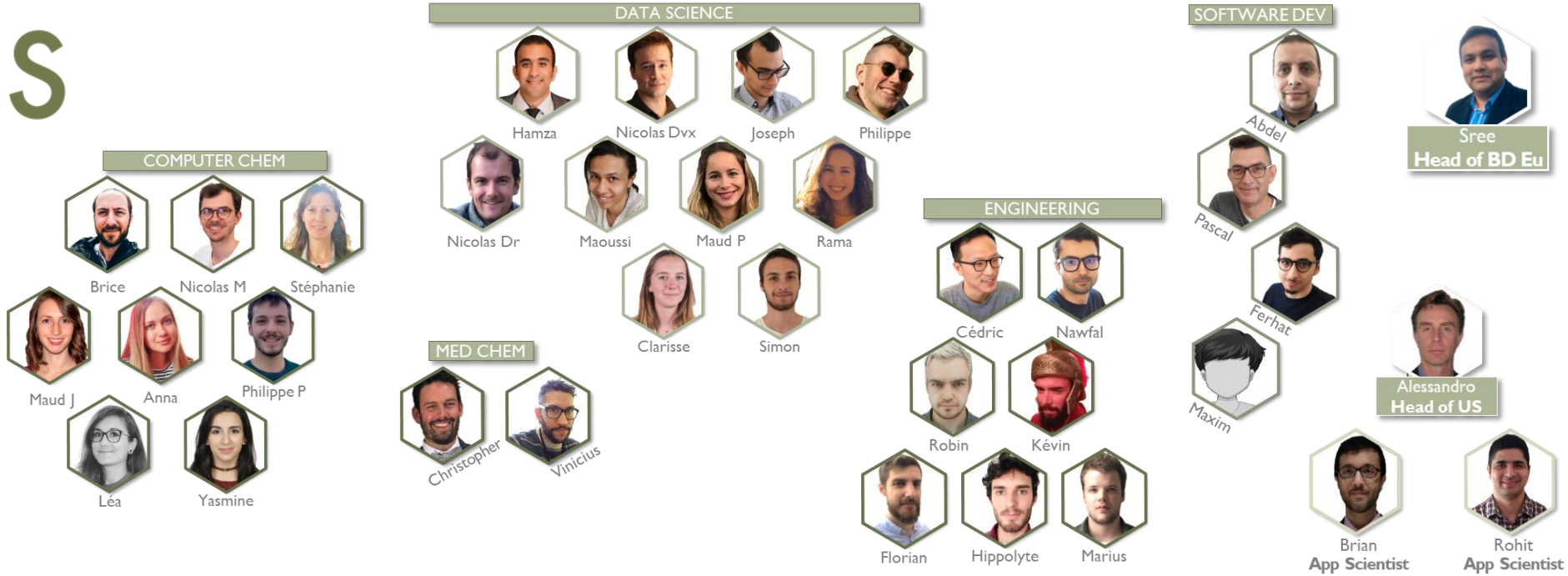
RSpred appears to outperform SA score by a very large margin on a complex MPO project

Take home message

- ✓ Using generative AI without synthetic accessibly heuristic leads to ugly molecules.
- ✓ Among the known heuristics SA score appear to be the best one to generate nice molecules
- ✓ RScore appears to be better than SA score producing more and divers easy to make molecules. This is maybe due to the fact that it runs a full retrosynthesis and make sure that the input molecule can be done from available starting material
- ✓ Contrary to other scores, RScore can be easily customizable either by modifying the scope of the chemistry but also the catalog of starting materials
- ✓ Despite the quality of the output of RScore, this score is time consuming and not really adapted for generative AI
- ✓ RSpred appears to reach similar performance compared to RScore but much faster and at much lower cost which makes a very attractive tool for generative AI
- ✓ RScore is accurate and very precise so we see lot of value in that score for post processing and prioritization of designs, where RSpred is very efficient for generative AI.

Thank you for your attention

The Iktos team!



Yann Gaston-Mathé
CEO
yann.gaston.mathe@iktos.com
+33 6 30 07 99 26

Nicolas Do Huu
CTO
nicolas.dohuu@iktos.com
+33 6 85 89 65 18

Quentin Perron
CSO
quentin.perron@iktos.com
+33 7 68 80 50 76