ETH zürich

De novo molecular design with machine intelligence

Gisbert Schneider

DCHAB
Department of Chemistry and Applied Biosciences

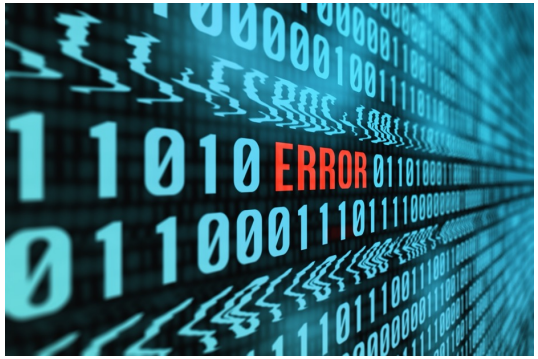
The dilemma of limited knowledge

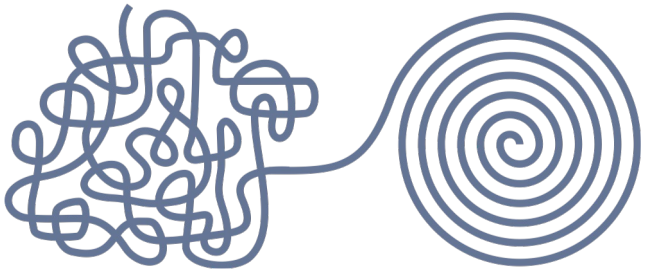
Image: Wikipedia, Musée Carnavalet

> **“Doctors pour drugs
> of which they know little,
> to cure diseases
> of which they know less,
> into human beings
> of whom they know nothing.**
>
> **Voltaire (1694−1778)**

ETH zürich

Drug discovery: At the edge of chaos





Images: Shutterstock

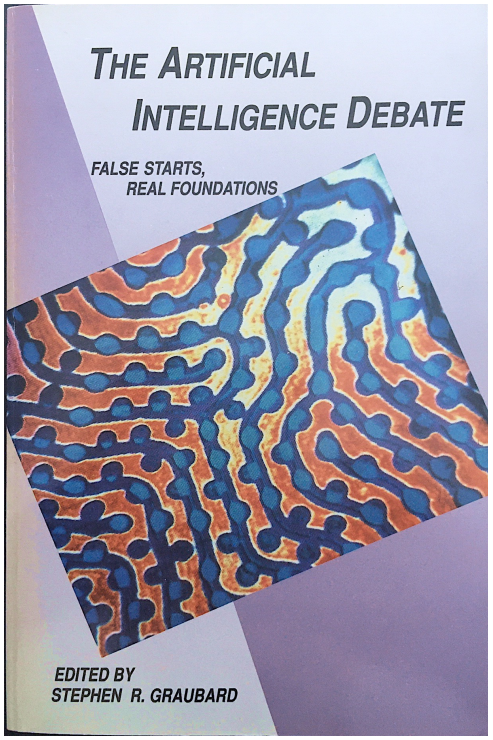
Error
- Models
- Data

Nonlinearity
- Drug action
- Non-additivity

Incompleteness
- Knowledge
- Partial predictability

Nat. Mach. Intell. **2019**, *1*, 128.
J. Med. Chem. **2016**, *59*, 4077.

Artificial intelligence to the rescue?



"Artificial intelligence has nothing new to offer [..] beyond the spectacle of ancient, well-drubbed errors replayed in a glitzy new medium.

Daniel D. Dennett (1988), p. 283

Play it again ...

> *We're now at a point where we have AI [systems] that are not directly programmed. They develop their own decision patterns.*

David Gunkel (2019), The National Post, Toronto

ETH zürich

Closing the loop: Feedback as a corrective

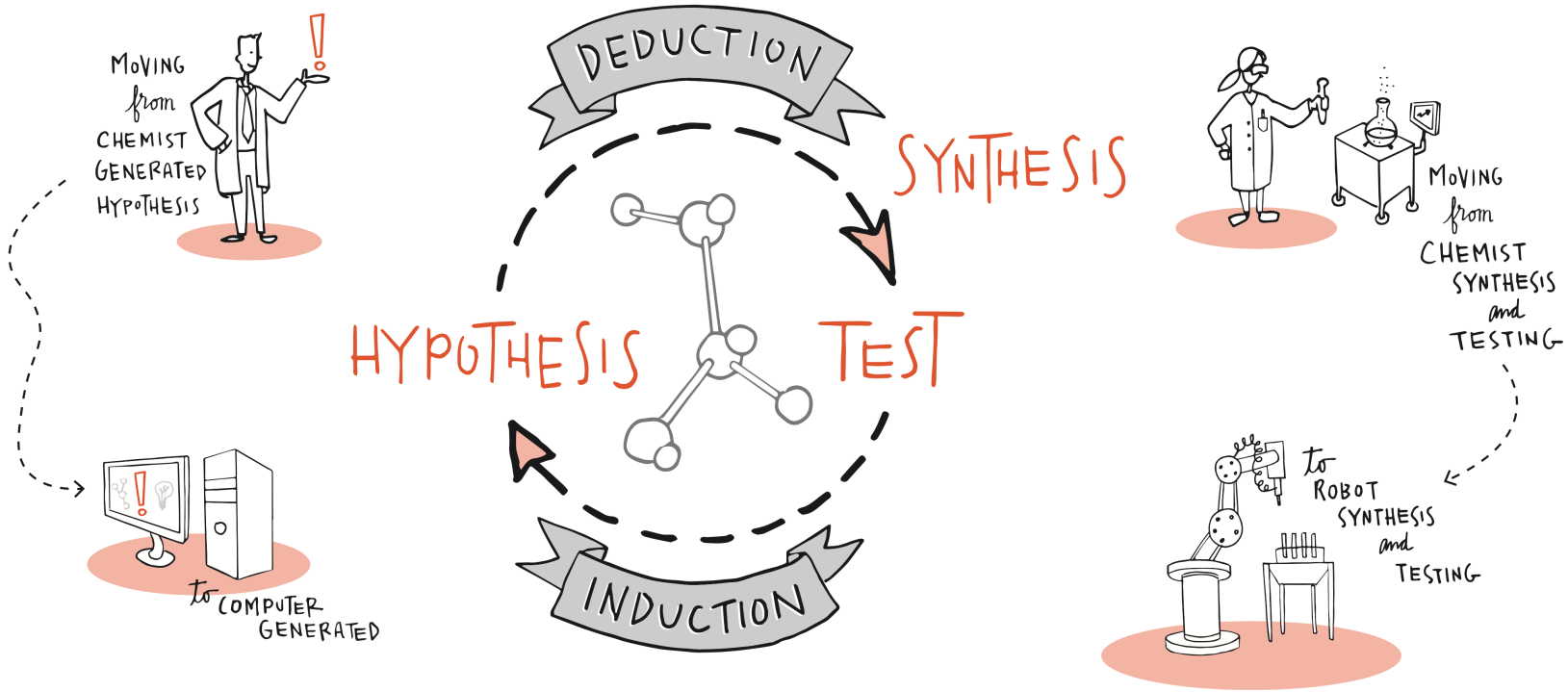


Image credit: Jack Burgess

Nat. Mach. Intell. **2019**, *1*, 128.
Nat. Rev. Drug Discov. **2018**, *17*, 97.

Hypothesis generation by ”de novo” molecular design

Structure Generator

- Chemist
- Rule-based (reactions, chemical transformations)
- Distribution-based (generative)

Scoring & Selection

- Experimental assay
- Explicit scoring (e.g., QSAR model)
- Implicit scoring (model-intrinsic)

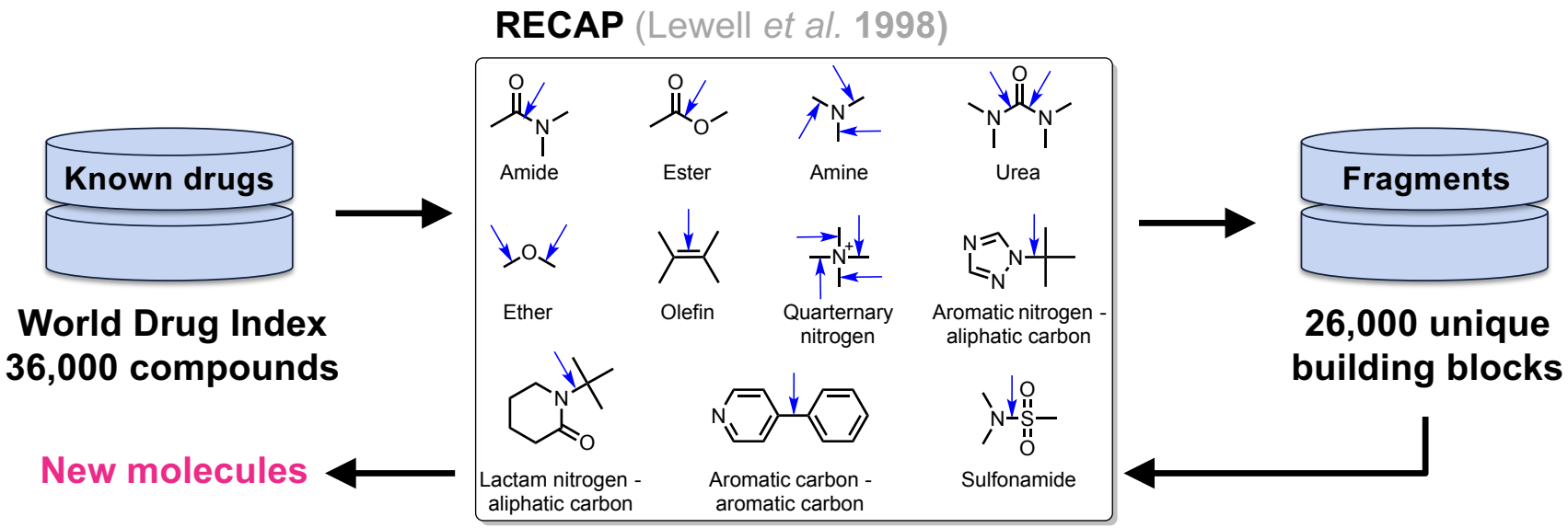
Optimizer

- Experimental design
- Stochastic, deterministic
- Active learning

Nat. Rev. Drug Discov. **2005**, *4*, 649.

Automated molecular design with pseudo-chemical transformations

TOPAS: Topology Assigning System



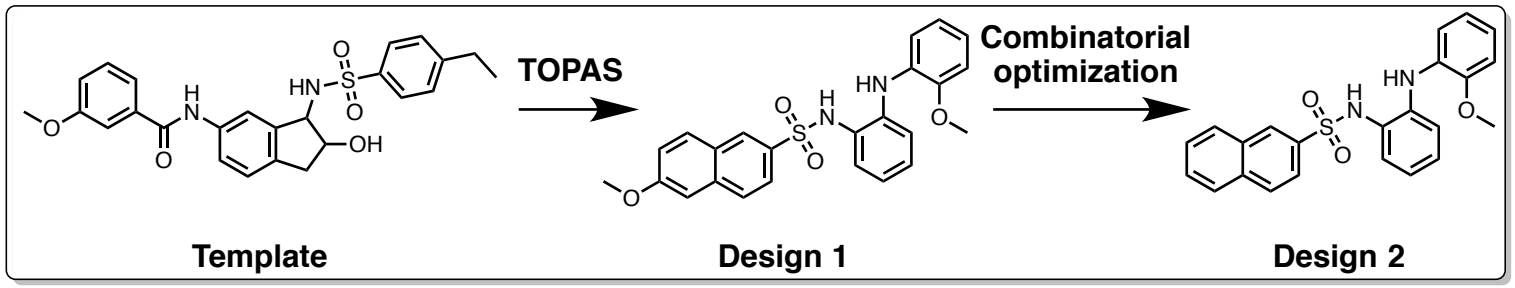
J. Comput. Aided Mol. Des. **2000**, *14*, 487.
Angew. Chem. Int. Ed. **2000**, *39*, 4130.

ETH zürich

Synthesizable bioactive compounds

CATS: Chemically Advanced Template Search

"Scaffold Hop"



First automated ligand-based de novo design

hKv1.5 channel blocker

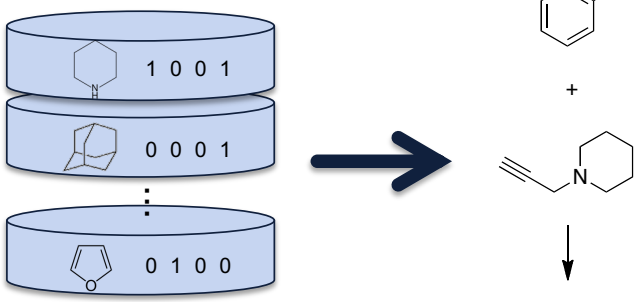
Template	Design 1	Design 2
IC_{50} = 0.11±0.02 µM	IC_{50} = 7.3±0.2 µM	IC_{50} = 0.5±0.1 µM
	3-step synthesis	

Angew. Chem. Int. Ed. **2000**, *39*, 4130.
Angew. Chem. Int. Ed. **1999**, *38*, 2894.

The virtual chemist: Reaction−based molecular scaffold hopping

DOGS: Design of Genuine Structures

Structure generation & optimization



Building blocks & Transformations

1 step: 7.8×10^{7}

2 steps: 2.6×10^{15}

3 steps: 7.6×10^{30}

Virtual products

Scoring & selection

Machine learning models

Similarity

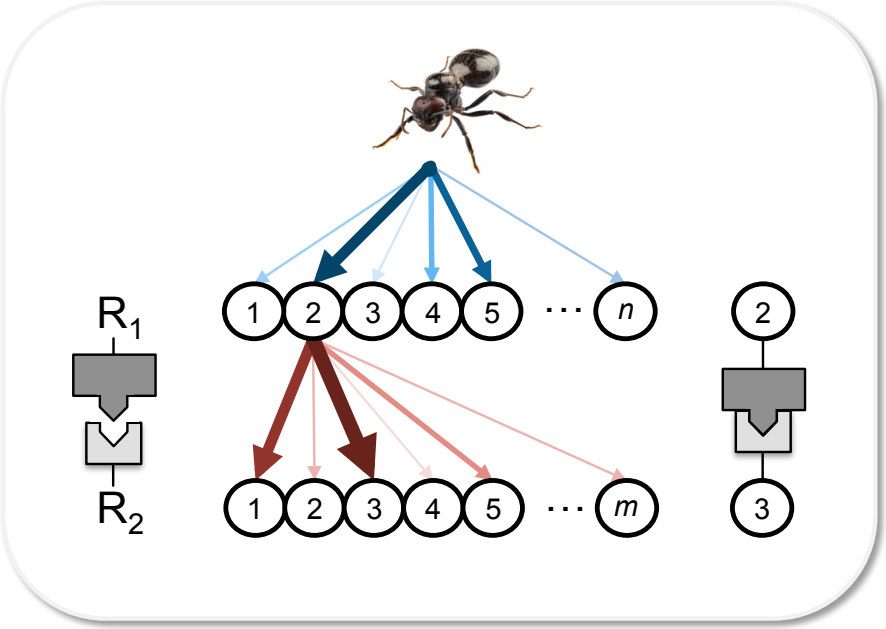
Heuristics, Intuition

Drug Discov. Today Technol. **2013**, *10*, e453.

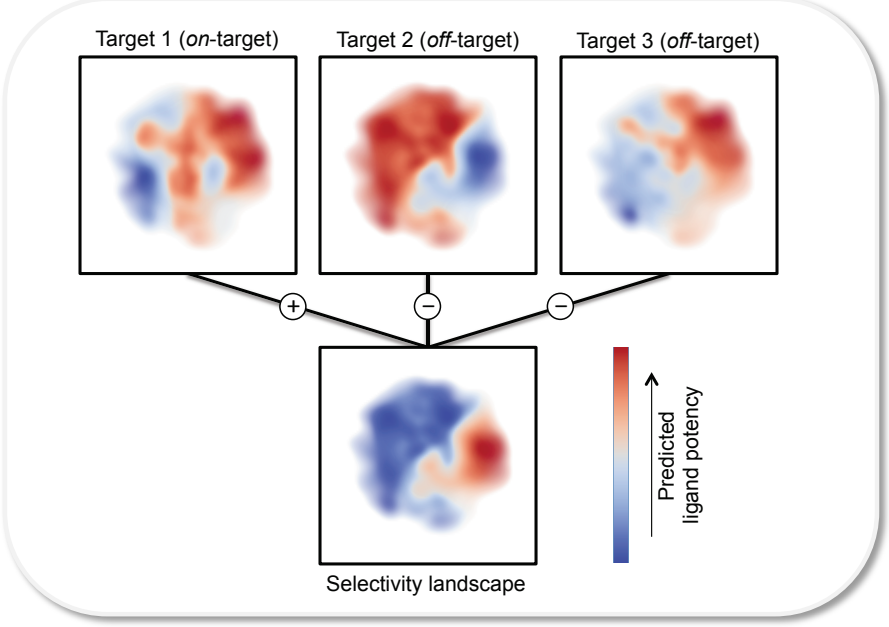
Combinatorial optimization with multiple objectives

MAntA: Molecular Ant Algorithm

Structure generation & optimization



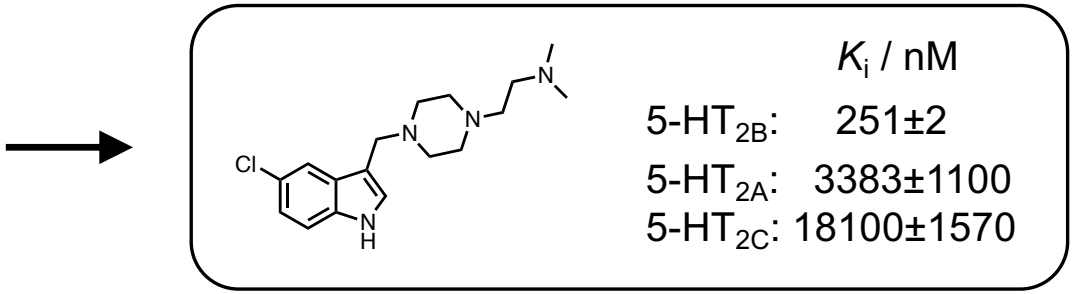
Scoring & selection

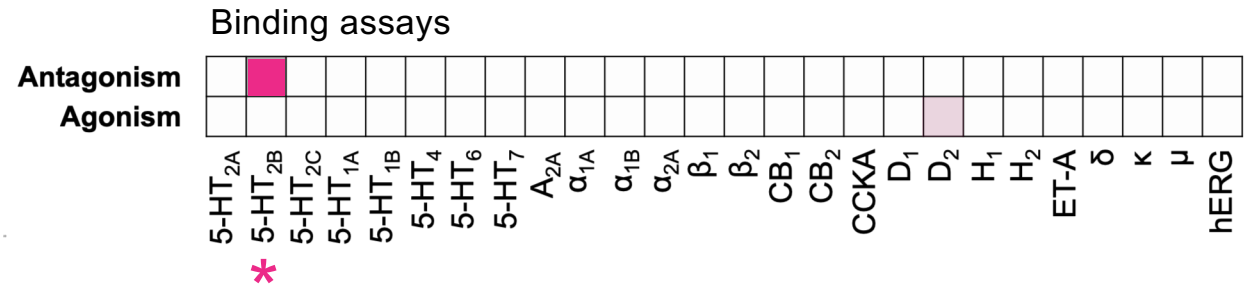


Future Med. Chem. **2017**, *9*, 381.
Angew. Chem. Int. Ed. **2014**, *53*, 4244.

Molecular de novo design with 26 objectives

- Goal: selective 5-HT_{2B} antagonist
- **MAntA** design approach
- Target prediction: GP models

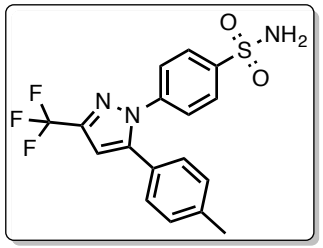




Angew. Chem. Int. Ed. **2015**, *54*, 1551.

The scoring challenge: Ligand-based prediction of drug targets

TIGER: Target Inference Generator

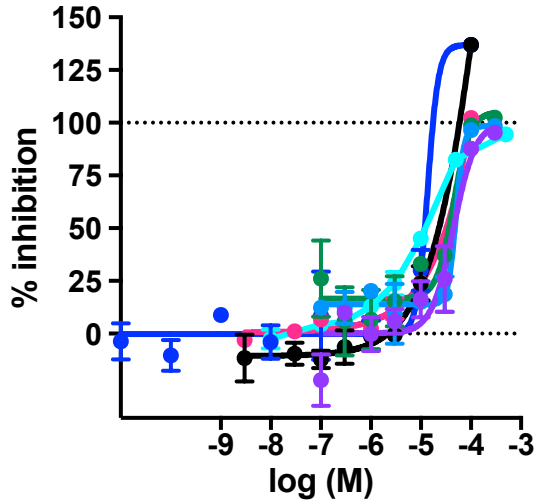


Celecoxib

COX2

IC_{50} = 0.002–4 µM
(K_B = 10–15 µM)

Gierse *et al.* **1999**
Bento *et al.* **2014**



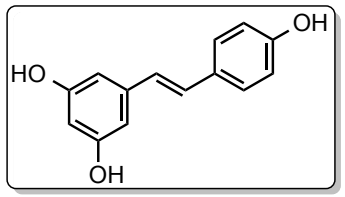
CysLT1	IC_{50} = 13 µM (K_B = 3 µM)
hERG	IC_{50} = 45 µM (K_B = 3 µM)
β1AR	IC_{50} = 50 µM (K_B = 11 µM)
β2AR	IC_{50} = 44 µM (K_B = 7 µM)
OX1	IC_{50} = 12 µM (K_B = 3 µM)
OX2	IC_{50} = 17 µM (K_B = 5 µM)
GC-R	IC_{50} = 11 µM (K_B = 6 µM)

✓ 11 of 20 (55%) predicted targets confirmed
✓ On- and off-targets identified

Angew. Chem. Int. Ed. **2017**, *56*, 11520.
Expert Opin. Drug Discov. **2017**, *12*, 271.

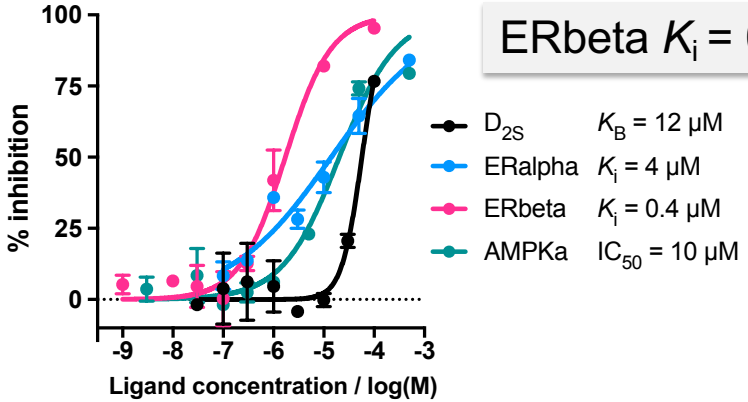
ETH zürich

Predicting the targets of bioactive natural products



Resveratrol
(cardiopreventive phytoalexin)

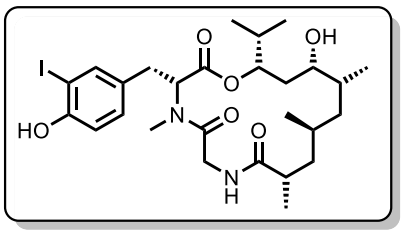
TIGER



ERbeta K_i = 0.4 µM

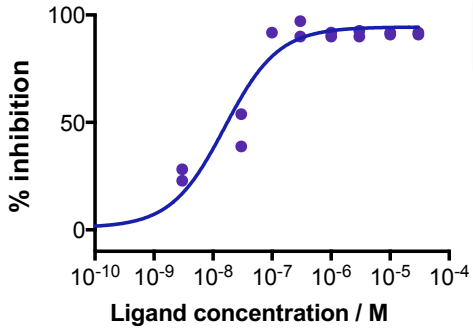
D_{2S} K_B = 12 µM
ERalpha K_i = 4 µM
ERbeta K_i = 0.4 µM
AMPKa IC_{50} = 10 µM

Anaspida sp. (sea hare)



Doliculide
(anticancer depsipeptide)

TIGER

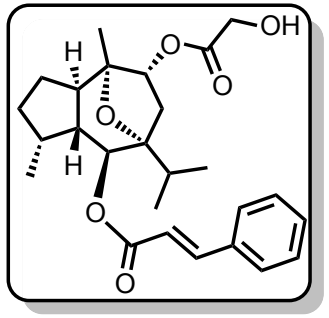


EP3 K_B = 6 nM

Angew. Chem. Int. Ed. **2017**, *56*, 11520.
Angew. Chem. Int. Ed. **2016**, *55*, 12408.
Nat. Chem. **2014**, *6*, 1072.

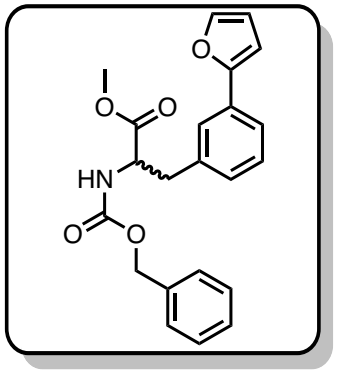
Ligand-based scaffold hopping from a structurally intricate natural product

Phyllantus Engleri
(potato bush)



Englerin A
(anticancer sesquiterpene)

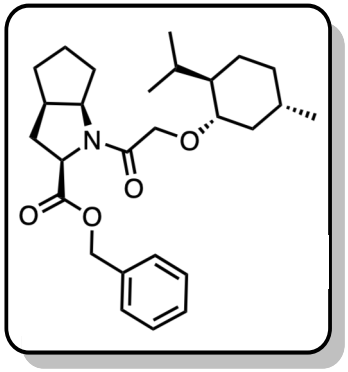
DOGS



Design 1

2D pharmacophore scoring (CATS)

K_B = 0.2 µM



Design 2

Shape-only scoring

K_B = 0.2 µM

TIGER

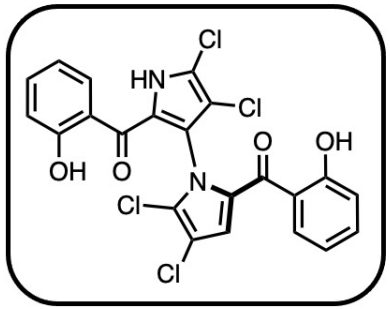
TRPM8 blocker
K_B = 0.4 µM

- ✓ 3-step synthesis as predicted
- ✓ The generated molecules “inherited” TRPM8 activity

ChemMedChem **2020**, *15*, 566.
Angew. Chem. Int. Ed. **2016**, *55*, 6789.

Ligand-based de novo design of marinopyrrole A mimetics

marine *Streptomyces*
(Al-Dhabi *et al.* 2018)

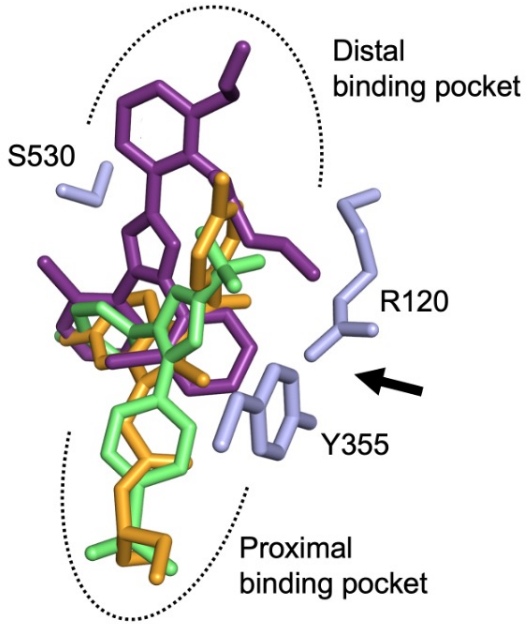


Marinopyrrole A
(anticancer, antiinfective)

TIGER

COX1 inhibitor
IC_{50} = 17 µM

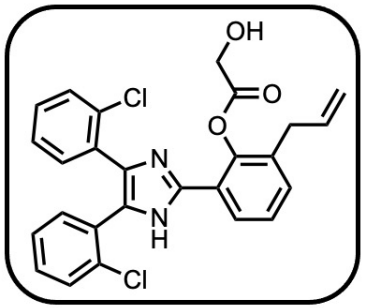
de novo design
celecoxib
indomethacin



*o*COX-1
PDB-ID: 7JXT

IC_{50} COX1 = 0.1 µM
IC_{50} platelets = 0.009 µM

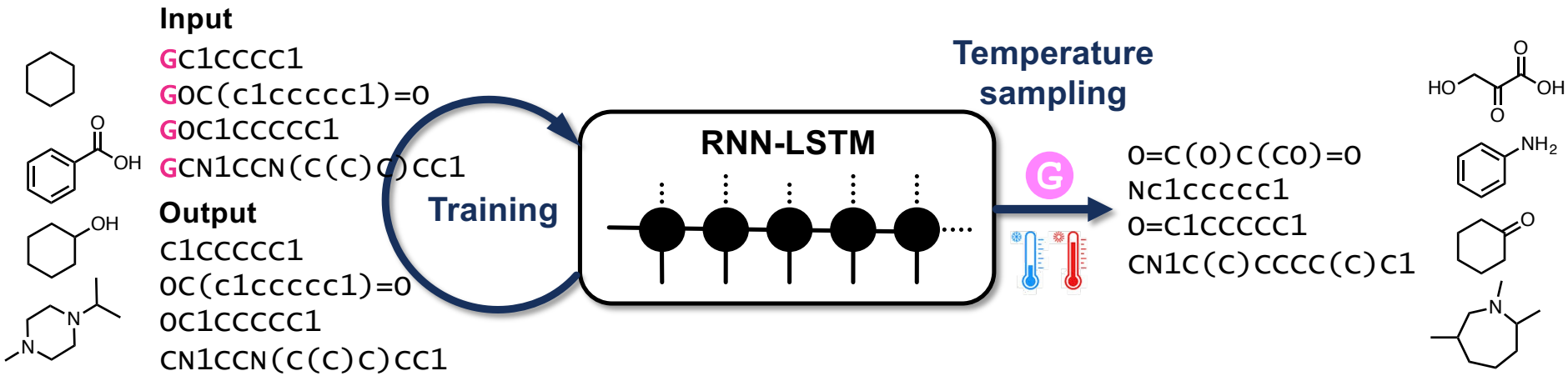
- ✓ 3-step synthesis as predicted
- ✓ Selective COX1 inhibition
- ✓ 7/8 targets shared with marinopyrrole A



De novo design

Adv. Sci. **2021**, 2100832.

Chemical language models: Recursive networks with LSTM cells



Little big data **Generative AI model** **New Molecules**

unidirectional bidirectional

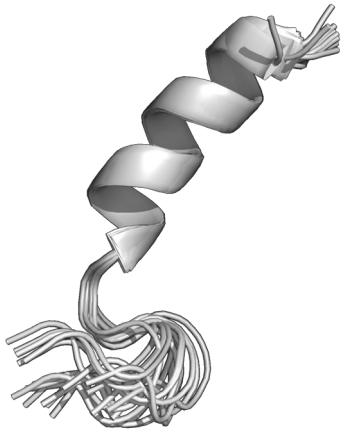
J. Chem. Inf. Model. **2020**, *60*, 1175.

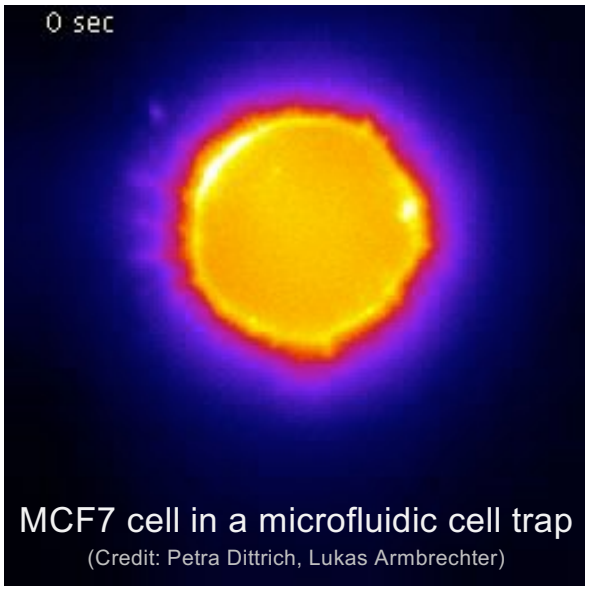
Generative peptide design

Known sequences

P(**X**,**Y**)

New sequence(s)



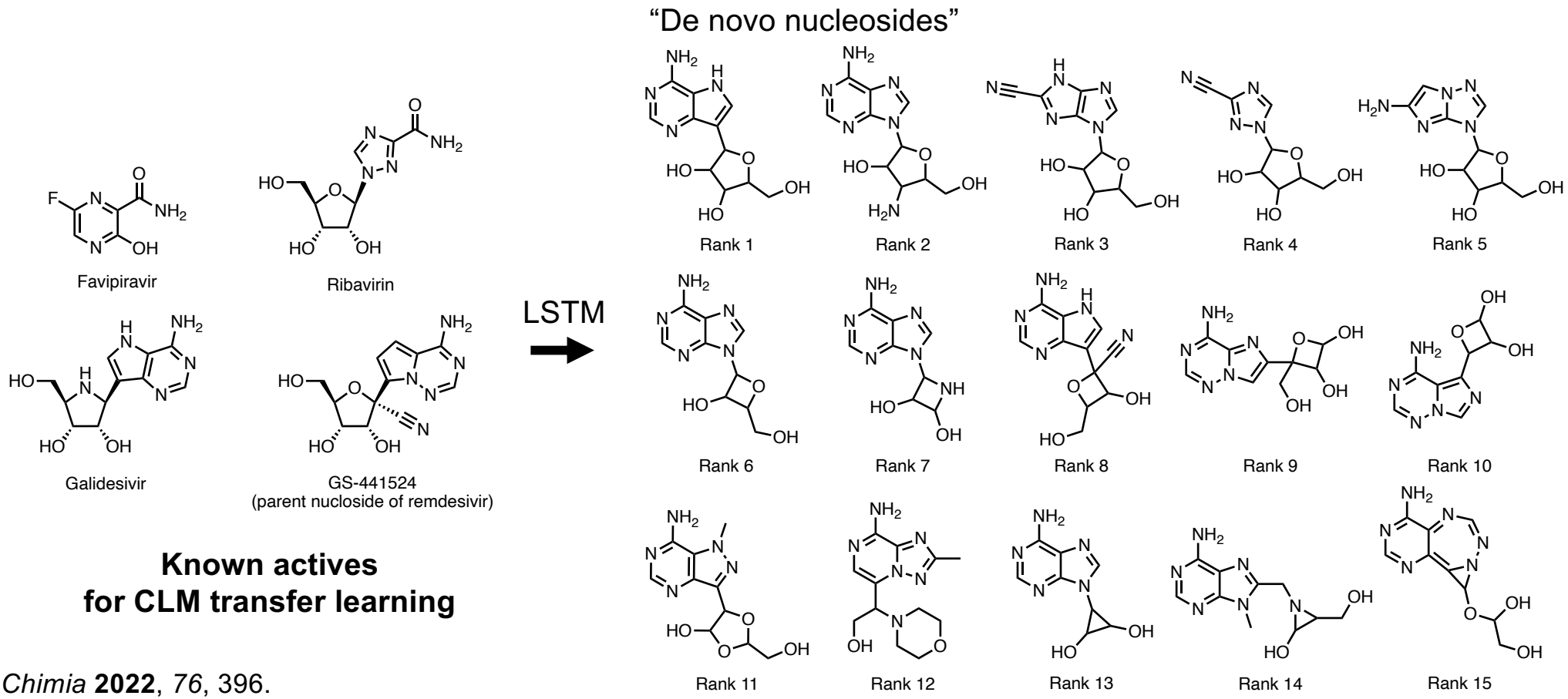


First generative model for peptide design:
Biophys. J. **1994**, *66*, 335.

Angew. Chem. Int. Ed. **2019**, *58*, 1674.
J. Chem. Inf. Model. **2018**, *58*, 472.

ETH zürich

Generative design of potential SARS-CoV-2 RdRp inhibitors

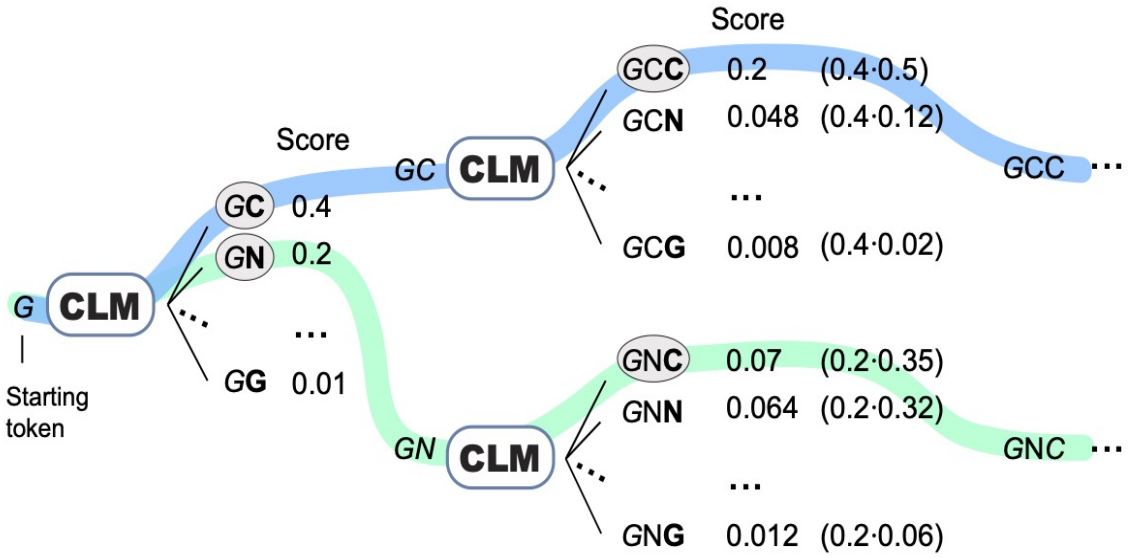


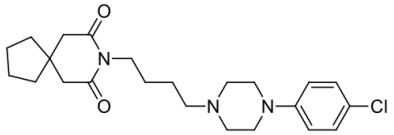
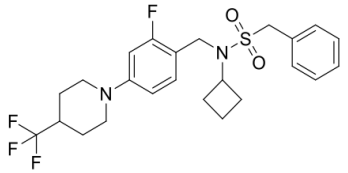
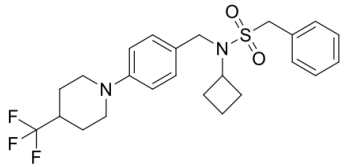
Known actives for CLM transfer learning

Chimia **2022**, *76*, 396.

ETH zürich

LSTM with transfer learning and beam-search scoring



De novo design	IC_{50}(RORγ)
	4.6±0.5 μM
	0.37±0.05 μM
	0.68±0.07 μM

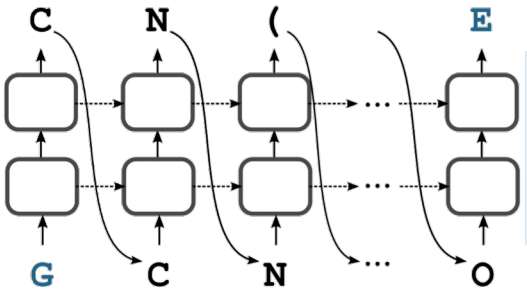
- Training on bioactive molecules from ChEMBL
- Transfer learning with LXR agonists
  (3 natural products, 1 synthetic)

Angew. Chem. Int. Ed. **2021**, *60*, 202104405.

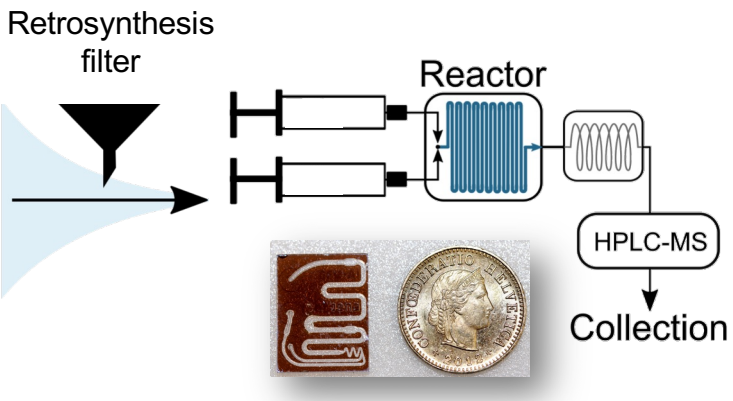
ETH zürich

Closing the loop: Design-Make-Test-Analyze

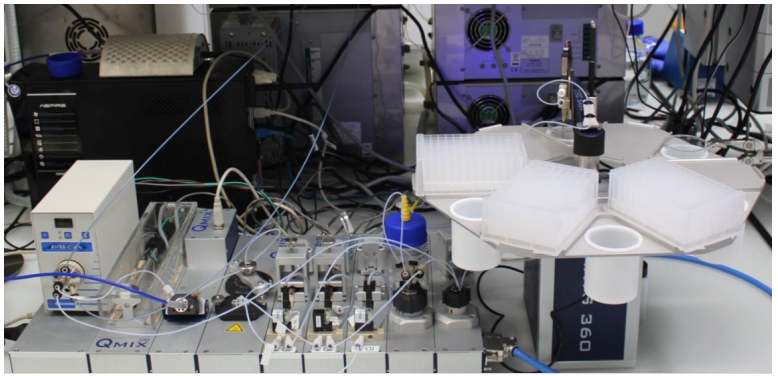
Chemical language model



On-chip synthesis



Automated synthesis with inline analytics



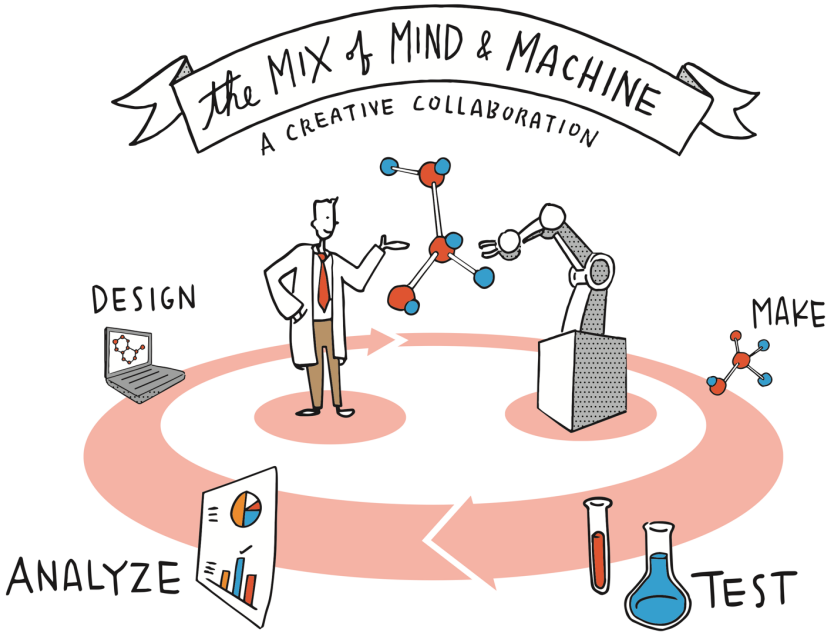
- Training on synthesizable compounds (vendor catalogs)
- Transfer learning with 40 LXR agonists
- Retrosynthesis filter for 17 reactions
- Target prediction

- 64% successfully synthesized
- 68% bioactive
- Fast, efficient, economic

Sci. Adv. **2021**, *7*, eabg3338.

ETH zürich

Takeaways



- Drug design with machine intelligence works
- Surprising solutions
- Faster/better decision making
- Most of the de novo designs are synthetically feasible, some have desired activity
- Bottlenecks: Molecular representation, scoring, experimental feedback
- We need methods that work with “little data”
- Let’s not fool ourselves: AI is not a magic hammer (“edge of chaos”)
- Progress by **deep learning & deep thinking**

Nat. Rev. Drug Discov. **2020**, *19*, 353.

ETH zürich

Molecular design with AI is a team effort

modlab

inSili.com

FNSNF Fonds national suisse Schweizerischer Nationalfonds Fondo nazionale svizzero Swiss National Science Foundation

DFG Deutsche Forschungsgemeinschaft

NOVARTIS | The Novartis Foundation

OPO STIFTUNG

BEILSTEIN INSTITUT

Marie Skłodowska-Curie Actions

GDCh

FCI Fonds der Chemischen Industrie

Aeschimann, W., Alamuri, P., Alanine, A., Alig, L., Allenspach, M. D., Atz, K., Bauer, C. A., Baumann, K., Bender, A., Bernegger, S., Bieler, M., Böcker, A., Böhm, H. J., Borchers, S., Brand, S., Bremm, M., Bruns, D., Button, A., Button, A. L., Byrne, R., Byvatov, E., Carrasco-Gomez, R., Cui, J., Deng, W., Derksen, S., Emanuelsson, O., Esser, J., Fechner, U., Feißt, C., Finkelmann, A. R., Fino, R., Fjell, C. D., Flesch, D., Franke, L., Friedrich, L., Fuchs, J. A., Funatsu, K., Gabernet, G., Gawehn, E., Geppert, T., Givehchi, A., Grabowski, K., Grisoni, F., Gulbakan, B., Gunesch, A. P., Gupta, A., Haehnke, V., Hajduk, J., Hanke, T., Hartenfeller, M., Henkel, A., Hieke, M., Hiss, J., Hiss, J. A., Hiß, J., Hofmann, B., Horvath, D., Hoy, B., Hähnke, V., Hüsch, J., Isert, C., Jagla, B., Jiménez-Luna, J., Jäger, N., Keppner, S., Kirchmair, J., Klenner, A., Knauer, S. K., Koch, C. P., Koch, G., Koeberle, A., Koenig, M., Krüger, B., Kumar, K. S., Kunze, J., Köhler, M., Lanig, H., Lee, M. L., Leimbacher, M., Leuner, K., Li, X., Lieder, F., Lin, Y. C., Lohmann, R., Lötsch, J., Löwer, M., Lüthi, H. P., Martin, O., Meissner, M., Merk, D., Miyao, T., Moret, M., Muratov, E. N., Müller, A. T., Neuhaus, C. S., Noeske, T., Paetz, J., Perković, M., Perna, A. M., Persch, E., Pillong, M., Proschak, E., Rabal, O., Reisen, F., Reker, D., Renner, S., Resch, E., Reutlinger, M., Rich, A. S., Roche, O., Rodrigues, T., Roersch, F., Rogers–Evans, M., Rupp, M., Rödl, C. B., Rörsch, F., Sabiani, S., Sander, K., Schmid, G., Schmidt, T. P., Schmuker, M., Schneider, P., Schuchhardt, J., Schulz, M. M. P., Schüller, A., Siemoneit, U., Spork, S., Spänkuch, B., Stahl, M., Stauch, B., Steri, R., Stutz, K., Tanrikulu, Y., Tausch, L., Tegtmeyer, N., Thelemann, J., Thieme, T. M., Todoroff, N., Urbanek, D. A., Wang, Q., Weisel, M., Werner, T., Werz, O., Wessler, S., Wisniewska, J. M., Wrede, P., Wu, N., Yang, X., Zander, J., Zettl, H., Zuegge, J.

… and many more.

ETH zürich