De Novo Drug Design

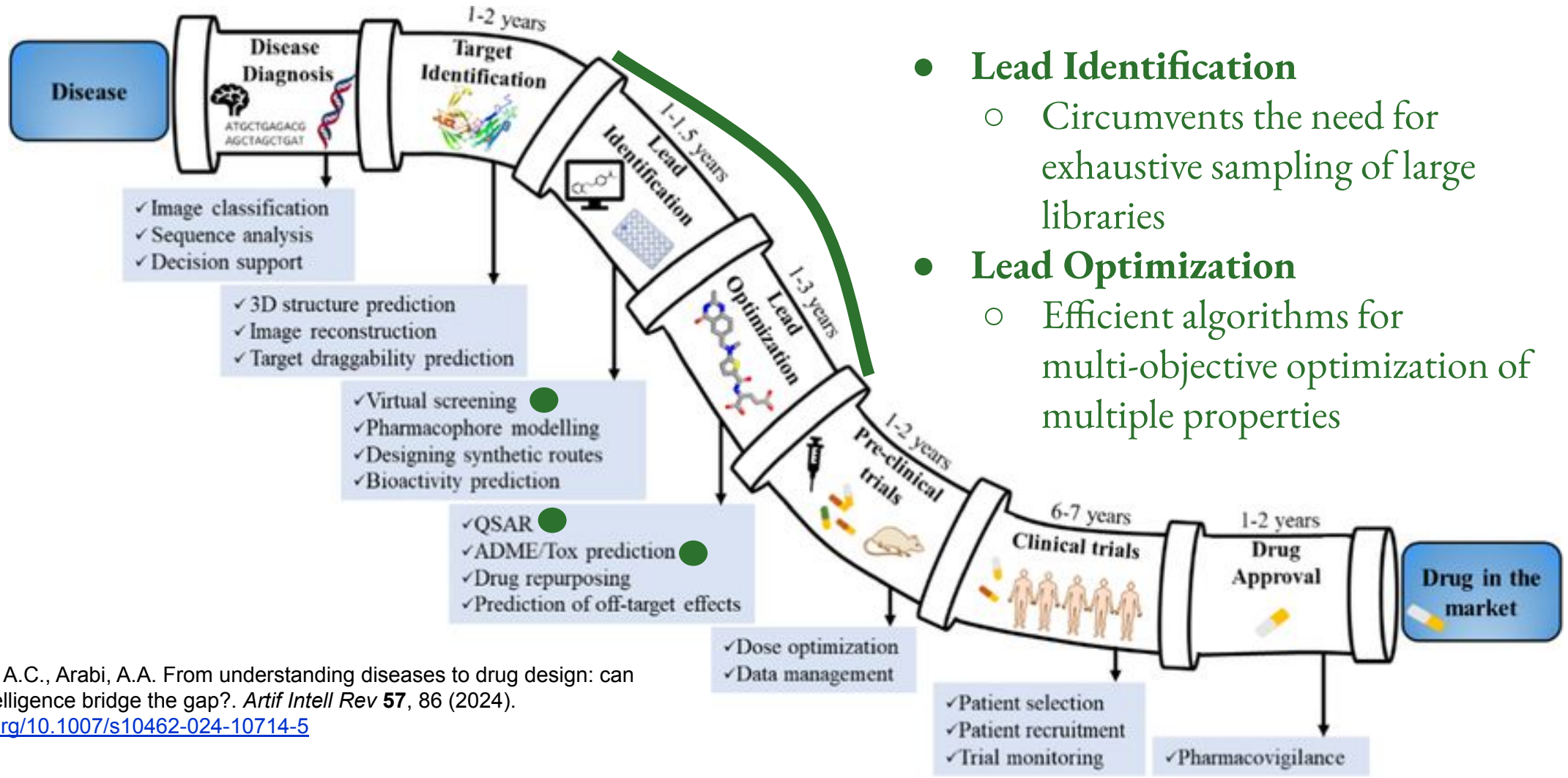
Martin Šícho (Martin.Sicho@vscht.cz) – 8ADD Olomouc – 2025-01-30

Universiteit Leiden

UNIVERSITY OF CHEMISTRY AND TECHNOLOGY PRAGUE

LACDR

De Novo Drug Design (DNDD) in the Pipeline



- **Lead Identification**
  - Circumvents the need for exhaustive sampling of large libraries
- **Lead Optimization**
  - Efficient algorithms for multi-objective optimization of multiple properties

Pushkaran, A.C., Arabi, A.A. From understanding diseases to drug design: can artificial intelligence bridge the gap?. *Artif Intell Rev* **57**, 86 (2024). https://doi.org/10.1007/s10462-024-10714-5

The Case for De Novo Drug Design

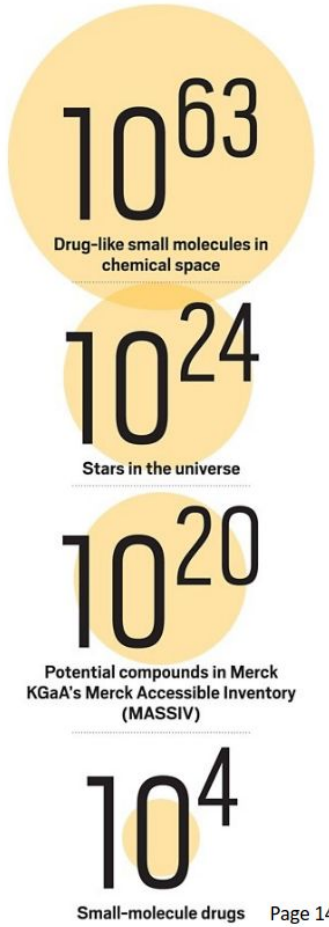
universität wien

Dimensions of the chemical space

Type	# molecules
Particles in the (observable) universe	10^{82}
Molecules < 1000 Da consisting of C, N, O, P, S, Hal, H	Up to 10^{180}
Drug-like molecules based on extrapolation on GDB-17[2] based on stitching together up to 30 carbon, nitrogen, oxygen, and sulfur atoms in different arrangements	 10^{33} 10^{63}
Merck Accessible Inventory (MASSIV)	10^{20}
Chemical universe database GDB-17: Listing all molecules up to 17 atoms[1]	166,400,000,000
Make-on-demand compounds in the public domain	> 5,000,000,000
On-stock compounds	230,000,000
Known natural products	700,000
Purchasable natural products	25,000[3]

[1] Ruddigkeit L et al. J Chem Inf Model 2012, 52, 2864–2875. doi: 10.1021/ci300415d
[2] Polishchuk PG, Madzhidov TI, Varnek A. J Comput-Aided Mol Des 2013, 27, 675–679.
[3] Chen Y. et al., J Chem Inf Model 2017, 57, 2099–2111.

27.01.2025 Figure: https://cen.acs.org/pharmaceuticals/drug-discovery/Hunting-drugs-chemical-space/100/i23



Kirchmair et al., 8ADD Olomouc, 2025, https://www.kfc.upol.cz/wp-content/uploads/2025/01/FOR-PUBLICATION_kirchmair_introduction-to-cheminformatics.pdf

Universiteit Leiden

UNIVERSITY OF CHEMISTRY AND TECHNOLOGY PRAGUE

LACDR

De Novo Drug Design: Yesterday and Today

1700 J. Med. Chem. 1993, 36, 1700–1710

GroupBuild: A Fragment-Based Method for *De Novo* Drug Design

Sergio H. Rotstein and Mark A. Murcko*

Vertex Pharmaceuticals Incorporated, 40 Allston Street, Cambridge, Massachusetts 02139-4211

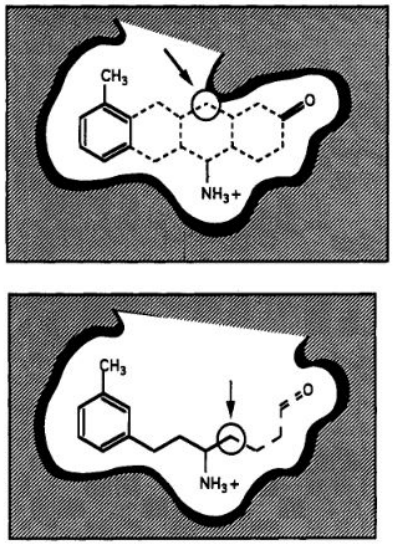
Received February 9, 1993

A novel method for *de novo* drug design, GroupBuild, has been developed to suggest chemically reasonable structures which fill the active sites of enzymes. The proposed molecules provide good steric and electrostatic contact with the enzyme and exist in low-energy conformations. These structures are composed entirely of individual functional groups (also known as "building blocks" or "fragments") which the program chooses from a predefined library. User-selected enzyme seed atom(s) may be used to determine the area(s) in which structure generation begins. Alternatively, GroupBuild may begin with a predocked "inhibitor core" from which fragments are grown. For each new fragment generated by the program, several thousand candidates in a variety of locations and orientations are considered. Each of these candidates is scored based on a standard molecular mechanics potential function. The selected fragment and orientation are chosen from among the highest scoring cases. Tests of the method using HIV protease, FK506 binding protein, and human carbonic anhydrase demonstrate that structures similar to known potent inhibitors may be generated with GroupBuild.

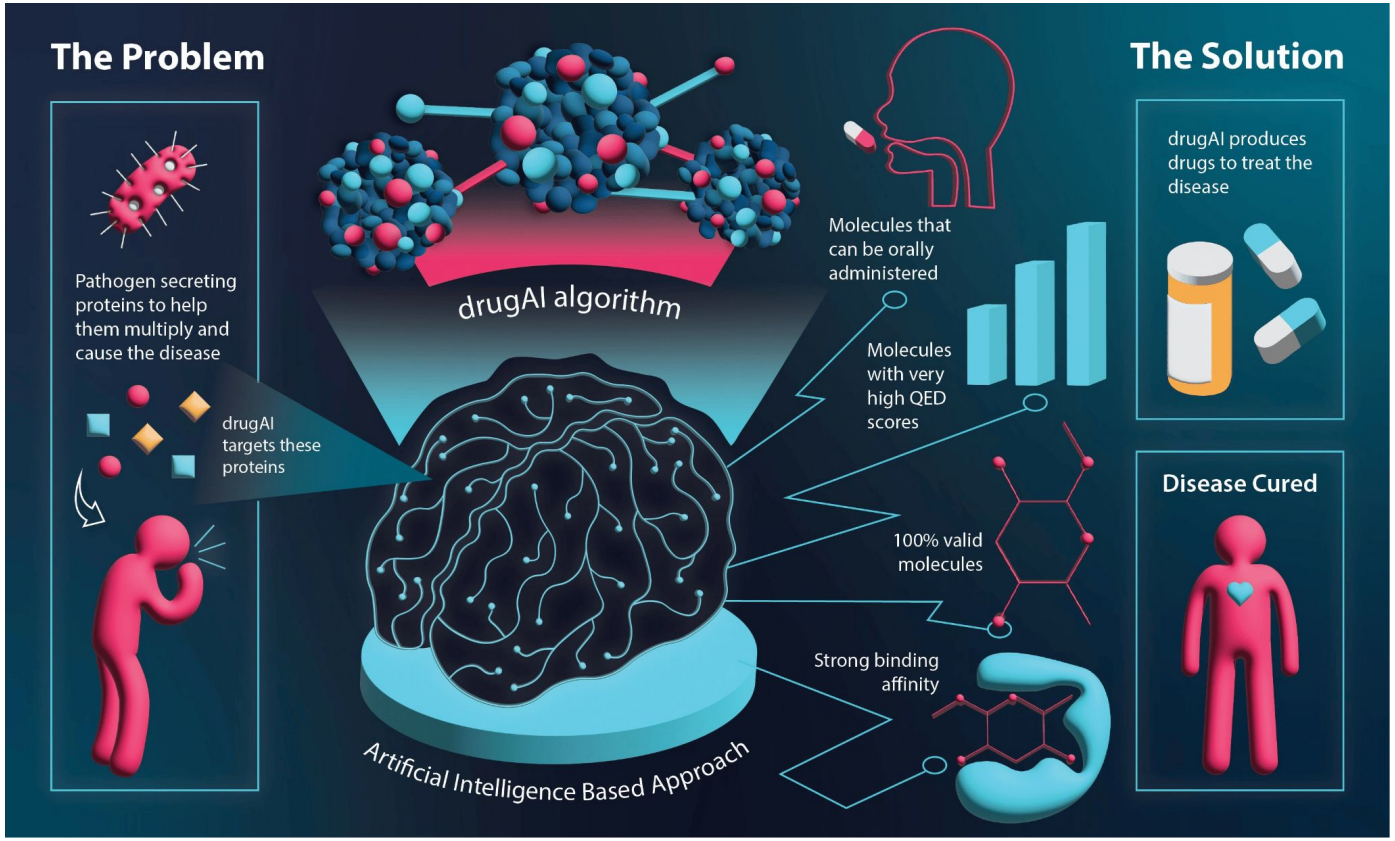
Table I. Current Fragment Library

Acid	Aldehyde	Amide
Amine	Benzene	Cyclohexane
Cyclopentane	Ethane	Ethylene
Hydroxy	Methoxy	Methane
Sulfone	Thiophene	

combine



Pharmaceuticals **2024**, *17*(2), 161; https://doi.org/10.3390/ph17020161

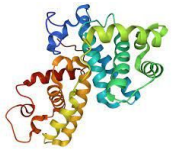


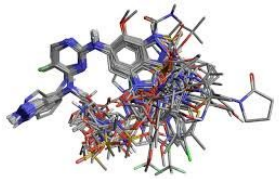
Universiteit Leiden

UNIVERSITY OF CHEMISTRY AND TECHNOLOGY PRAGUE

LACDR

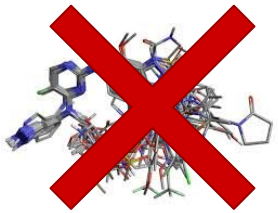
When and How?





- in pocket generation[1,2]
  - fragment-based vs. atom-based
- 2D ligand-based generation with 3D structure scoring[1]

- 2D/3D ligand-based generation with scoring[4]
  - QSAR
  - Pharmacophore Model
  - Shape Similarity
  - ...



- in pocket generation[1,2]
  - fragment-based vs. atom-based
- stable diffusion[3]

- "multi-modal" generative models[5,6]
- DEL and combinatorial library design
- make-on-demand libraries

1. https://cheminformantics.blogspot.com/2024/12/structure-aware-generative-molecular.html?m=1
2. https://www.biosolveit.de/application-academy/chemical-space-docking/
3. https://github.com/arneschneuing/DiffSBDD
4. Liu X, IJzerman AP, van Westen GJP. Computational Approaches for De Novo Drug Design: Past, Present, and Future. Methods Mol Biol. 2021;2190:139-165. doi:10.1007/978-1-0716-0826-5_6
5. Bernatavicius A, Šícho M, Janssen APA, Hassen AK, Preuss M, van Westen GJP. AlphaFold Meets De Novo Drug Design: Leveraging Structural Protein Information in Multitarget Molecular Generative Models. J Chem Inf Model. 2024;64(21):8113-8122. doi:10.1021/acs.jcim.4c00309
6. Bernatavicius A, Šícho M, Janssen APA, Hassen AK, Preuss M, van Westen GJP. AlphaFold Meets De Novo Drug Design: Leveraging Structural Protein Information in Multitarget Molecular Generative Models. J Chem Inf Model. 2024;64(21):8113-8122. doi:10.1021/acs.jcim.4c00309

In Pocket 3D Generation

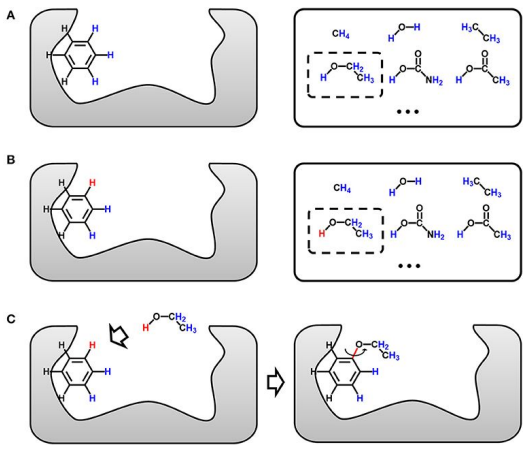


Fig 1: LigBuilder[2]

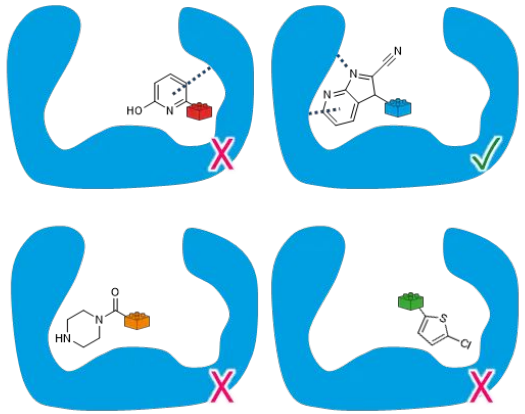


Fig 2: Chemical Space Docking[3]

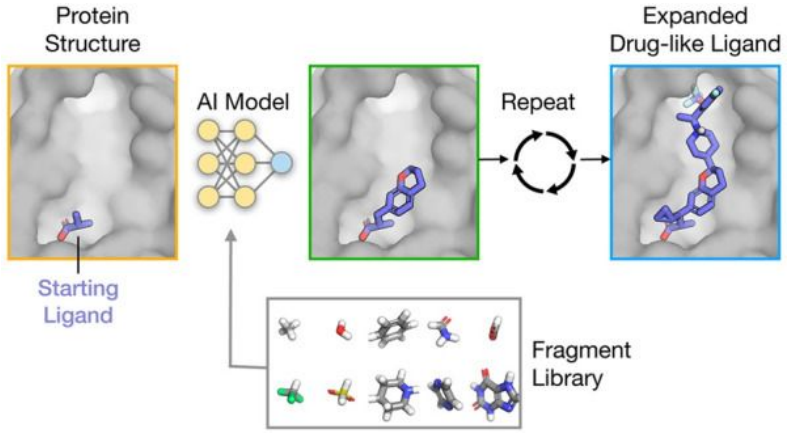


Fig 3: FRAME[5]

- Oldest methods with similar approaches appearing over the years[1]
- Employ genetic algorithms and heuristics[2,3] → obvious and transparent solution to the problem
- generative deep learning can be used as well[4,5]
- **Downsides:**
  - Library of fragments/synthons is required -> limits search space, but often necessary due to complexity in 3D
  - DL models are black boxes + produced geometries and structures are often synthetically or physically impossible

1. J. Med. Chem. 1993, 36, 12, 1700–1710
2. Yuan Y, Pei J, Lai L. LigBuilder V3: A Multi-Target de novo Drug Design Approach. Front Chem. 2020;8:142. Published 2020 Feb 28. doi:10.3389/fchem.2020.00142
3. https://www.biosolveit.de/application-academy/chemical-space-docking/
4. J. Med. Chem. 2022, 65, 13, 9478–9492
5. Powers AS, Yu HH, Suriana P, Koodli RV, Lu T, Paggi JM, Dror RO. Geometric Deep Learning for Structure-Based Ligand Design. ACS Cent Sci. 2023 Nov 17;9(12):2257-2267. doi: 10.1021/acscentsci.3c00572. PMID: 38161364; PMCID: PMC10755842.
6. J. Chem. Inf. Model. 2024, 64, 6, 1794–1805

2D Graph Generation

- By far the most explored and abundant set of methods
- **Fragment growing or joining** applicable to 2D graphs as well
  - GraphGA[2]
  - **CReM[3]**
- Deep learning models most popular recently
  - **Recurrent Neural Networks (RNNs)[1,4]**
  - Variational Autoencoders (VAEs)[1,4]
  - Generative Adversarial Networks (GANs)[1,4]
  - Normalizing flow models[4]
  - **Transformers[5]**
  - ...
- **Reinforcement Learning (RL)[1]**

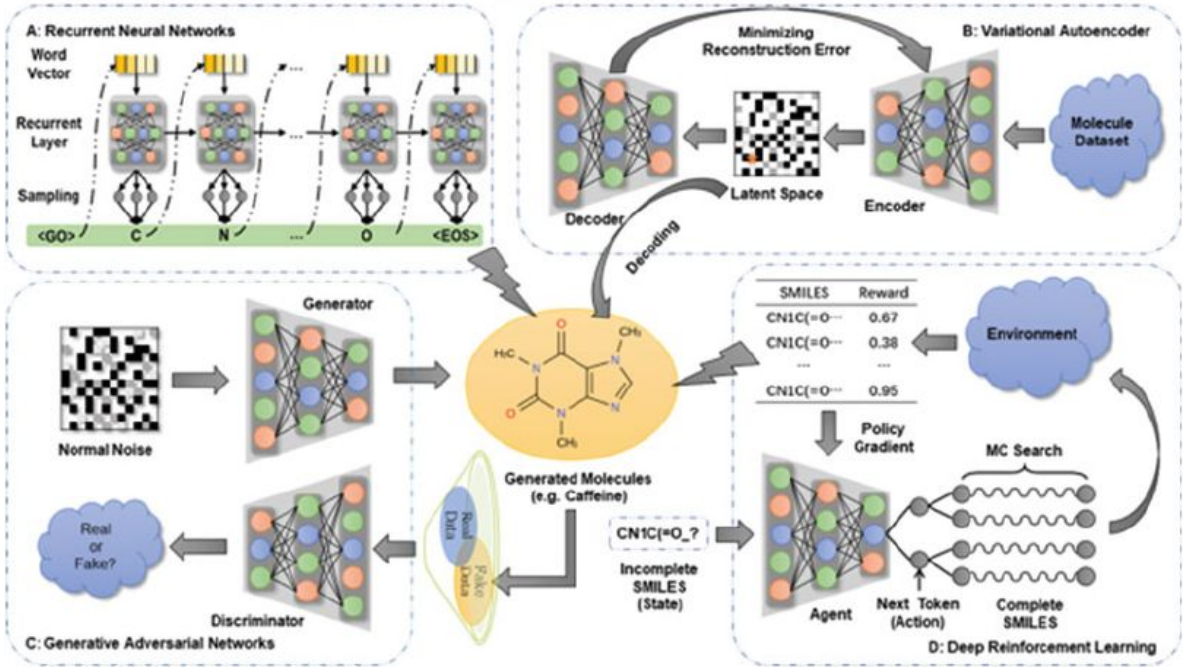
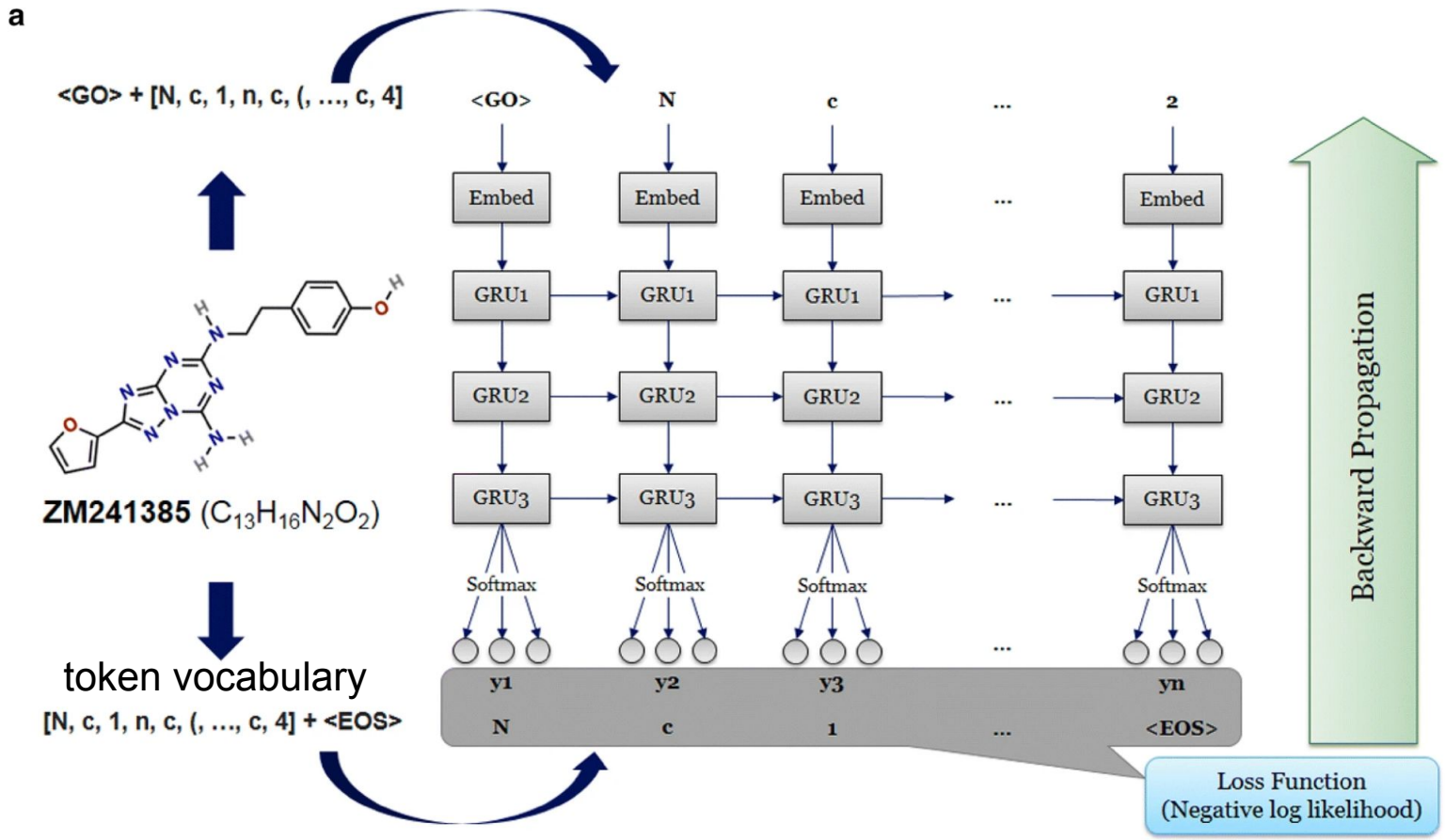


Fig 1: Overview of Common DL Architectures[1].

1. Liu X, IJzerman AP, van Westen GJP. Computational Approaches for De Novo Drug Design: Past, Present, and Future. Methods Mol Biol. 2021;2190:139-165. doi:10.1007/978-1-0716-0826-5_6
2. Jensen JH. A graph-based genetic algorithm and generative model/Monte Carlo tree search for the exploration of chemical space. Chem Sci. 2019 Feb 11;10(12):3567-3572. doi: 10.1039/c8sc05372c. PMID: 30996948; PMCID: PMC6438151.
3. Polishchuk, P. CReM: chemically reasonable mutations framework for structure generation. J Cheminform 12, 28 (2020). https://doi.org/10.1186/s13321-020-00431-w
4. Bilodeau C, Jin W, Jaakkola T, Barzilay R, Jensen KF. Generative models for molecular discovery: Recent advances and challenges. WIREs Comput Mol Sci. 2022; 12:e1608. https://doi.org/10.1002/wcms.1608
5. Šícho M, Luukkonen S, van den Maagdenberg HW, Schoenmaker L, Béquignon OJM, van Westen GJP. DrugEx: Deep Learning Models and Tools for Exploration of Drug-Like Chemical Space. J Chem Inf Model. 2023;63(12):3629-3636. doi:10.1021/acs.jcim.3c00434

Recurrent Neural Networks (RNNs)

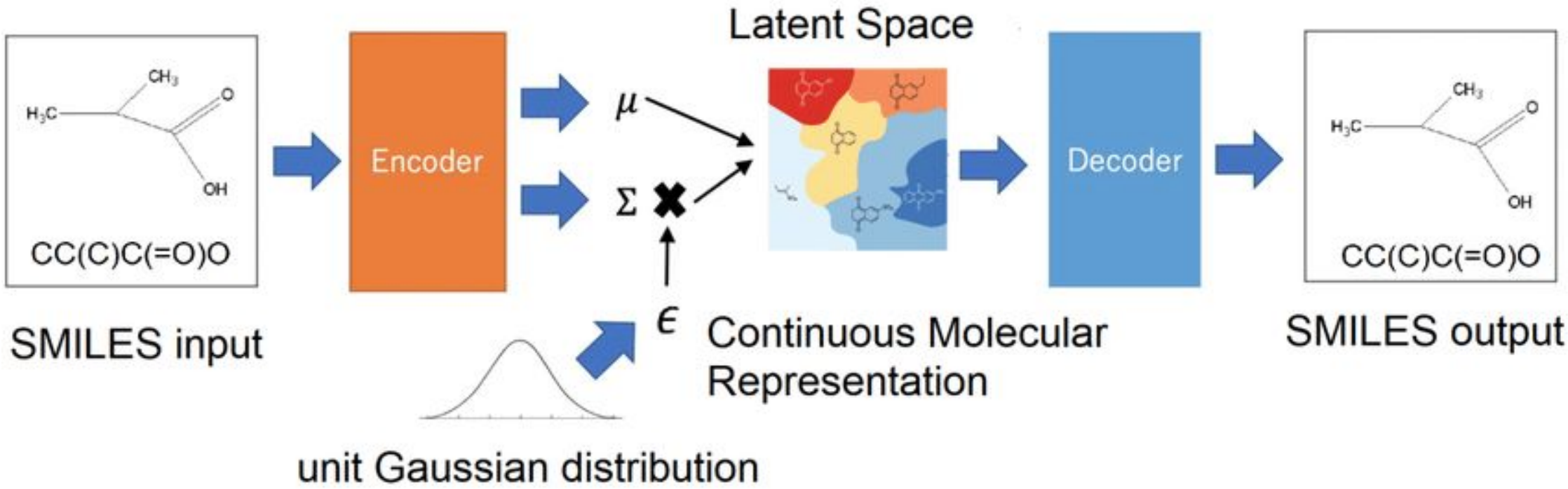


- RNNs have a **notion of 'memory'** using specialized neurons
  - GRU or LSTM cells
- for most tasks in NLP domain surpassed by transformers, but still prevalent in de novo drug design
  - RNNs **generally fail on processing longer sequences** (short memory)

1. Liu X, Ye K, van Vlijmen HWT, IJzerman AP, van Westen GJP. An exploration strategy improves the diversity of de novo ligands using deep reinforcement learning: a case for the adenosine A_{2A} receptor. J Cheminform. 2019 May 24;11(1):35. doi: 10.1186/s13321-019-0355-6. PMID: 31127405; PMCID: PMC6534880.
2. https://www.geeksforgeeks.org/rnn-vs-lstm-vs-gru-vs-transformers/

Variational Autoencoders (VAEs)

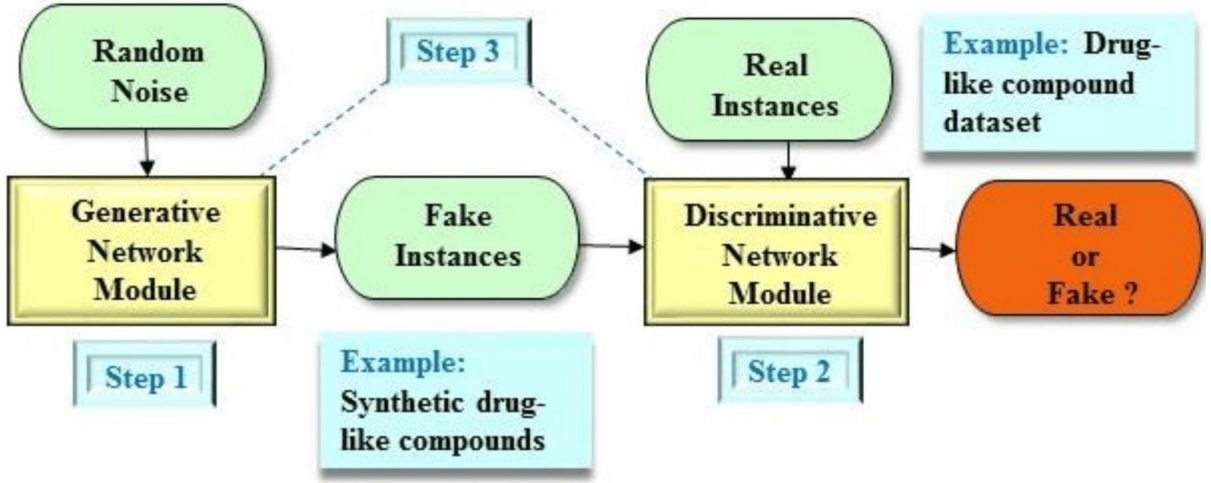
- Enable **integration of (multi-)objective optimization** into the generative process



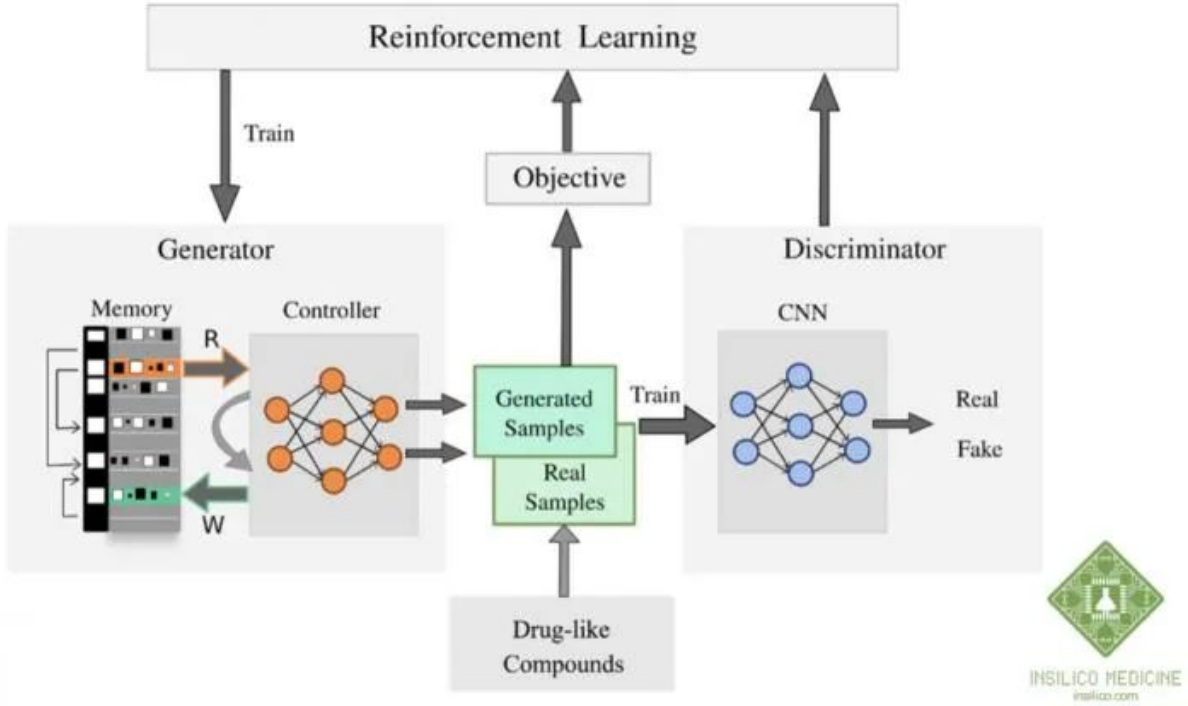
R. Wei and A. Mahmood, "Recent Advances in Variational Autoencoders With Representation Learning for Biomedical Informatics: A Survey," in *IEEE Access*, vol. 9, pp. 4939-4956, 2021, doi: 10.1109/ACCESS.2020.3048309.

Generative Adversarial Networks (GANs)

- Specialized training strategy that **introduces a discriminator network to distinguish "real" and "fake" instances**
- Can be combined with various optimization strategies for (multi-)objective optimization



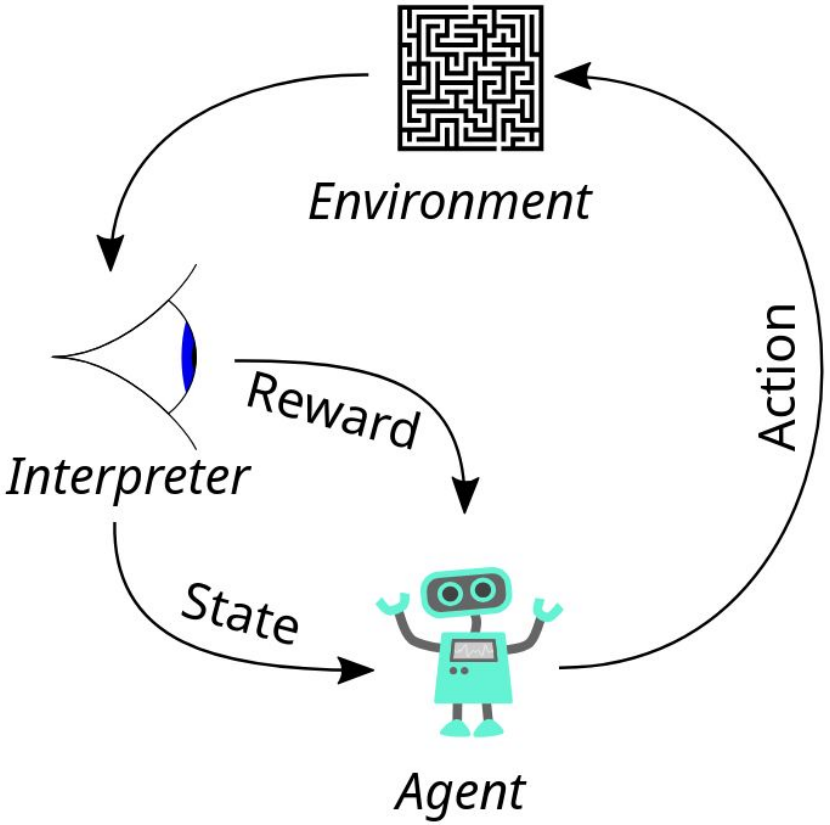
Molecules **2020**, *25*(14), 3250; https://doi.org/10.3390/molecules25143250



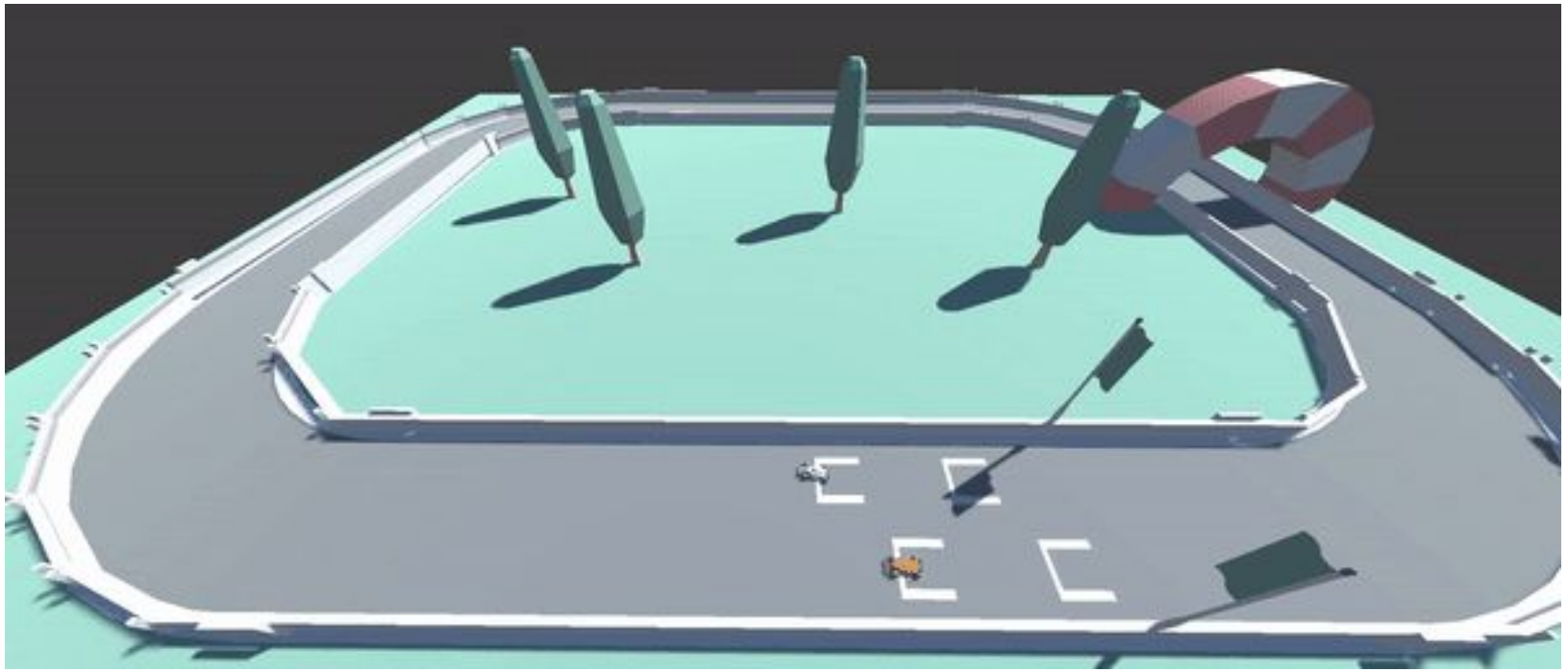
J. Chem. Inf. Model. 2018, 58, 6, 1194–1204

Reinforcement Learning (RL)

policy = sequence of **actions** the **agent** takes in the **environment** (goal = select best policy in terms of **reward**)



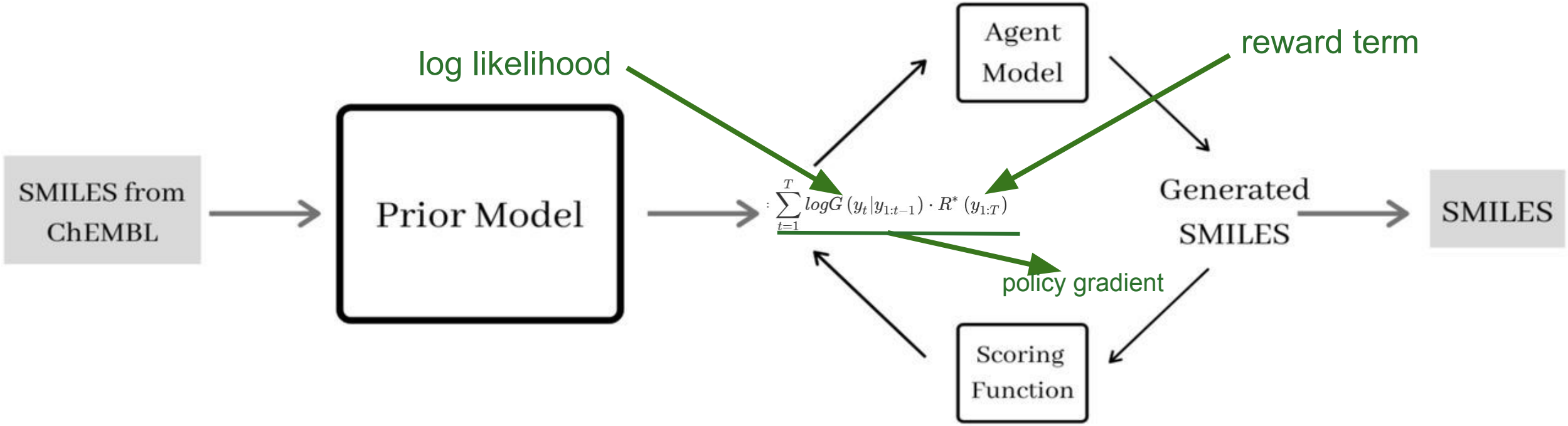
https://en.wikipedia.org/wiki/Reinforcement_learning



https://monidp.medium.com/self-driving-car-with-reinforcement-learning-in-unity-88458d13fcd1

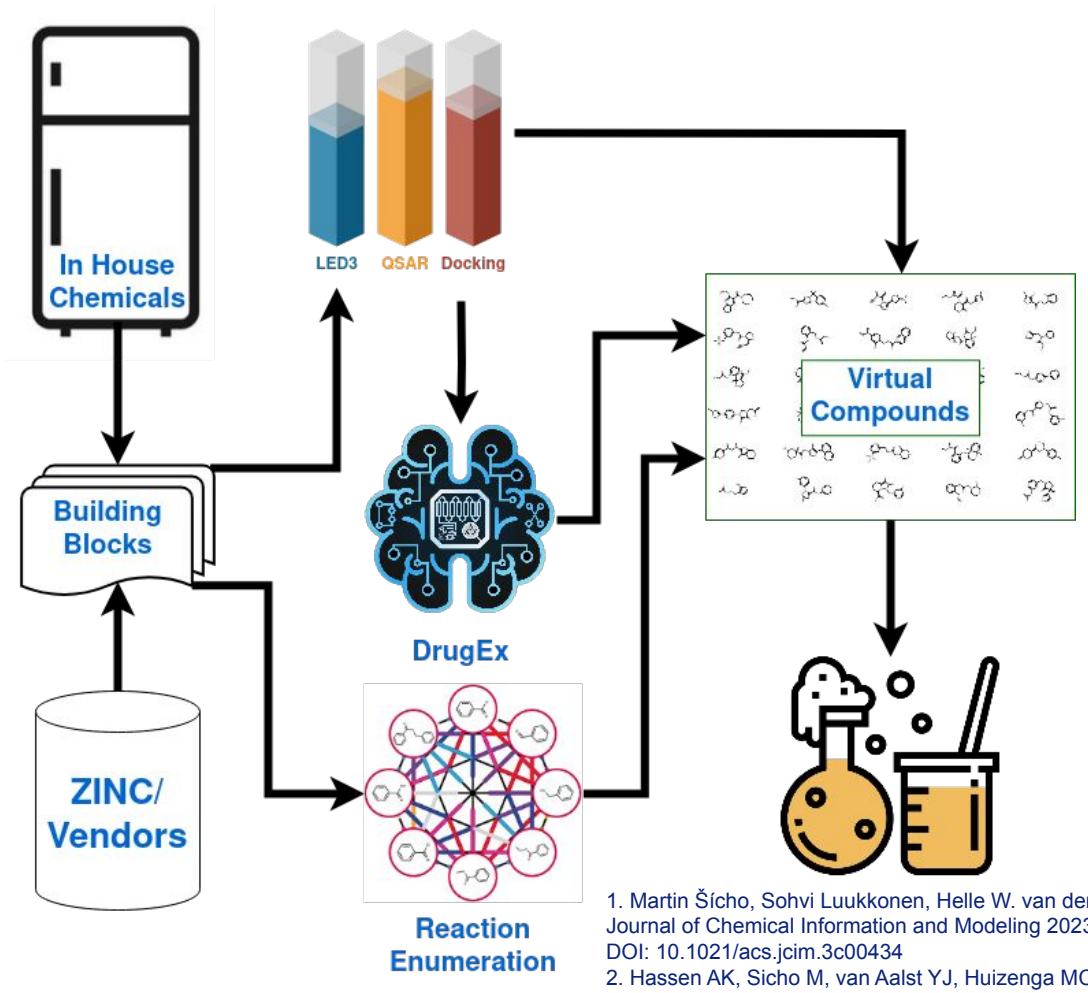
Reinforcement Learning in Molecular Generation

- RL is popular to integrate (multi-)objective optimization with generative models
  - Can be used with all model architectures -> **policy gradient**
- **BUT**, it is **highly stochastic** (can lead to unstable training) + unclear parameter optimization strategy



Testing the Limits of SMILES-based *De Novo* Molecular Generation with Curriculum and Deep Reinforcement Learning., Maranga Mokaya, Fergus Imrie, Willem P. van Hoorn, Aleksandra Kalisz, Anthony R. Bradley, Charlotte M. Deane, bioRxiv 2022.07.15.500218; doi: https://doi.org/10.1101/2022.07.15.500218

Putting Molecular Generation to Practice





- DrugEx (Molecular Generation)[1]
  - https://github.com/CDDLeiden/DrugEx
- LED3Score (SA Scoring)[2]
  - https://github.com/AlanHassen/led3_score
- QSPRPred (QSPR Modelling)[3]
  - https://github.com/CDDLeiden/QSPRPred
- Spock (Molecular Docking/SBDD)
  - https://github.com/CDDLeiden/spock
  - (available soon)
- GenUI (GUI)
  - https://github.com/martin-sicho/genui



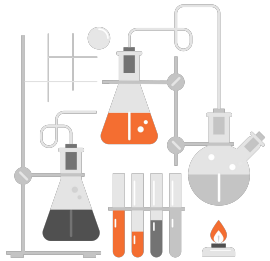



1. Martin Šícho, Sohvi Luukkonen, Helle W. van den Maagdenberg, Linde Schoenmaker, Olivier J. M. Béquignon, and Gerard J. P. van Westen
Journal of Chemical Information and Modeling 2023 63 (12), 3629-3636
DOI: 10.1021/acs.jcim.3c00434
2. Hassen AK, Sicho M, van Aalst YJ, Huizenga MCW, Reynolds DNR, Luukkonen S, et al. Generate What You Can Make: Achieving in-house synthesizability with readily available resources in de novo drug design. ChemRxiv. 2024; doi:10.26434/chemrxiv-2024-wtjt6 This content is a preprint and has not been peer-reviewed.
3. van den Maagdenberg HW, Šícho M, Alencar Araripe D, Luukkonen S, Schoenmaker L, Jespers M, et al. QSPRpred: a Flexible Open-Source Quantitative Structure-Property Relationship Modelling Tool. ChemRxiv. 2024; doi:10.26434/chemrxiv-2024-m9989 This content is a preprint and has not been peer-reviewed.

De Novo Drug Design Case Study (Monoglyceride Lipase, MGLL)

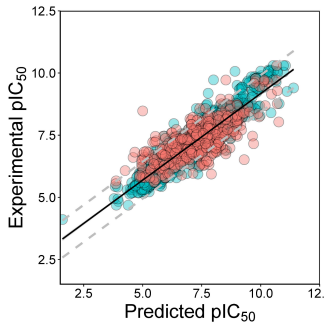
LED3Score

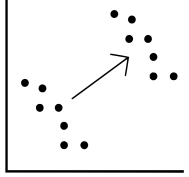
Given a set of building blocks, can we find a synthetic route to the given compound?

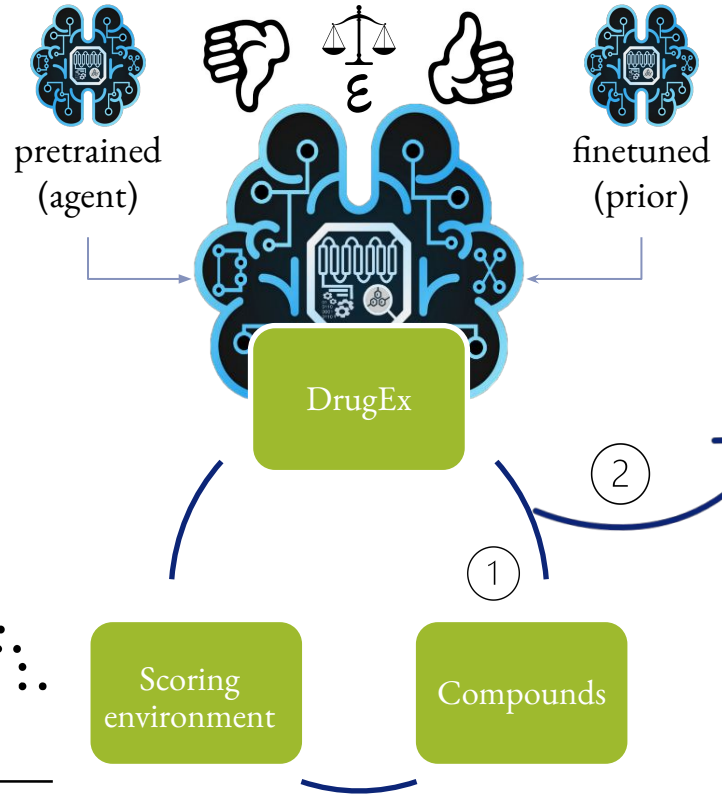


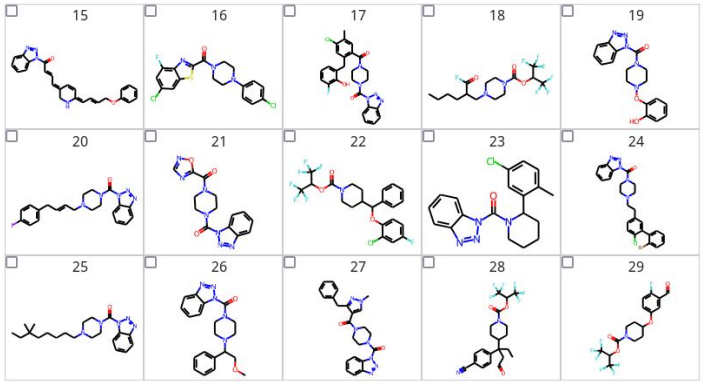
QSAR Model

Is the generated compound likely to bind?

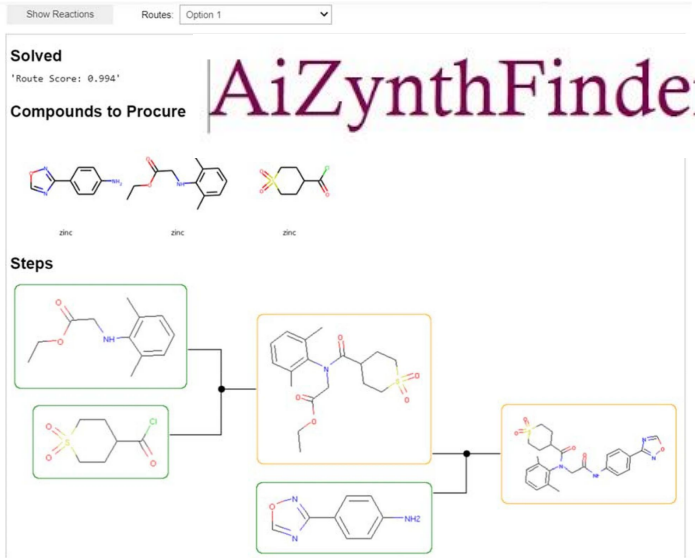








3



Reinforcement Learning Loop

DrugEx: *J. Chem. Inf. Model.* 2023, 63, 12, 3629–3636
LED3Score: *ChemRxiv. 2024; doi:10.26434/chemrxiv-2024-wtjt6 This content is a preprint and has not been peer-reviewed.*

Universiteit Leiden

UNIVERSITY OF CHEMISTRY AND TECHNOLOGY PRAGUE

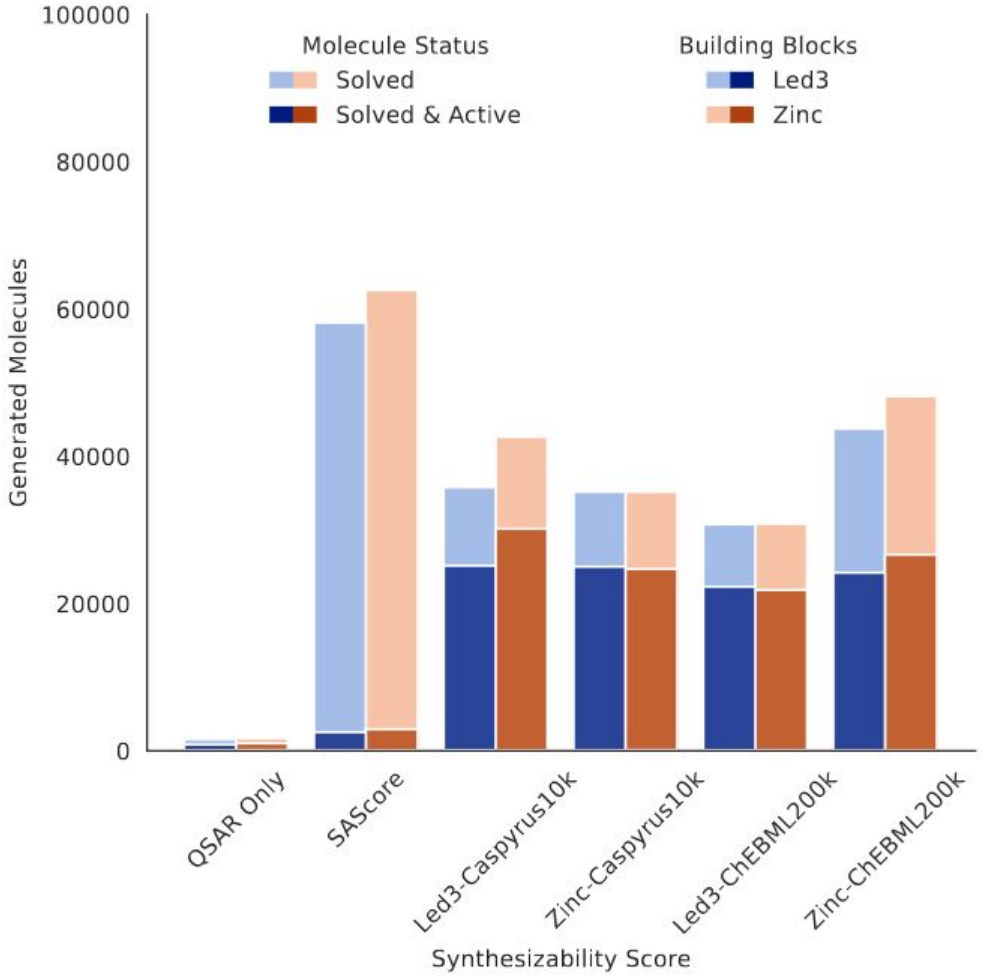
LACDR

De Novo Drug Design Case Study (Monoglyceride Lipase, MGLL)

- Workflow[1]:
  a. Train DrugEx with six different SA scores as an objective:
     - **1. None** (baseline), only the QSAR model
     - **2. SAScore** by Ertl et al.[2]
     - **ML-based:**
       - **3. LED3_casp10k**
       - **4. LED3_chembl200k**
       - **5. ZINC_casp10k**
       - **6. ZINC_chembl200k (RAScore reproduction)**
  b. Generate **100,000** structures for each of the 6 cases.
  c. Solve routes for all 6 cases with AiZynthFinder.
  d. Determine desirability of the generated structures.
  e. Answer questions:
     - **Q1: How many desired compounds with solved routes can we obtain?**
     - **Q2: What is the prediction error of the ML-based scores on the generated molecules?**
     - **Q3: Can we pick and synthesize new active molecules?**

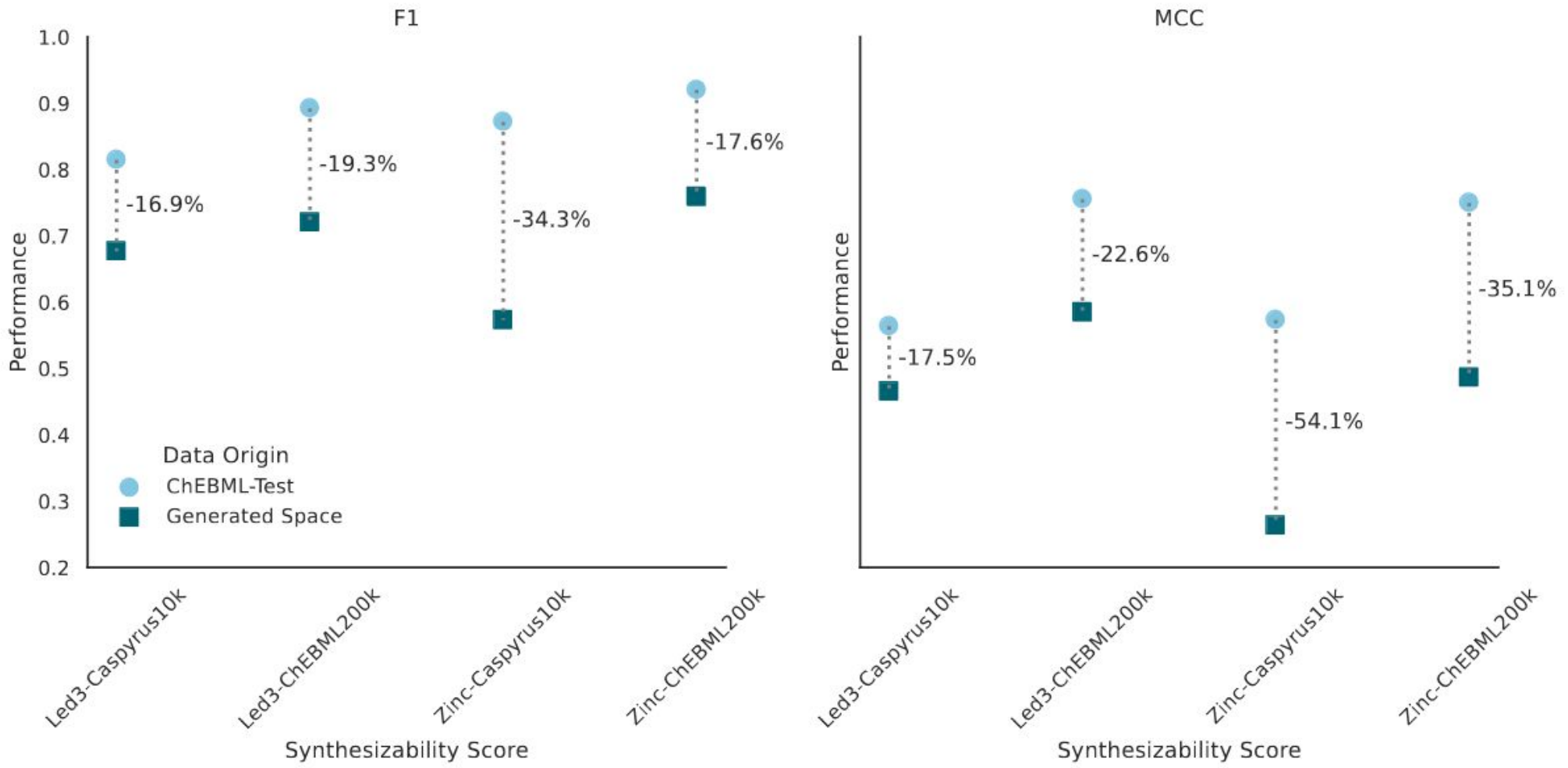
1. Hassen AK, Sicho M, van Aalst YJ, Huizenga MCW, Reynolds DNR, Luukkonen S, et al. Generate What You Can Make: Achieving in-house synthesizability with readily available resources in de novo drug design. ChemRxiv. 2024; doi:10.26434/chemrxiv-2024-wtjt6 This content is a preprint and has not been peer-reviewed. https://doi.org/10.26434/chemrxiv-2024-wtjt6
2. Ertl, P., Schuffenhauer, A. Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. J Cheminform 1, 8 (2009). https://doi.org/10.1186/1758-2946-1-8

Q1: Predicted Desirability (Synthesizable & Active)

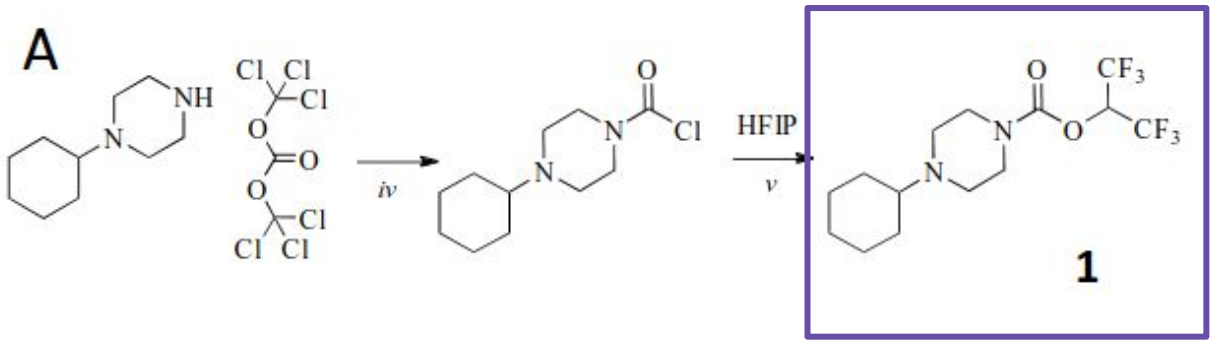


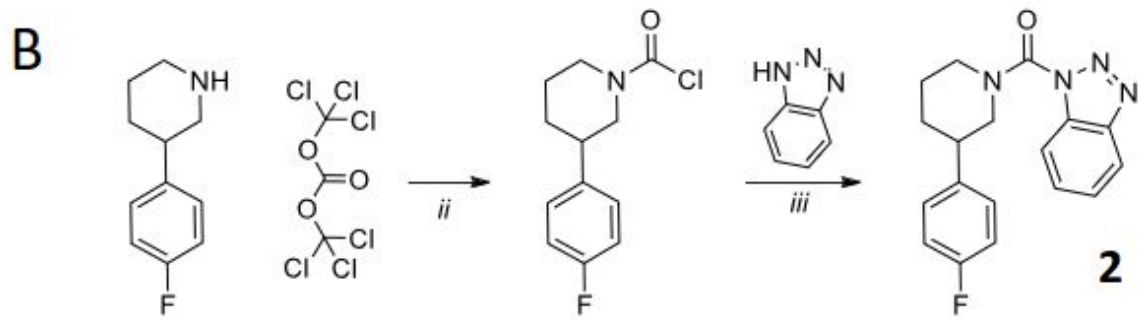
- Synthetic accessibility is important to account for
  - **QSAR baseline without SA nearly zero solved routes**
- **SAScore resulted in poor optimization of the QSAR objective**
- Building block set size does not matter much
  - ZINC and LED3 showed comparable results for all ML-based scores

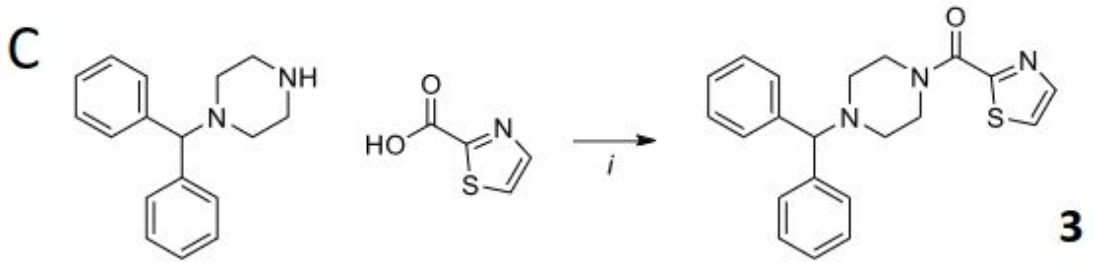
Q2: Predictive Performance on Generated Compounds

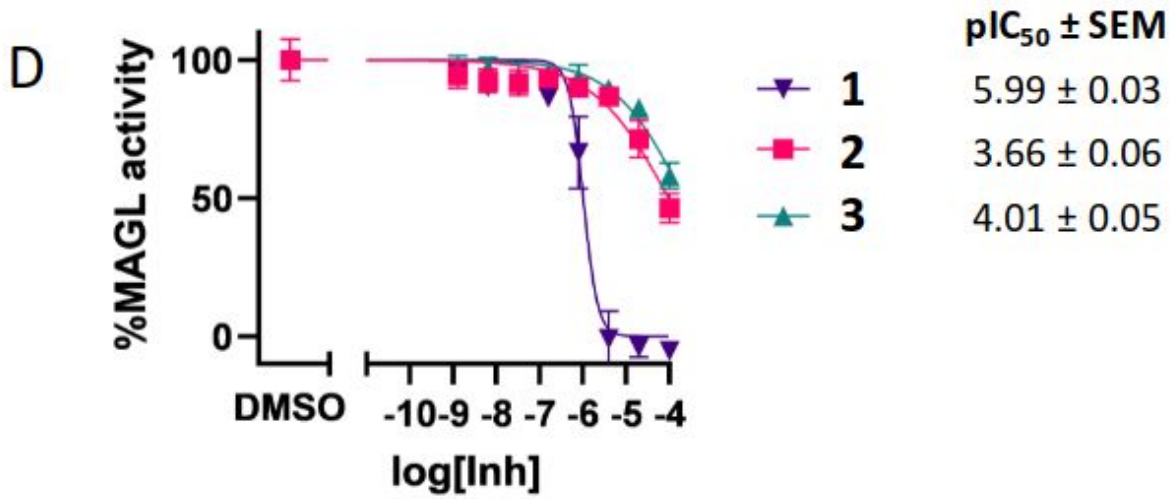


Q3: Experimental Validation









Chemokine Receptors (CCRs) in Cancer

BRAIN TUMOR	
CXCR4	PRX177561 + Bevacizumab + Sunitinib POL5551 + aVEGF AMD3465 USL311 + Lomustine* AMD3100*
ACKR3	X7Ab + Temozolomide

BREAST CANCER	
CCR1	CCX9588 + anti-PD-L1
CCL2	CNTO888 + Radiotherapy
CXCR2	Reparixin + PTX
CXCR4	LY2510924* Balixafortide + Eribulin* USL311 + Lomustine*

LUNG CANCER	
CCR4	AF399/420/1802
CXCR4	AMD3100 + VIC-008 LY2510924*

HEPATOCELLULAR CARCINOMA	
CCR2	747 + Sorafenib RDC018

PANCREATIC TUMOR	
CCR2	PF-04136309 + nab-PTX + GEM* PF-04136309 + GEM* PF-04136309 + FX* CCX872 + anti-PD-1 CCX872 + FX
CXCR2	CXCR2$^{-/-}$ CXCR2$^{-/-}$ + anti-PD-1 SB225002 + RS504393 + FX AZD5069 * Reparixin + PTX*
CXCR4	AMD3100 + anti-PD-L1

COLON and GASTRIC CANCER	
CCR1	BL5923
CCR4	AF399/420/1802
CCR5	Maraviroc + Chemotherapy*
CCR7	siRNA
CXCR2	Reparixin + 5-fluorouracil
CXCR4	LY2510924*

RENAL CARCINOMA	
CCR4	Affi 5

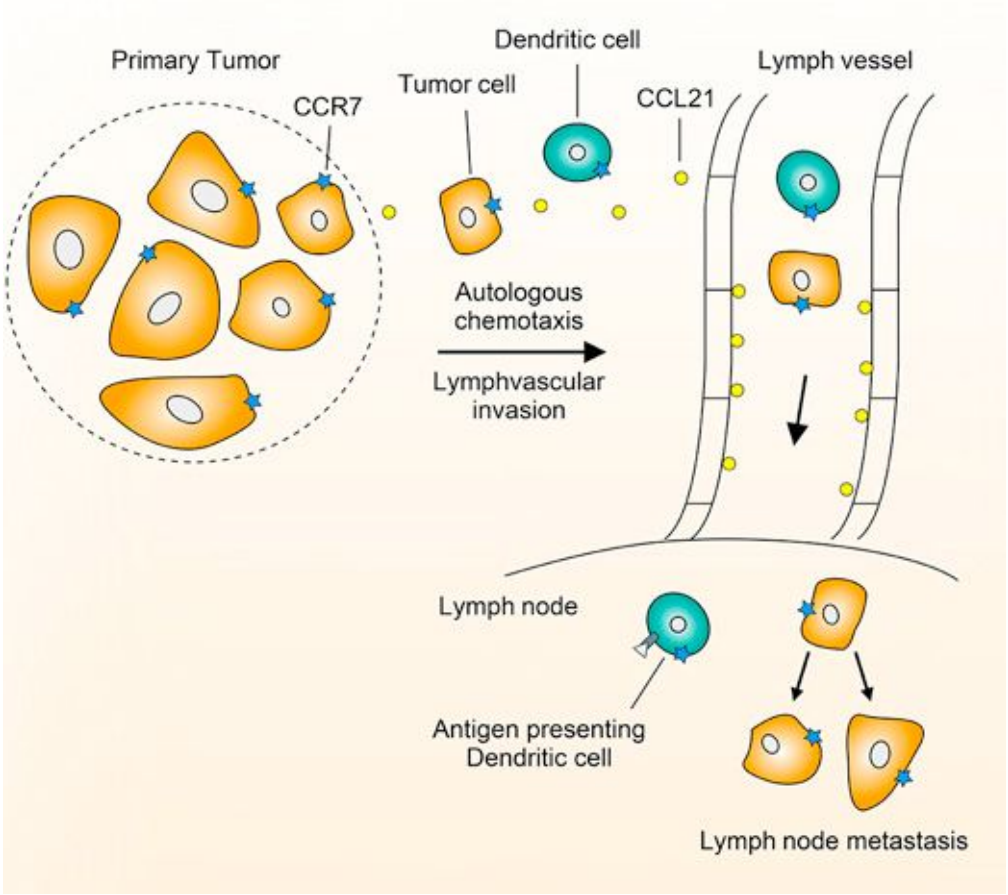
HEMATOLOGIC MALIGNANCIES	
CCR1	CCX721
CCR4	Mogamulizumab* Anti-CCR4 CAR T-cells
CCR7	MSM R707
CXCR4	AMD3100 + Ara-C LY2510924* PF-06747143 BKT140 + Rituximab AMD3100 AMD3100 + anti-PD-1

OVARIAN and PROSTATE CANCER	
CCR2	iCCR2
CCL2	CNTO888*
CCR7	siRNA
CXCR2	SB225002 + Sorafenib SB265610 + Docetaxel
CXCR4	LY2510924* AMD3100

SKIN TUMOR	
CCR4	AF399/420/1802
CXCR2	Navarixin + anti-MEK

Chemokine receptor inhibitors in cancer. Inhibitors in preclinical models and clinical trials.

From: Mollica Poeta V. et al., Front. Immunol. 10:379. doi: 10.3389/fimmu.2019.00379



CCR7 directs cells to organs that express their ligands (CCL21 and CCL19)

From: Jaeger K. et al., Cell, 5:178, doi: 10.1016/j.cell.2019.07.028

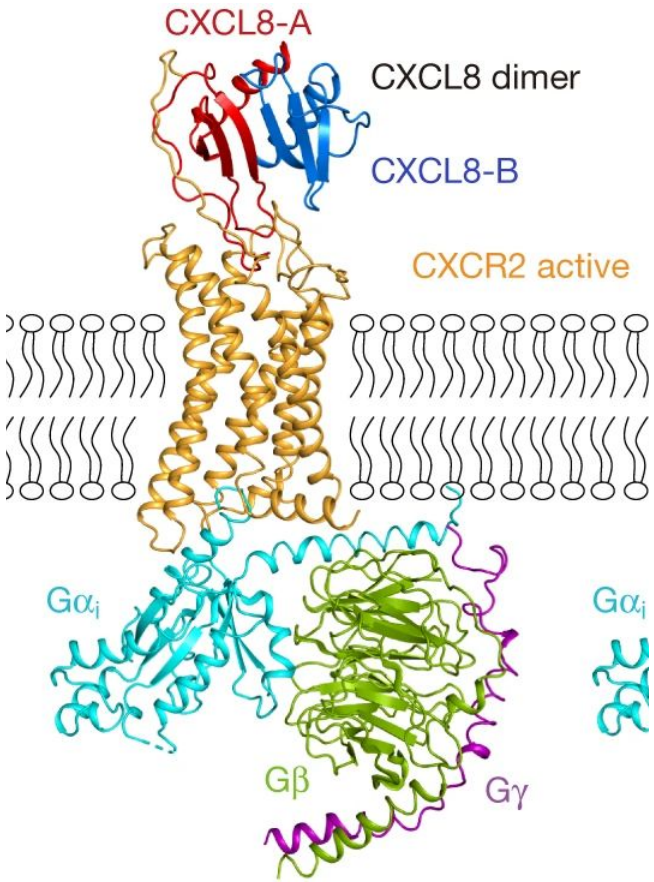
Universiteit Leiden

UNIVERSITY OF CHEMISTRY AND TECHNOLOGY PRAGUE

LACDR

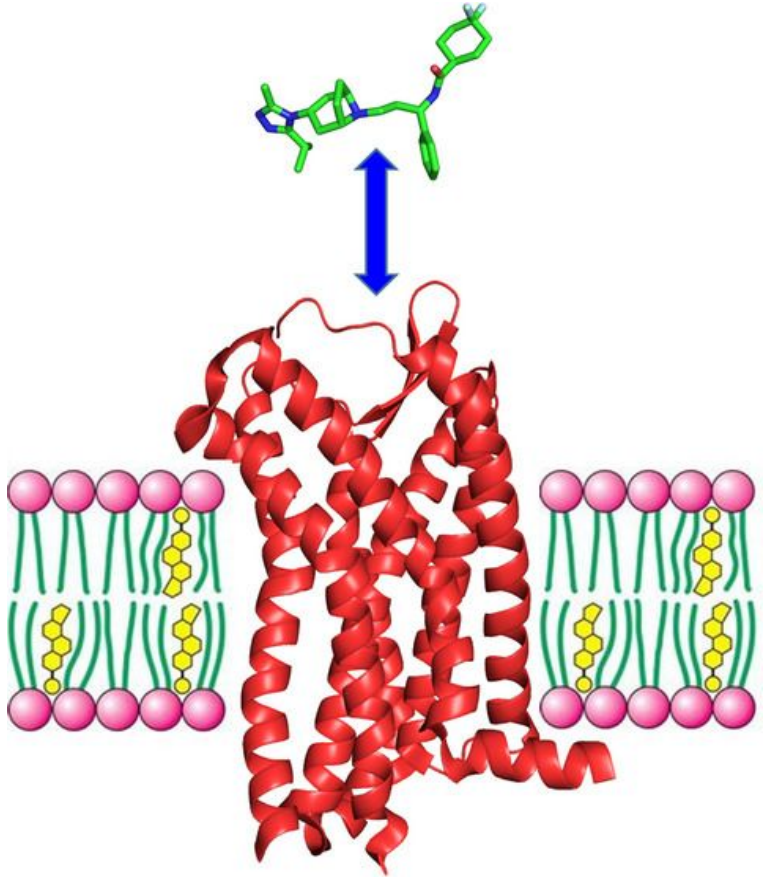
Activation and Deactivation of CCRs

Activation



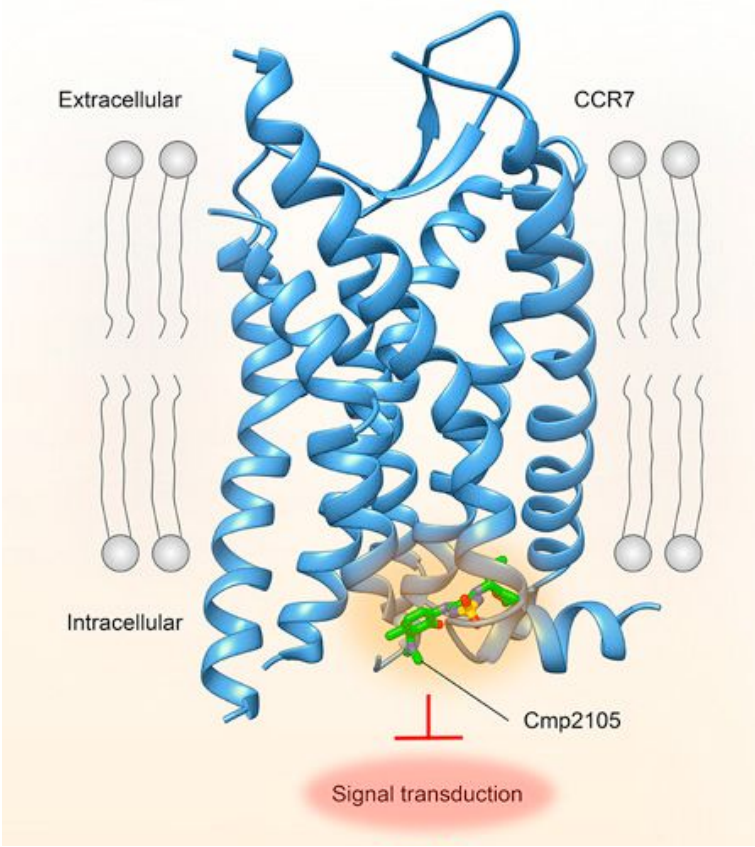
Endogenous activation via chemokine.
From: Liu, K. et al., Nature 585:126-135,
doi: 10.1038/s41586-020-2492-5

Deactivation - Orthosteric/Extracelular Allosteric



Maraviroc, extracellular allosteric antagonist of CCR5.
From: Calmet, P. et al., FEBS J, 287:2367-2385.
doi: 10.1111/febs.15145

Deactivation - Intracellular Allosteric



Cmp2105, intracellular allosteric antagonist of CCR7.
From: Jaeger K. et al., Cell, 5:178, doi: 10.1016/j.cell.2019.07.028

Universiteit Leiden

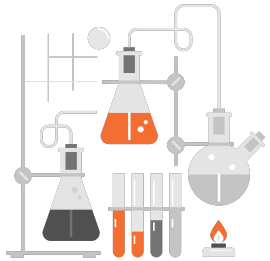
UNIVERSITY OF CHEMISTRY AND TECHNOLOGY PRAGUE

LACDR

Molecular Docking in De Novo Drug Design Towards Intracellular Allosteric Ligands of CCR2

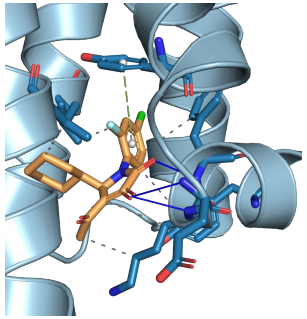
ZINCScore

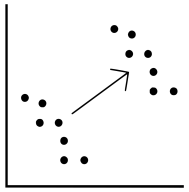
Given a set of building blocks, can we find a synthetic route to the given compound?

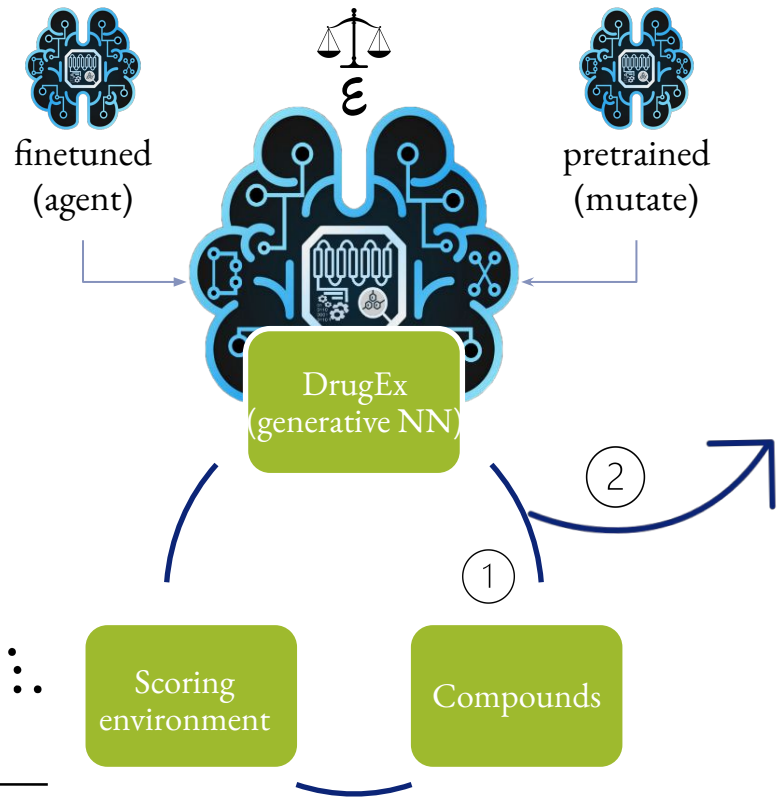


Molecular Docking

Score on interactions the ligand can make

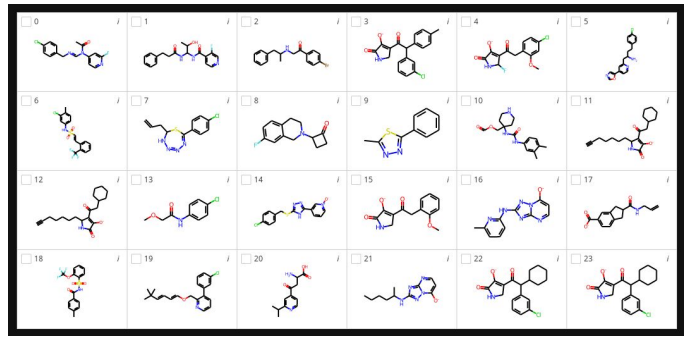




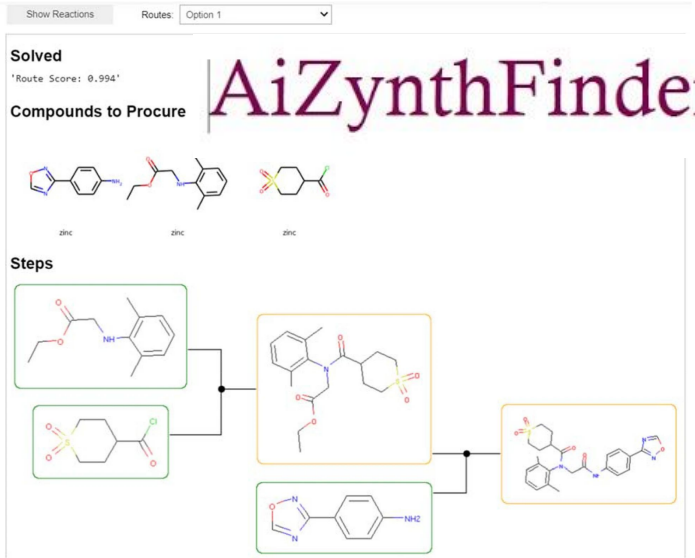


Reinforcement Learning Loop

Sicho et al. *Optimizing Molecular Interactions in De Novo Drug Design: Structure-Based Generation of Intracellular Allosteric Ligands for CCR2 with Transformers, Reinforcement Learning and Docking*, 2025, Unpublished.



3



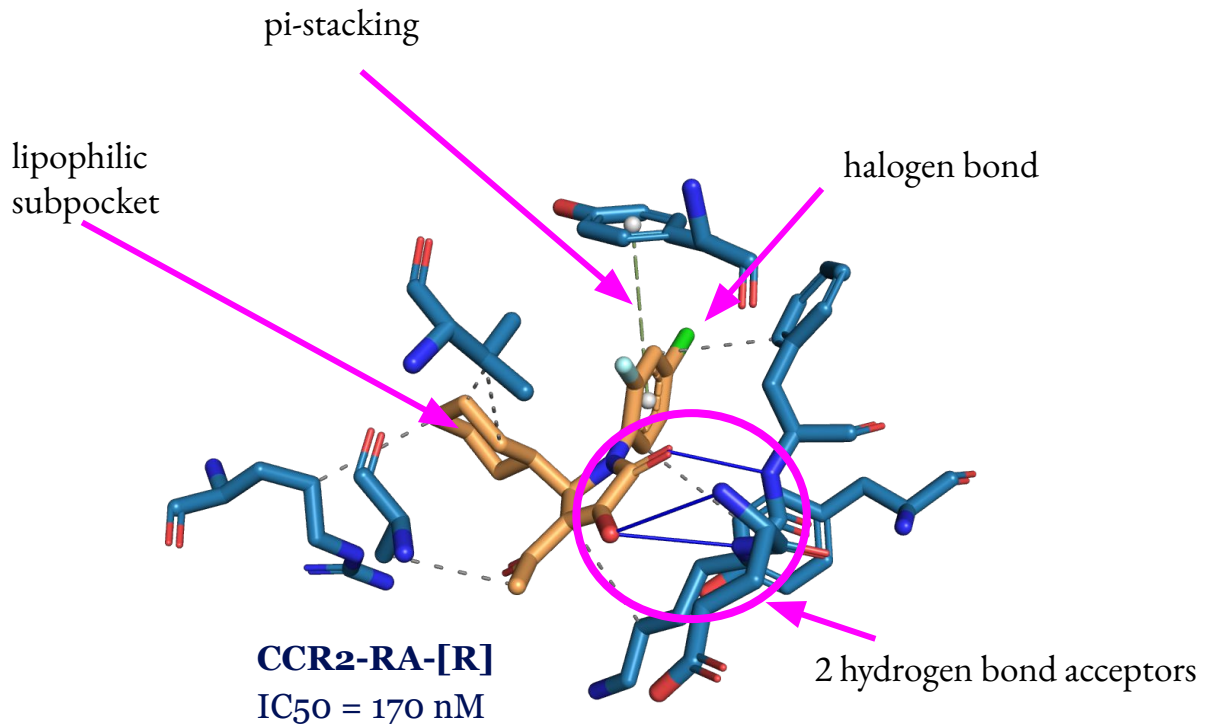
Universiteit Leiden

UNIVERSITY OF CHEMISTRY AND TECHNOLOGY PRAGUE

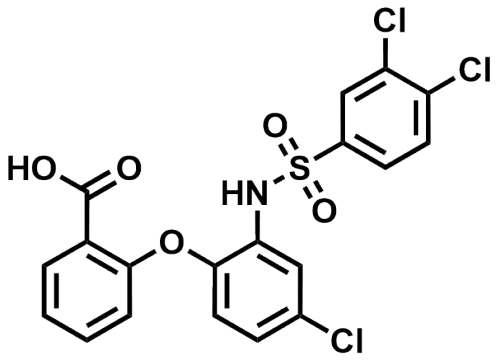
LACDR

The Binding Pocket

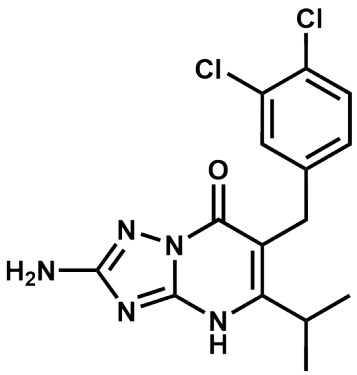
- **One crystal structure with an intracellular allosteric ligand (CCR2-RA-[R]):**
  - https://www.rcsb.org/structure/5T1A
  - downsides:
    - slightly lower resolution (2.81 Å)
    - some residues incomplete
    - mutations of some residues
  - upsides
    - important residues in the binding site are complete and have a meaningful orientation towards the ligand
    - most of the questionable residues are not directly in the binding site
- **Usable for docking after cleanup and some repairs**
  - add incomplete residues
  - reverse mutations close enough to the binding site with a plausible rotamer of the wild type amino acid



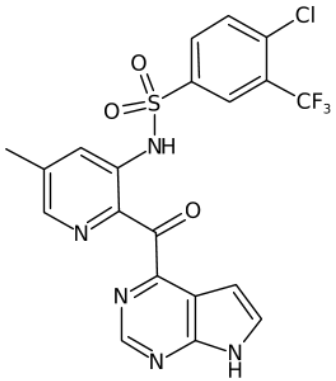
Known Allosteric CCR2 Ligands



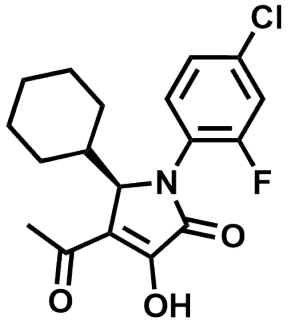
SD-24
Ki = 3.2 nM



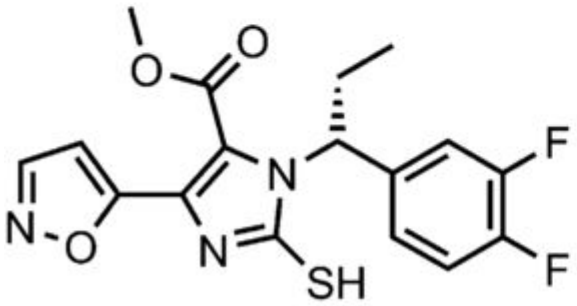
Compound 39
Ki = 1.6 nM



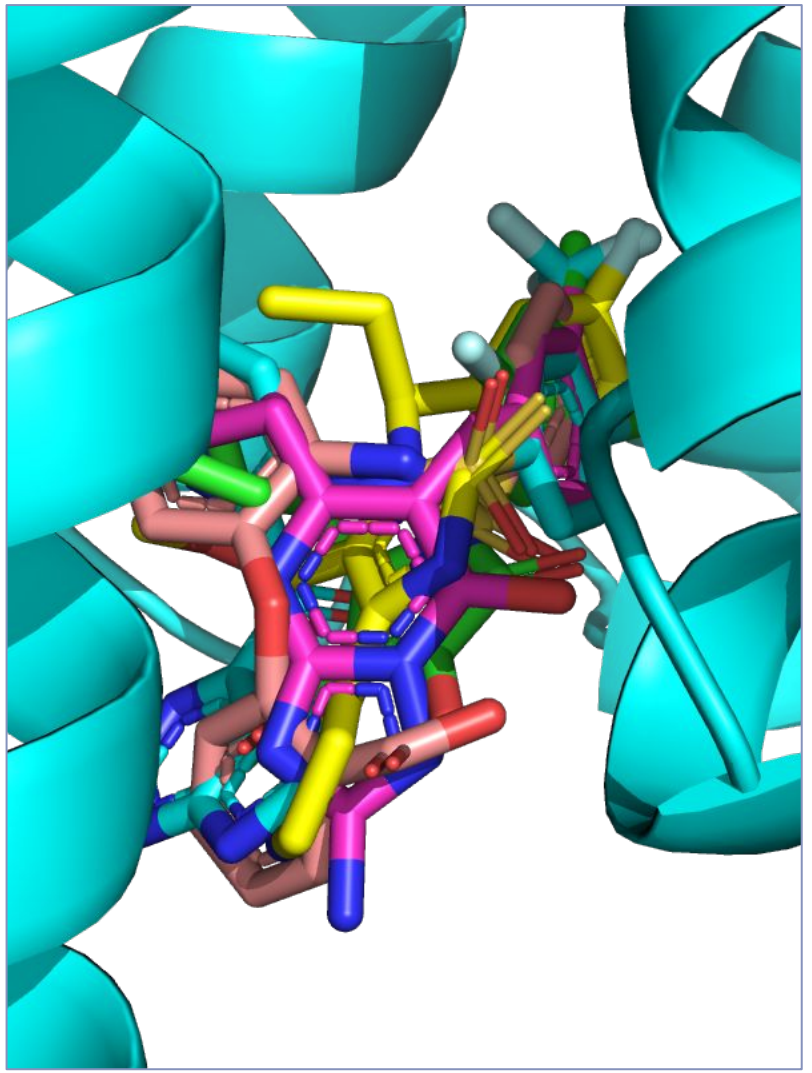
CCX140
IC50 = ???



CCR2-RA-[R]
IC50 = 12 nM



JNJ-27141491
IC50 = 13 nM



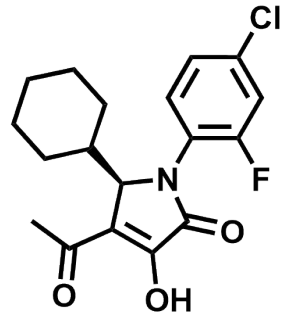
Universiteit Leiden

* data from ChEMBL

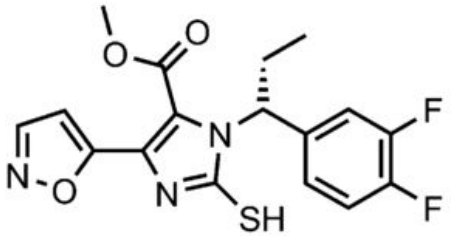
UNIVERSITY OF CHEMISTRY AND TECHNOLOGY PRAGUE

LACDR

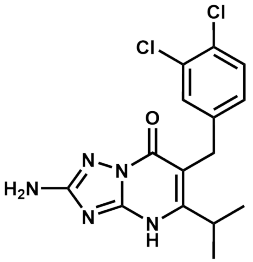
Docking of Known Ligands (AutoDock Vina)



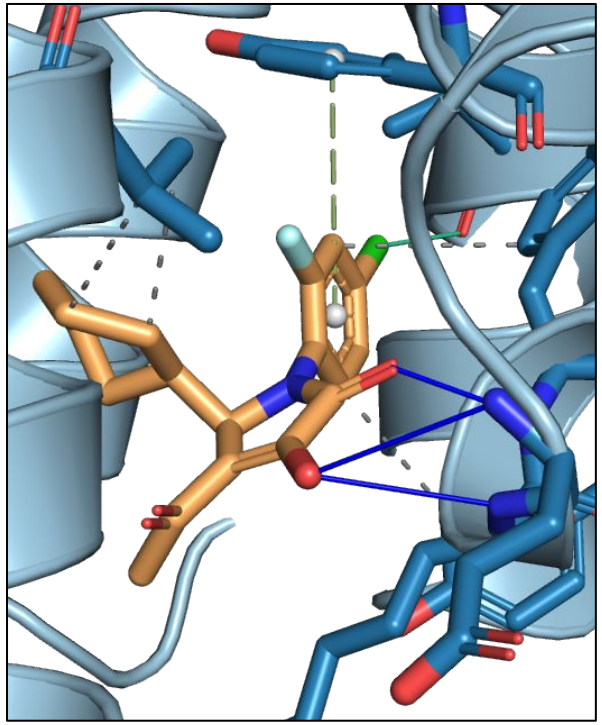
CCR2-RA-[R]

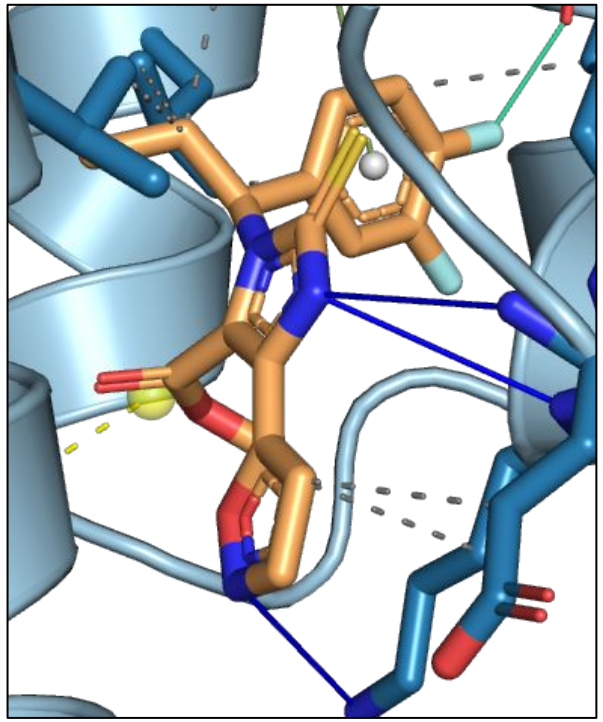


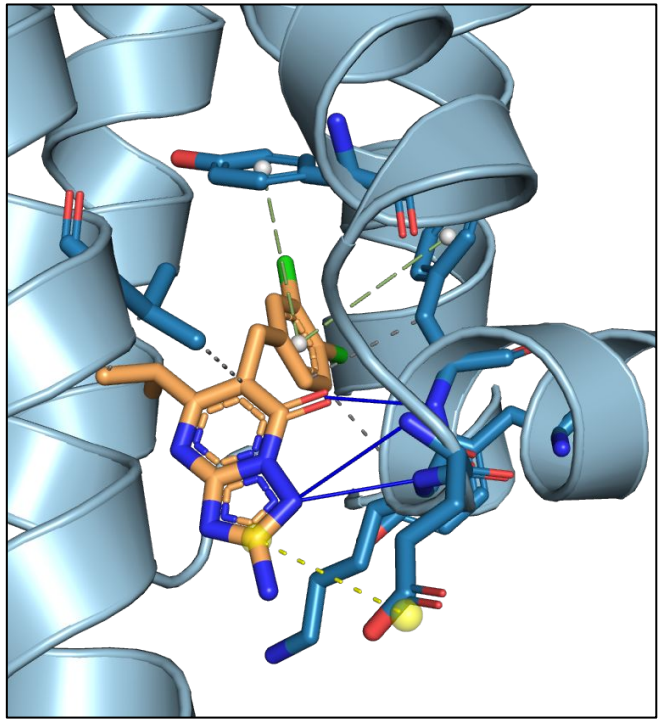
JNJ-27141491



Compound 39

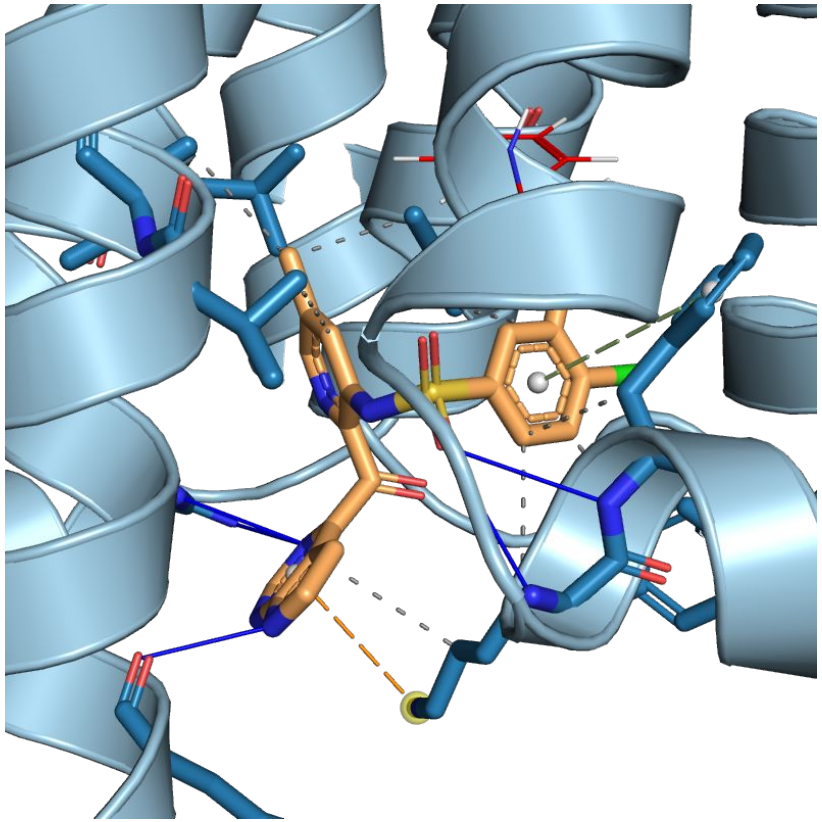


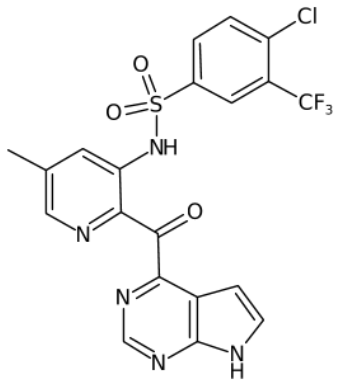




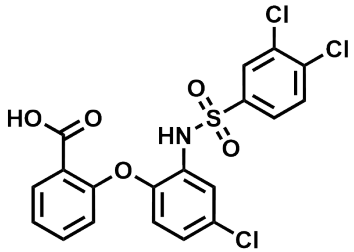
Interactions: https://github.com/pharmai/plip

Docking of Known Ligands (AutoDock Vina)

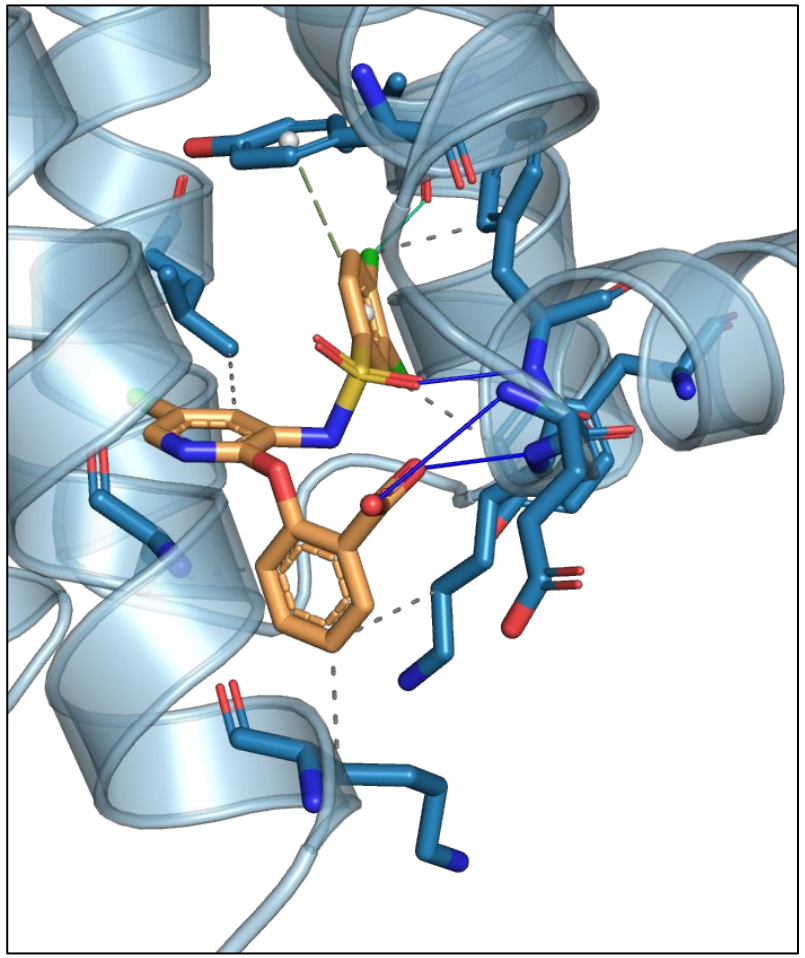




CCX140

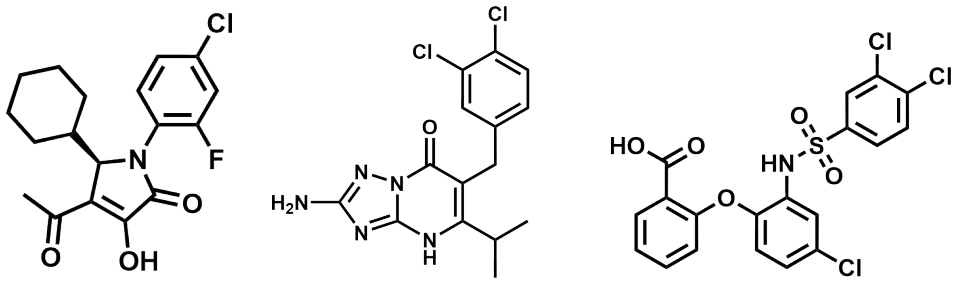


SD-24



Interactions: https://github.com/pharmai/plip

Scoring the Interactions



- Determined by **5 most active compounds from each scaffold group**
- Interactions manually divided into groups:
  - **Required**
    - Occur in all scaffold groups
  - **Essential**
    - Believed to be important for increased activity
    - pi-stacking interactions with key residues
  - **Important**
    - Known parts of the binding site that many of the high affinity/potency ligands exploit, but each different way
    - especially lipohilic interactions
  - **Interesting/New**
    - interactions that the top ligands have, but not all of them + potentially interesting residues to hit

_hbondd_LYS_311_A	15.0
_hydroph_LEU_81_A	13.0
_hbondd_PHE_312_A	13.0
_hbondd_GLU_310_A	12.0
_halogenbond_VAL_63_A	12.0
_hydroph_PHE_312_A	12.0
_hydroph_VAL_244_A	11.0
_hydroph_TYR_315_A	11.0
_pistack_TYR_305_A	11.0
_hydroph_LYS_311_A	9.0
_hydroph_ALA_241_A	7.0
_hydroph_LYS_237_A	5.0
_hydroph_TYR_305_A	4.0
_pistack_PHE_312_A	4.0
_hbonda_LYS_237_A	4.0
_hydroph_LEU_67_A	4.0
_hydroph_ARG_138_A	3.0
_hbondd_ARG_138_A	3.0
_hydroph_VAL_63_A	2.0
_saltbridge_GLU_310_A	2.0

SCORE = $$W = \frac{\sum_{i=1}^{n} w_i X_i}{\sum_{i=1}^{n} w_i}$$

all active ligands (pchembl >= 6.5)

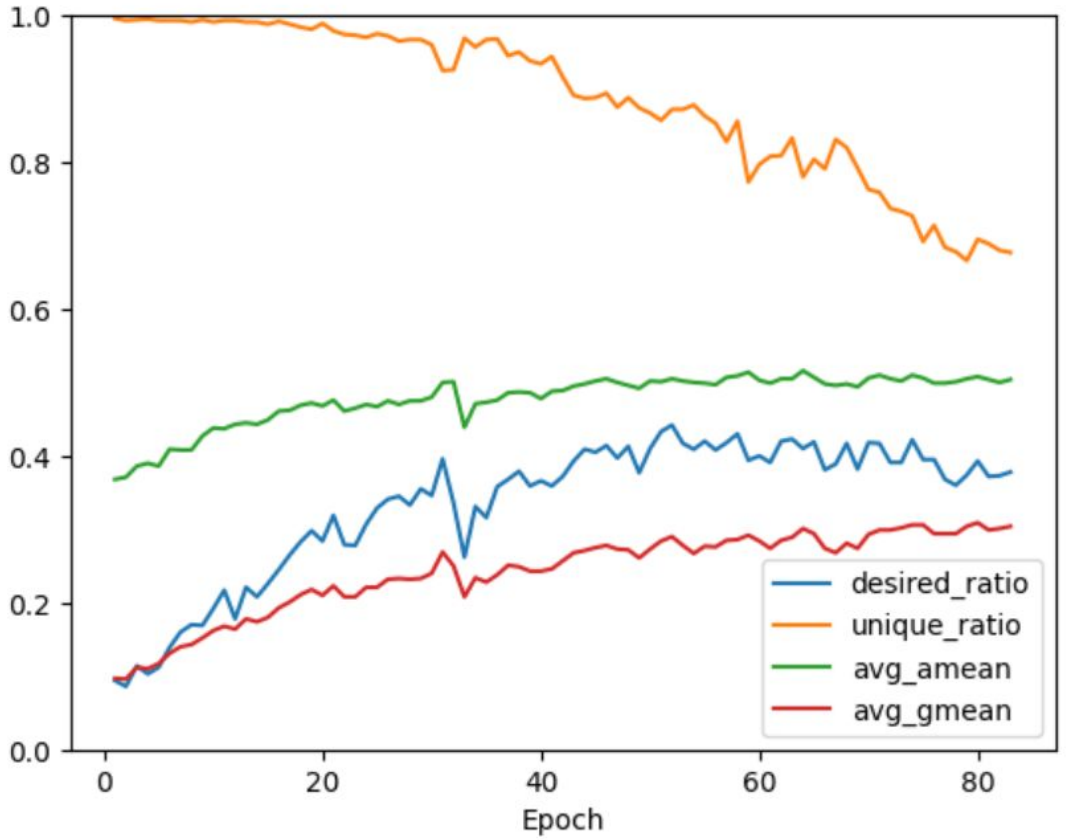
_hbondd_LYS_311_A	179.0
_hydroph_PHE_312_A	156.0
_hydroph_LEU_81_A	145.0
_hbondd_PHE_312_A	142.0
_pistack_TYR_305_A	142.0
_hydroph_TYR_315_A	136.0
_hydroph_VAL_244_A	125.0
_hydroph_LYS_311_A	123.0
_hydroph_ALA_241_A	94.0
_hbondd_GLU_310_A	93.0
_hbondd_ARG_138_A	83.0
_hydroph_LEU_67_A	75.0
_pistack_PHE_312_A	61.0
_halogenbond_VAL_63_A	59.0
_hydroph_VAL_63_A	55.0
_hydroph_LYS_237_A	54.0
_hydroph_THR_77_A	53.0
_hydroph_TYR_305_A	48.0
_hydroph_ARG_138_A	36.0
_hydroph_LEU_134_A	27.0
_hbonda_LYS_237_A	26.0
_hbondd_ARG_238_A	26.0
_pication_LYS_311_A	23.0
_hydroph_ILE_245_A	22.0
_pication_ARG_138_A	20.0
_saltbridge_GLU_310_A	19.0

Reinforcement Learning

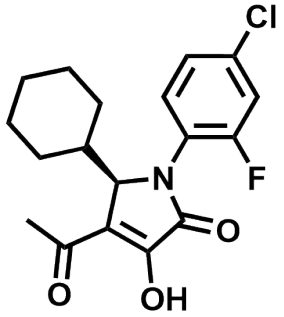
Generative Workflow

- generate 10,000 molecules and score them
- keep ligands with:
  - **ZINCScorer > 0.75**
  - Required and Essential IFPScore component = 1 (**hydrogen bonds with conserved residues + pi-stacking**)
  - => **96 structures**

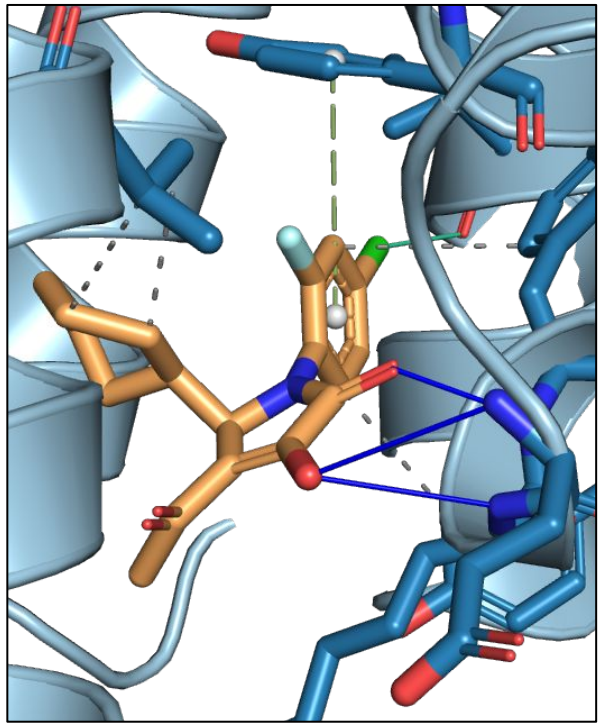
=> after manual prioritization: synthesis of 4 distinct scaffolds (2 easy, 2 hard)

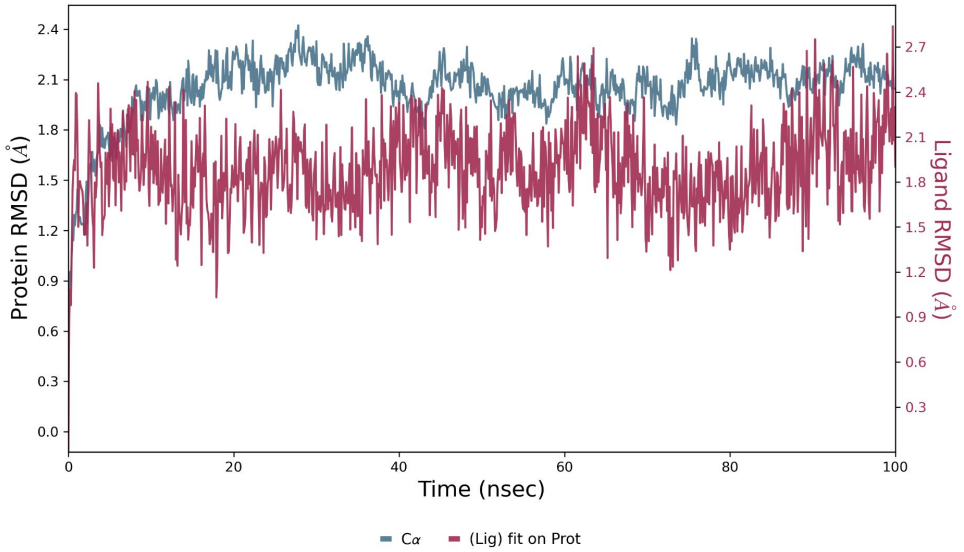


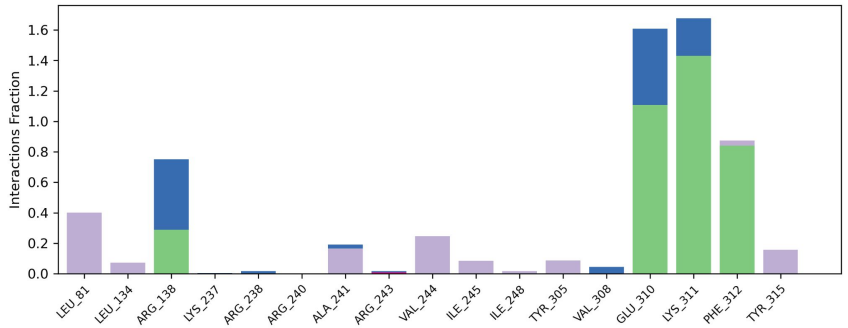
Molecular Dynamics with CCR2-RA-[R]

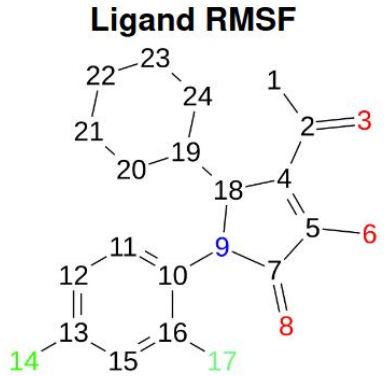


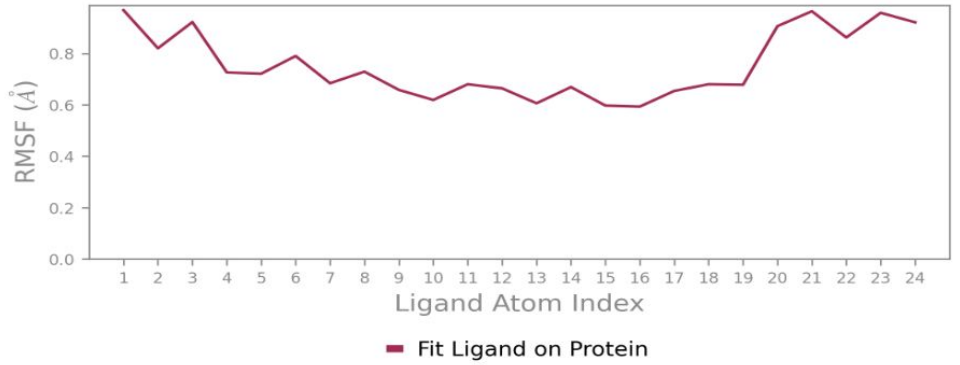
CCR2-RA-[R]





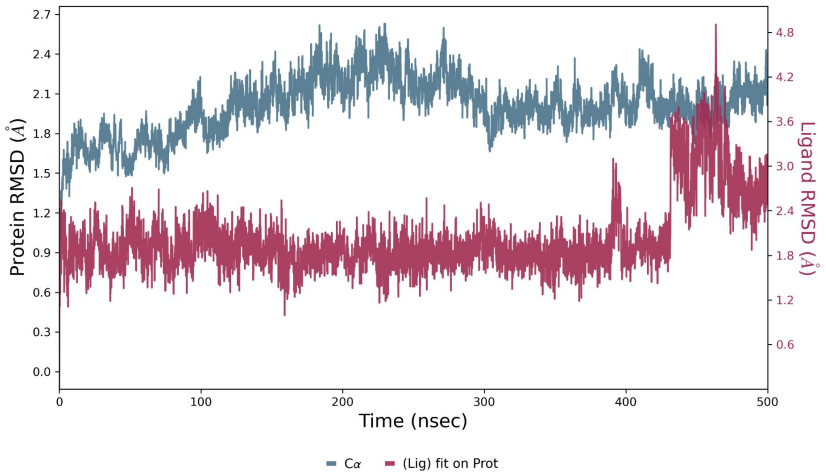


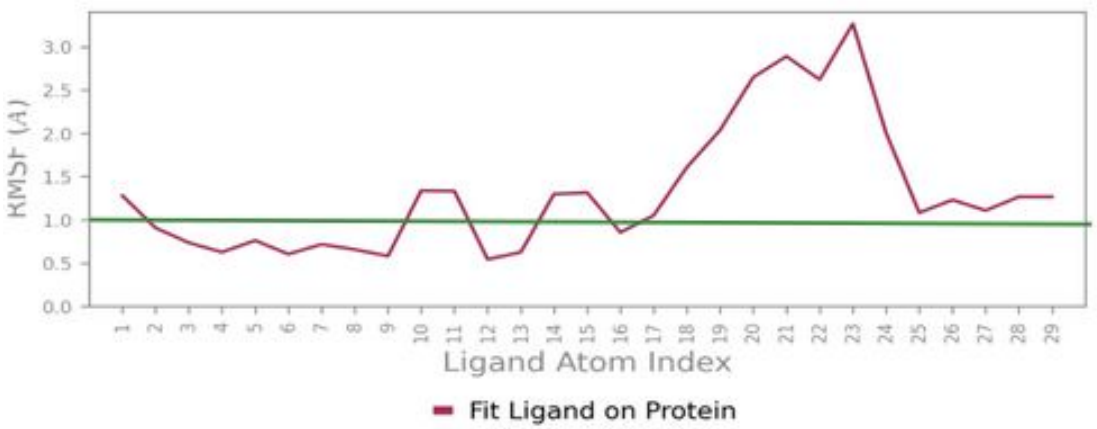


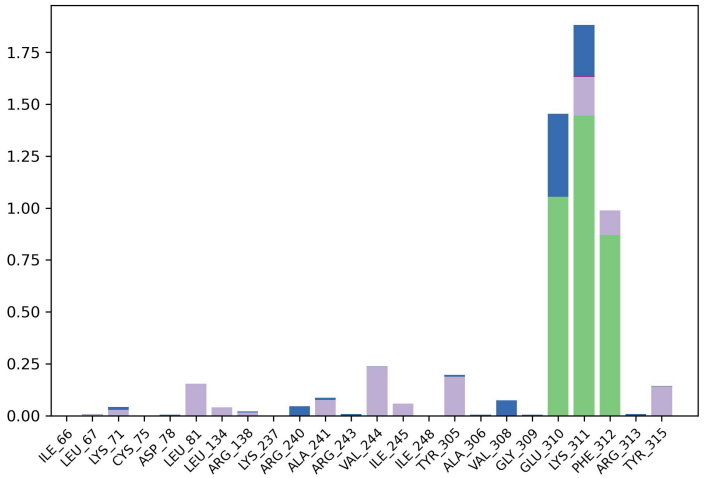


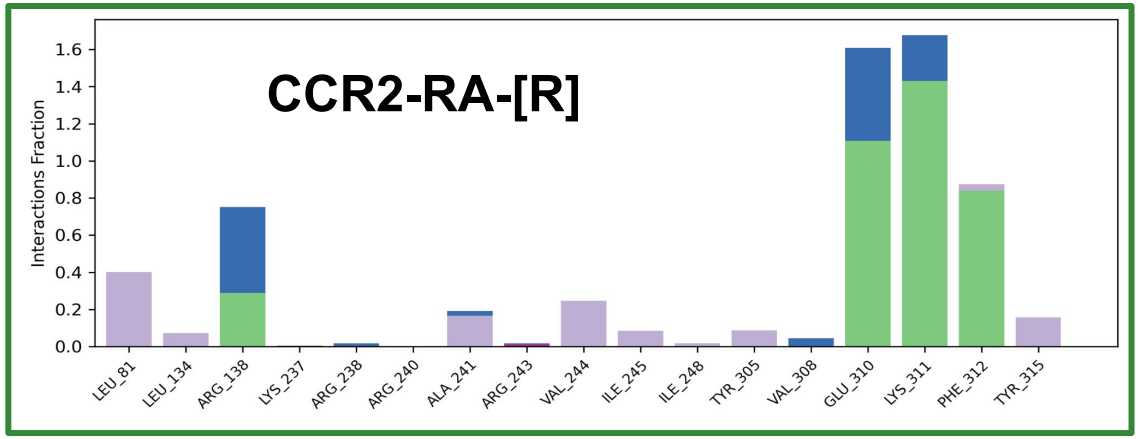
H-bonds Hydrophobic Ionic Water bridges

Candidate v4_007_a

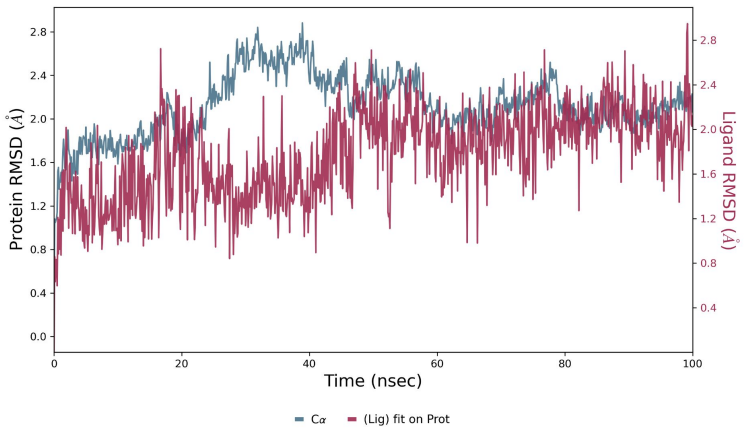


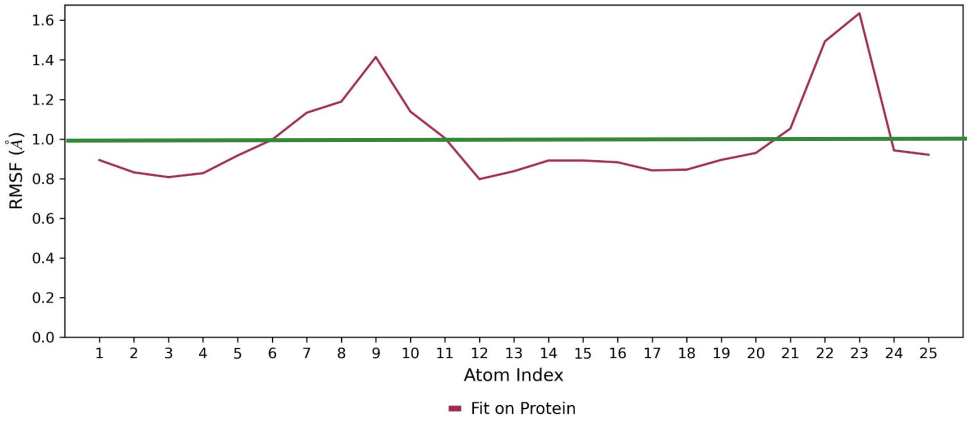


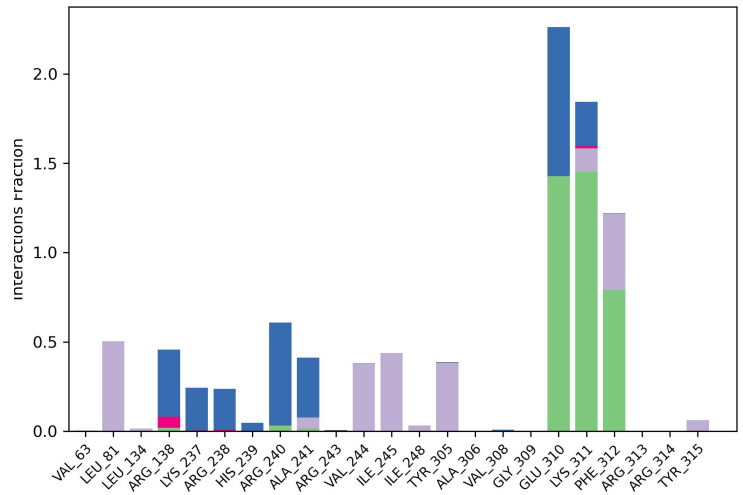


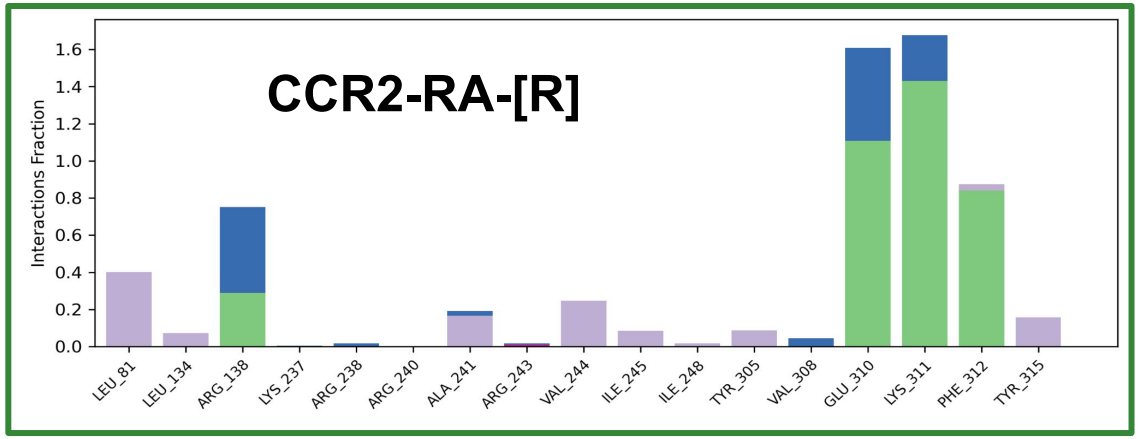


Candidate v4_100_d5_i









Hydrophobic Ionic Water bridges

CCR2-RA-[R] Displacement Assay

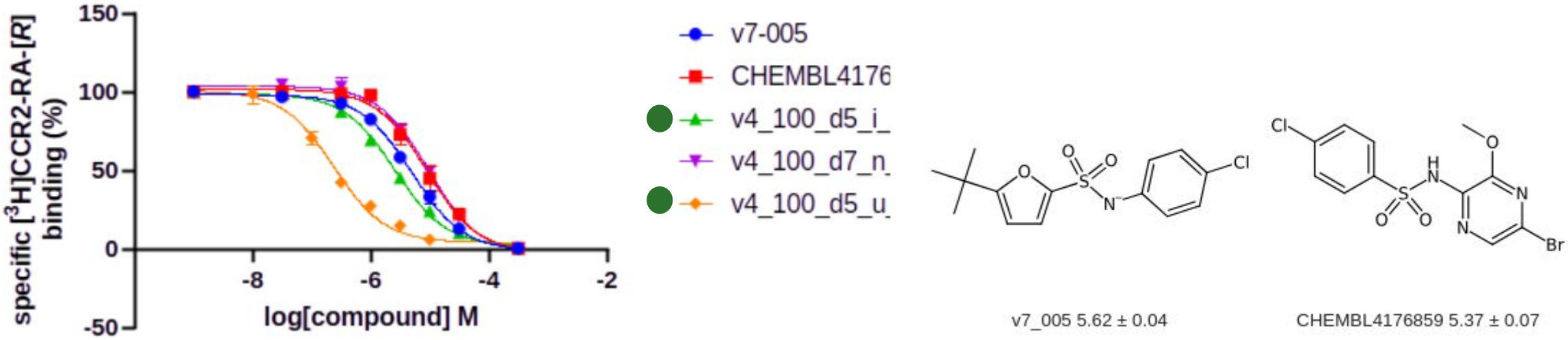
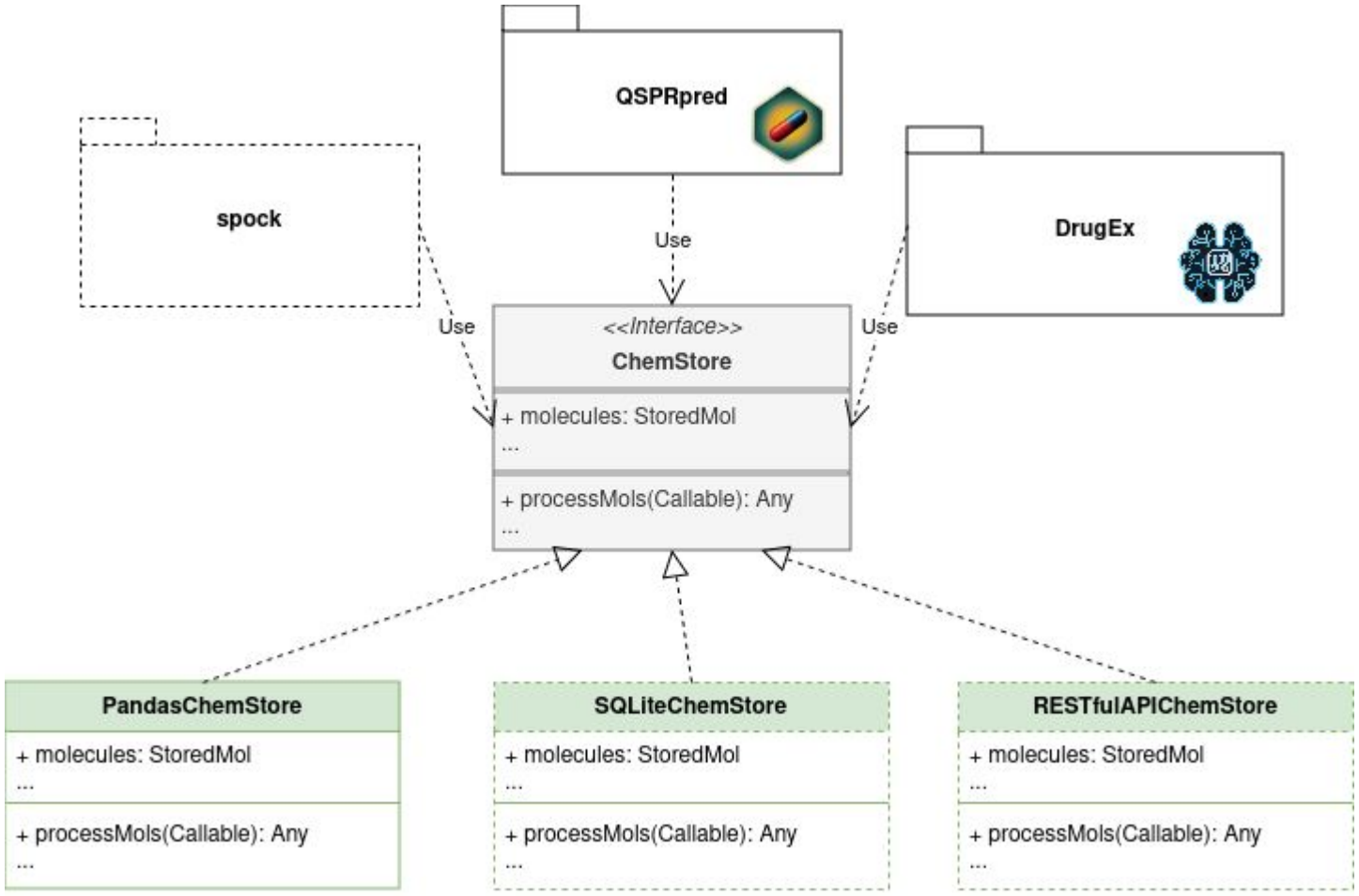


Figure : Data represent mean ± SEM of three independent experiments performed in duplicate. Total binding (TB) is set to 100% and nonspecific binding (NSB) to 0%. pKi ± SEM was determined from three independent experiments performed in duplicate or % displacement at 10 µM ± SD of two independent experiments performed in duplicate. The compounds were tested at 10 µM on 20 µg U2OS-CCR2 membranes and [3H]CCR2-RA-[R] with a concentration of ~6.5 nM. If [3H]CCR2-RA-[R] displacement is ≥50%, the compounds were tested with ranging concentrations. For pKi calculations a kD value of 6.3 nM was used.

v4_100_d5_u (Ki ~ 100 nM)
v4_100_d5_i (Ki ~ 1 µM)

Software Development Perspective

https://github.com/CDDLeiden/QSPRpred/tree/dev



- **ChemStore**
  - efficient processing
    - multi-CPU
    - Dask
    - ...
  - molecule representation hierarchy
    - standardization
    - unique identification
    - conformers
    - tautomers
    - ...
  - multiple implementations
    - Pandas
    - SQL Databases
    - ....

Conclusions

- DNDD is a large and historically rich field
  - It has seen a **significant boost in the last years from generative DL models**
- **Challenges:**
  - **Synthetic accessibility** and overall stability of generated structures
    - Plausible pose generation for in pocket generators
  - Validation and benchmarking
    - **Prospective validation with follow up wet lab experiments paramount**
  - Multi-modal models for zero-shot predictions

Acknowledgments

LACDR Colleagues

- **Supervisors & Consultants:**
  - **Gerard van Westen**
  - **Willem Jespers**
  - Anthe Janssen
  - Laura Heitman
- **Postdocs**
  - Sohvi Luukkonen

LACDR Students

- **PhD Students**
  - **Helle van den Maagdenberg**
  - Olivier Béquignon
  - **Alan Kai Hassen**
  - Andrius Bernatavicius
  - Yorick van Aalst
  - Remco van den Broek
- **Master Students**
  - Chara Spyropoulou
  - Sem Egbers

GitHub

VŠCHT DICH

- **Daniel Svozil**
- **Wim Dehaen**
- **Valeria Fil**
- **Petr Palivec**
- **Jozef Fulop**
- **Asia Ceklarz**

GAČR
GRANTOVÁ AGENTURA ČESKÉ REPUBLIKY

CZ OPENSCREEN

Universiteit Leiden

UNIVERSITY OF CHEMISTRY AND TECHNOLOGY PRAGUE

LACDR

Thank you.

Universiteit Leiden

UNIVERSITY OF CHEMISTRY AND TECHNOLOGY PRAGUE

LACDR

Bij ons leer je de wereld kennen