The application of AI-driven Drug Discovery technology for molecular optimization of nuclear receptor ligands

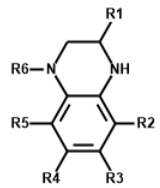
Rafał A. Bachorz

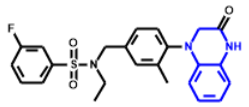
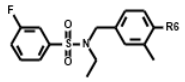
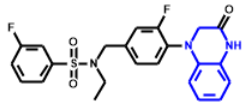
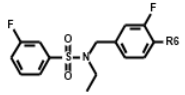
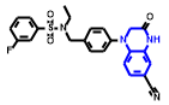
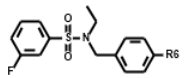
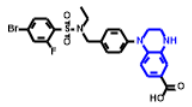
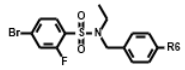
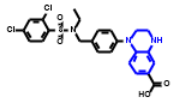
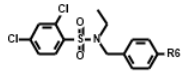
Agenda

- ADMET Predictor® as a *de novo* drug design environment
  - Available capabilities
  - Properties
- AIDD: Artificial Intelligence Drug Design
  - Key principles
  - Chemically intelligent SMIRKS transformation
- Practical use case: nuclear receptors agonists
  - Proposed workflow
  - QSAR model
  - Results
- Summary

ADMET Predictor®

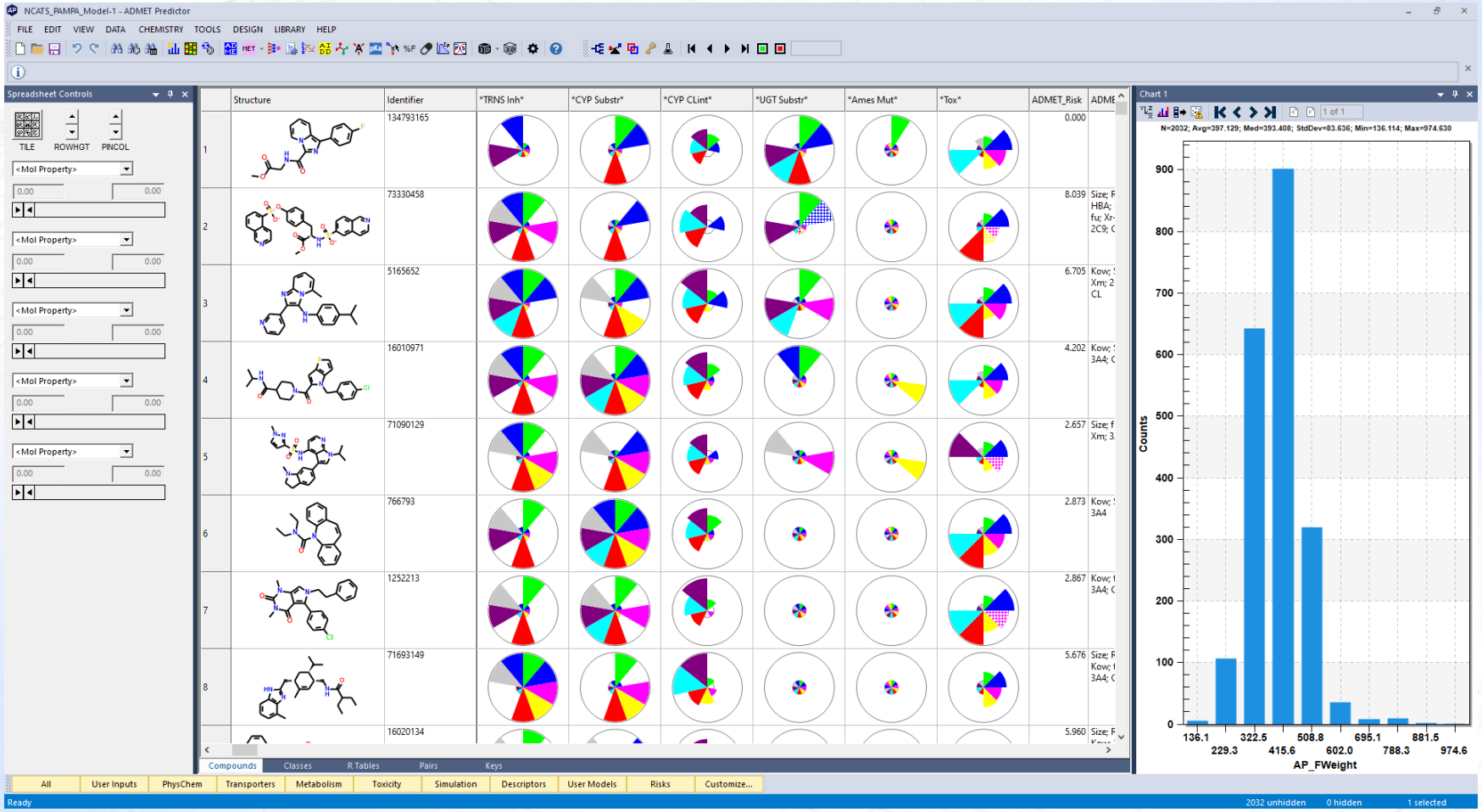
Scaffold



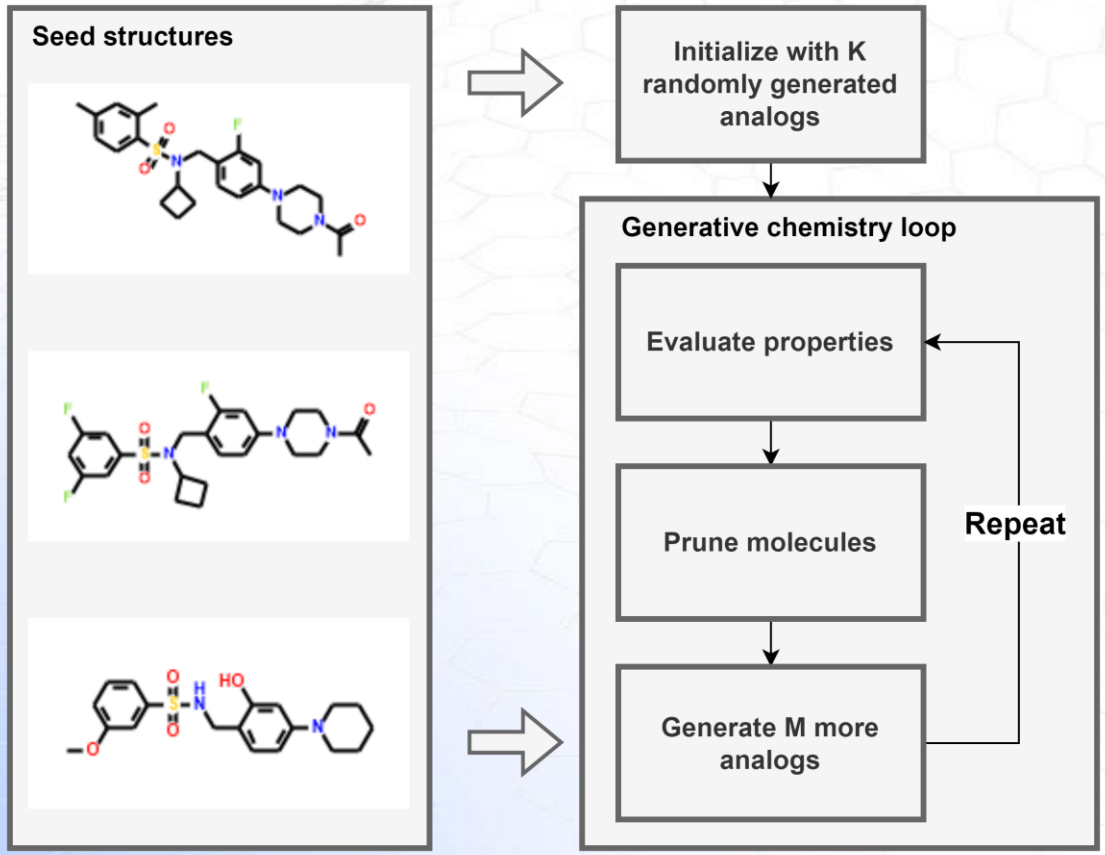
	Structure	Identifier	R1 (2)	R2 (3)	R3 (4)	R4 (4)	R5 (3)	R6 (30)
26		8848	=O	H	H	H	H	
27		7895	=O	H	H	H	H	
28		10093	=O	H	C#N	H	H	
29		7215	H	H		H	H	
30		6594	H	H		H	H	

- **Property prediction software** (QSAR/QSPR)
  - Predicts over 140 ADMET properties from chemical structure
  - Identifies ADMET liabilities in the form of numeric risk scores
- The **HTPK Simulation Module** lets you predict fraction absorbed and bioavailable, as well as other PK parameters, using a virtual human, mouse or rat simulation
- **ADMET Modeler Module** lets you build your own models using our advanced molecular and atomic descriptors
- **MedChem Studio Module** lets you prioritize lead series, discover SAR trends, and design novel compounds
- **AIDD Module** lets you generate new molecules in the frame of multicriteria optimization

ADMET Predictor®

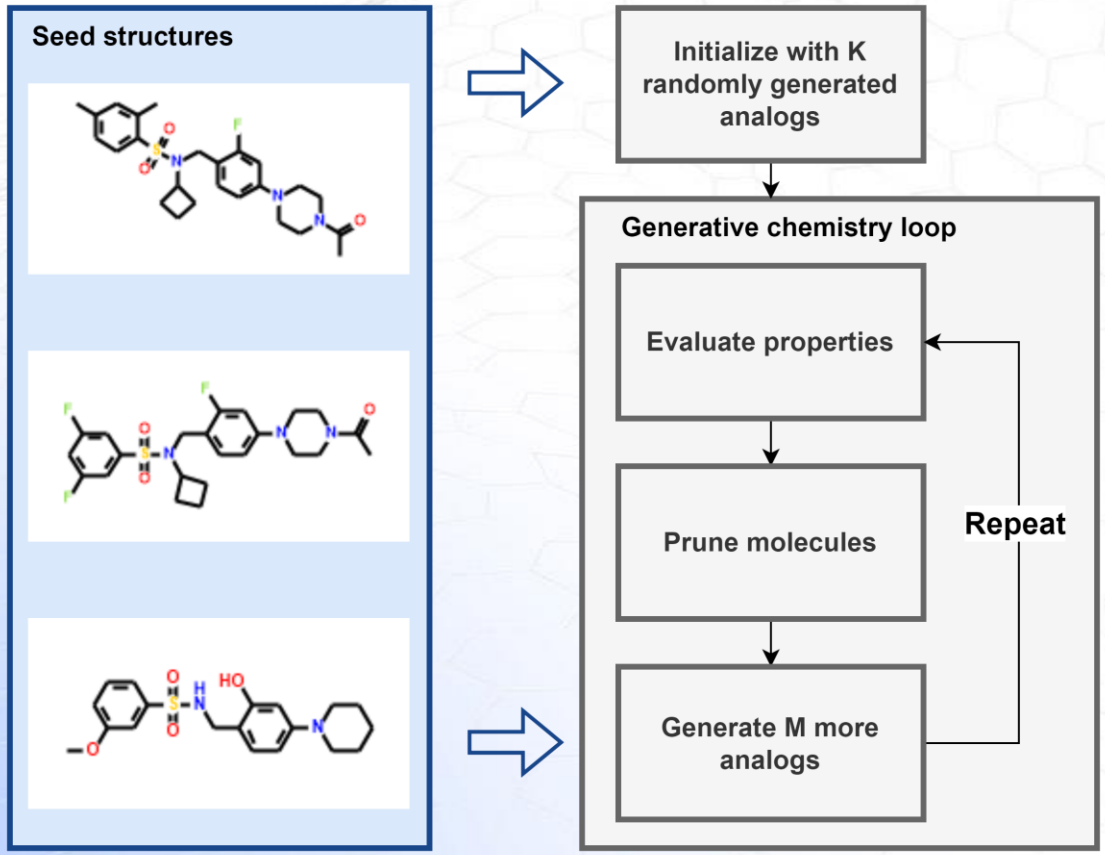


AIDD procedure



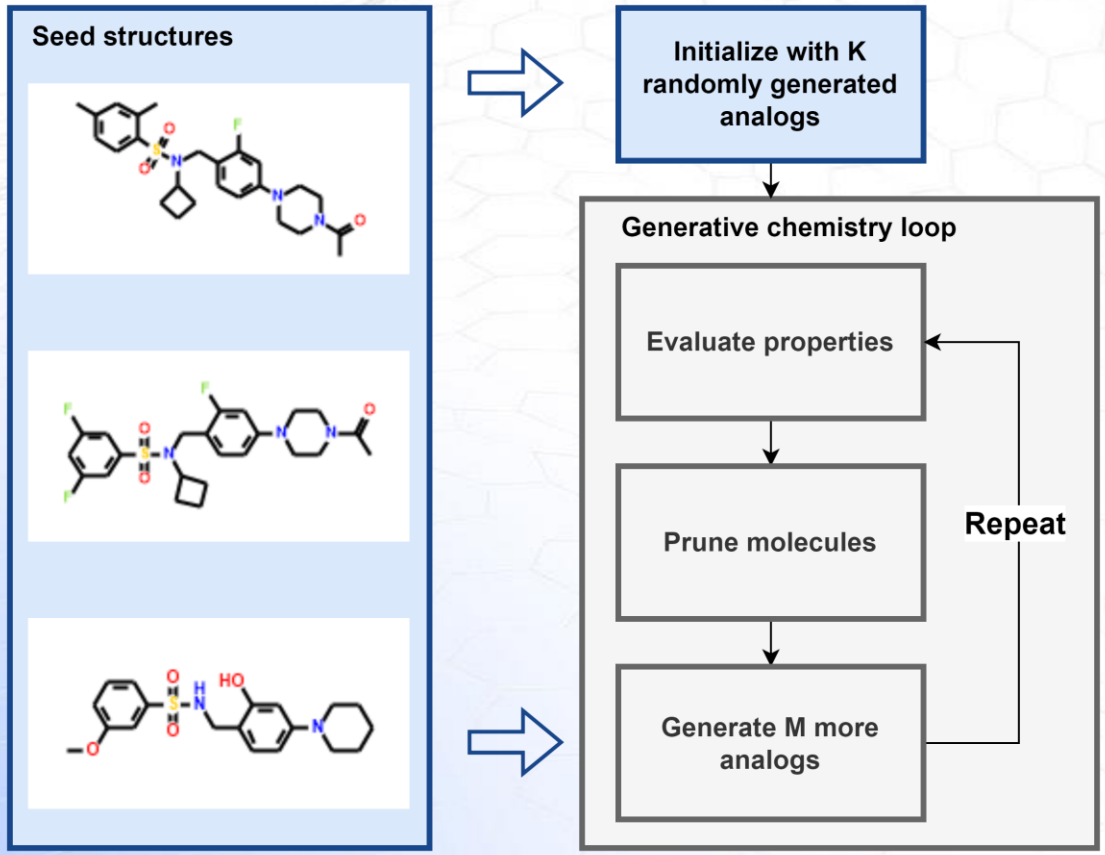
- Logical flowchart of AIDD algorithm
- Seed structures
- Population initialization
- Property evaluation
- Population pruning
- Generation of new compounds

AIDD procedure



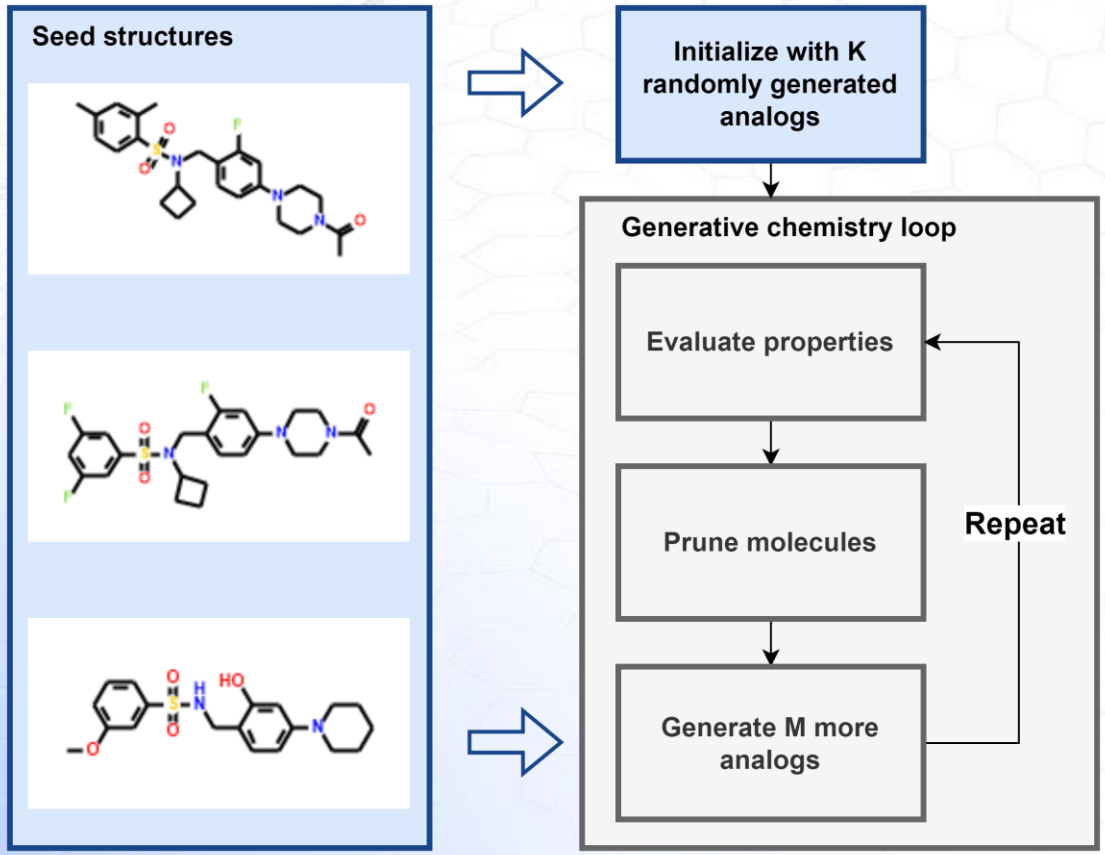
- **Seed structures**
- Can be chosen in an arbitrary manner
- Usual choice is to take some well-known ligands
- Some modifications can be introduced, e.g. bioisosteric replacements
- It is also possible to restrict the generative chemistry with scaffold definition

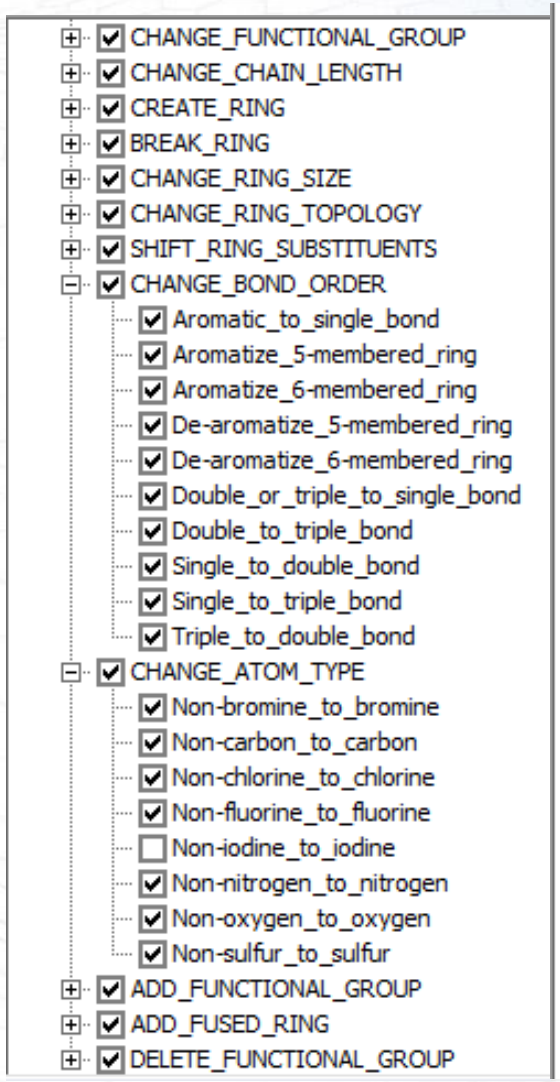
AIDD procedure

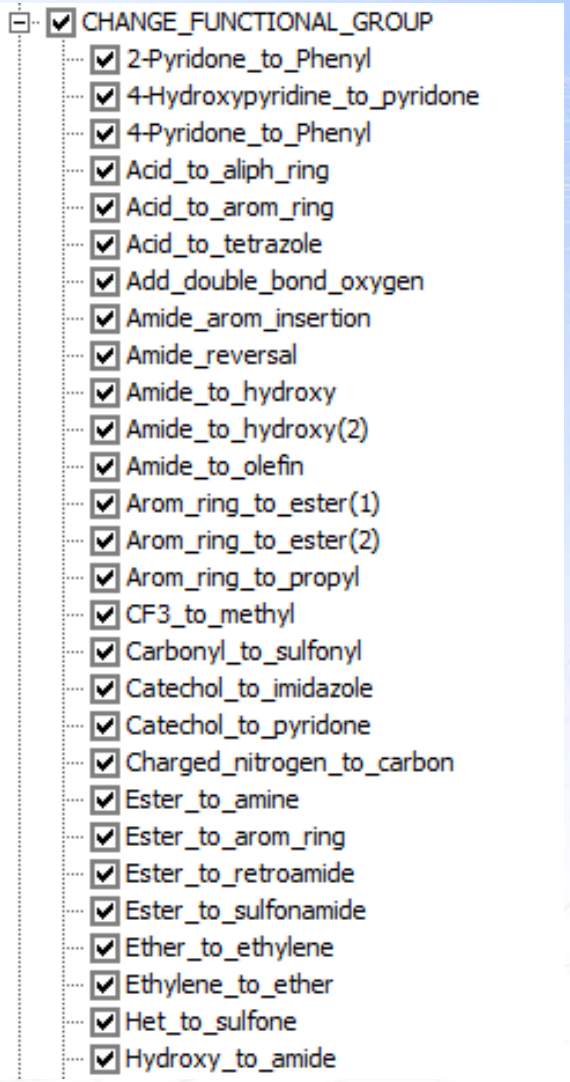


- Population initialization
- SMIRKS-based transformation
- Currently ca. 150 transforms available
- The pre-created population is the subject of further extension in the generative chemistry loop
- Efficient strategy of chemical space exploration

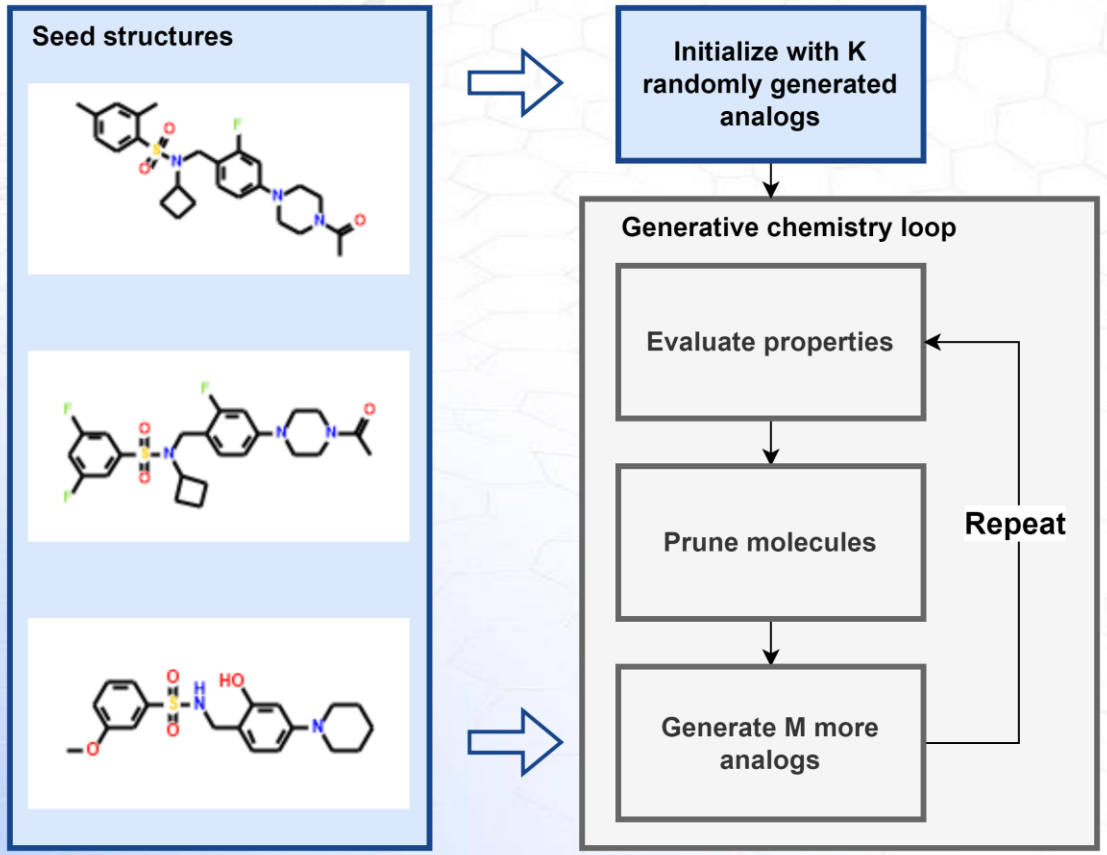
AIDD procedure





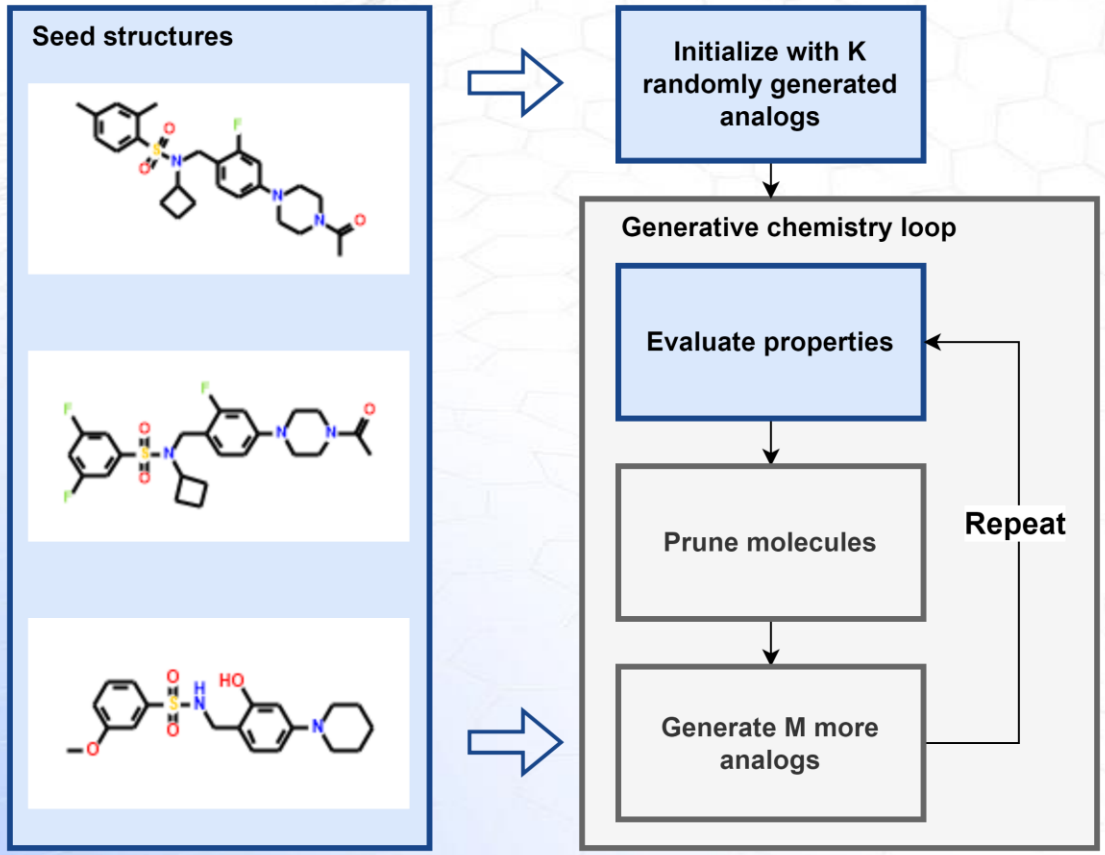


AIDD procedure



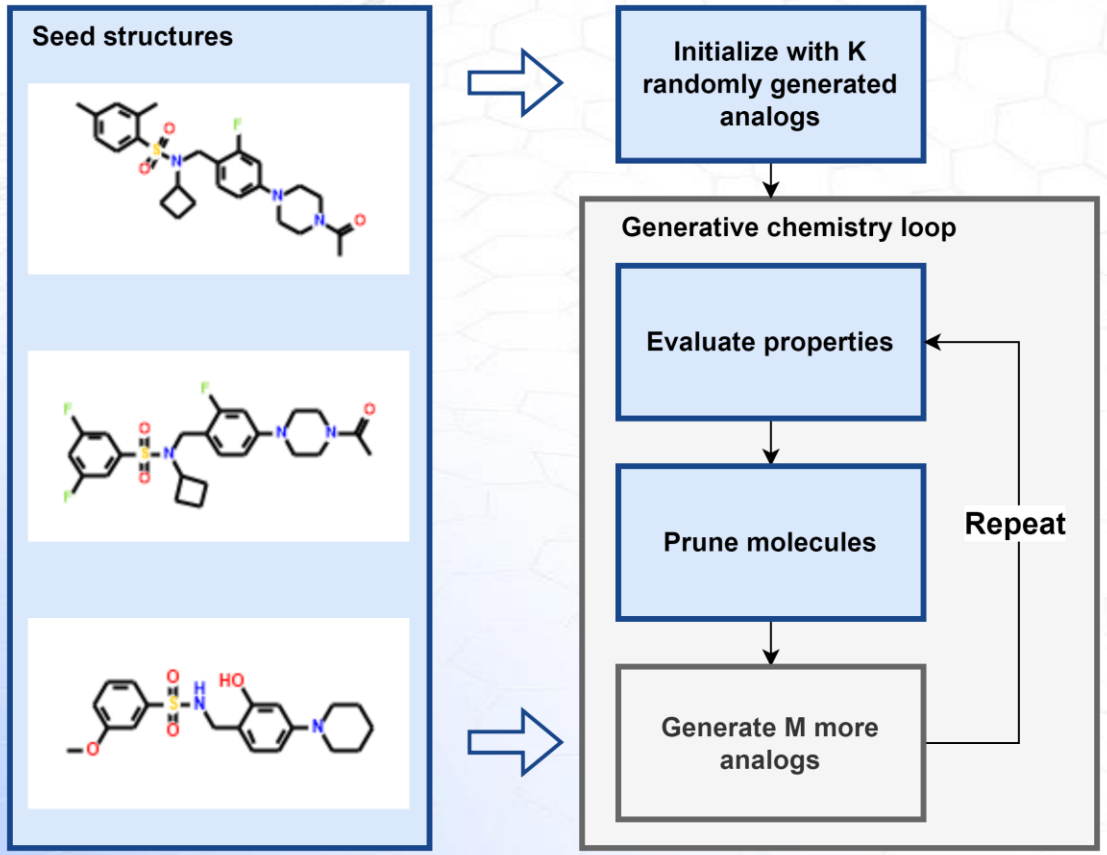
- Population initialization
- SMIRKS-based transformation
- Currently ca. 150 transforms available
- The pre-created population is the subject of further extension in the generative chemistry loop
- Efficient strategy of chemical space exploration

AIDD procedure



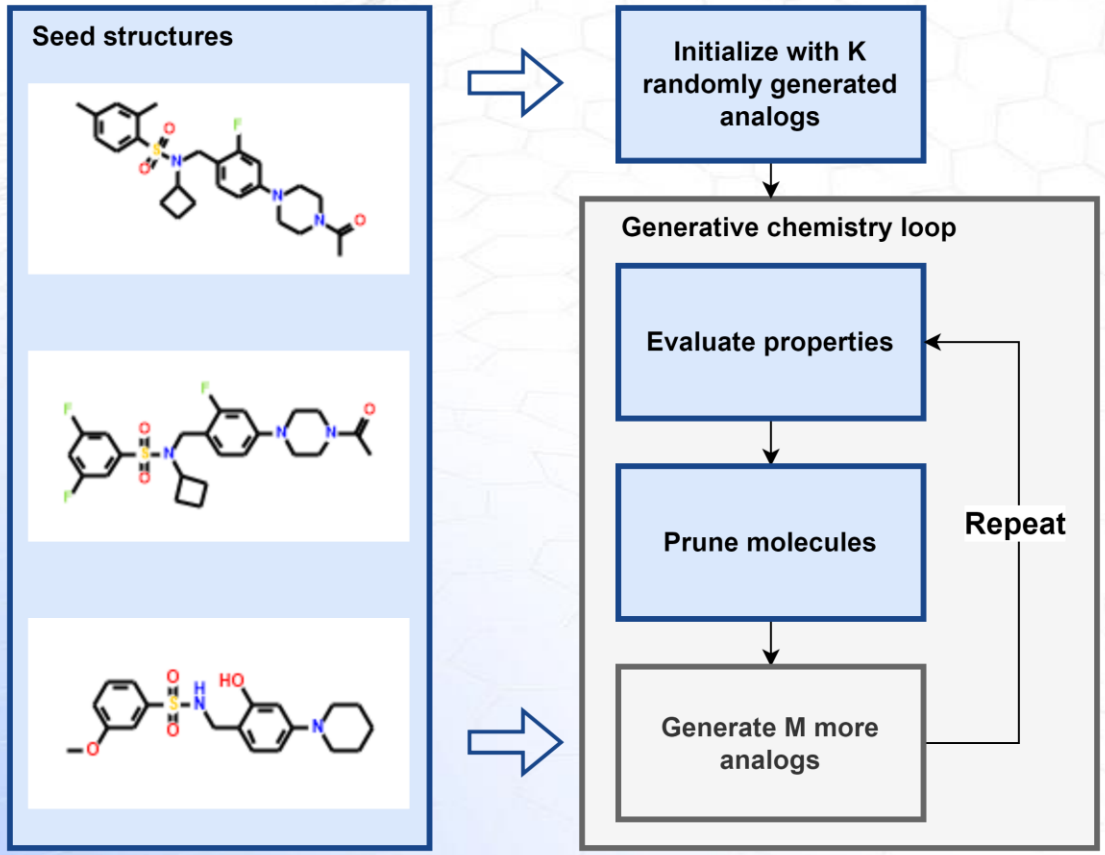
- Property evaluation
- Multiple ADMET Predictor properties available, e.g.:
  - Fraction absorbed
  - Fraction bioavailable
  - Blood-to-plasma ratio (RBP) for different species
  - Various flavours of solubilities (aqeuous, in simulated fasted gastric fluid, octanol-water partition coefficient, etc.)
  - SynthDiff+ reflecting the synthetic difficulty
  - 3D similarity to predefined ligand
  - Docking score obtained with external tools
  - Biological activity based on custom QSAR model

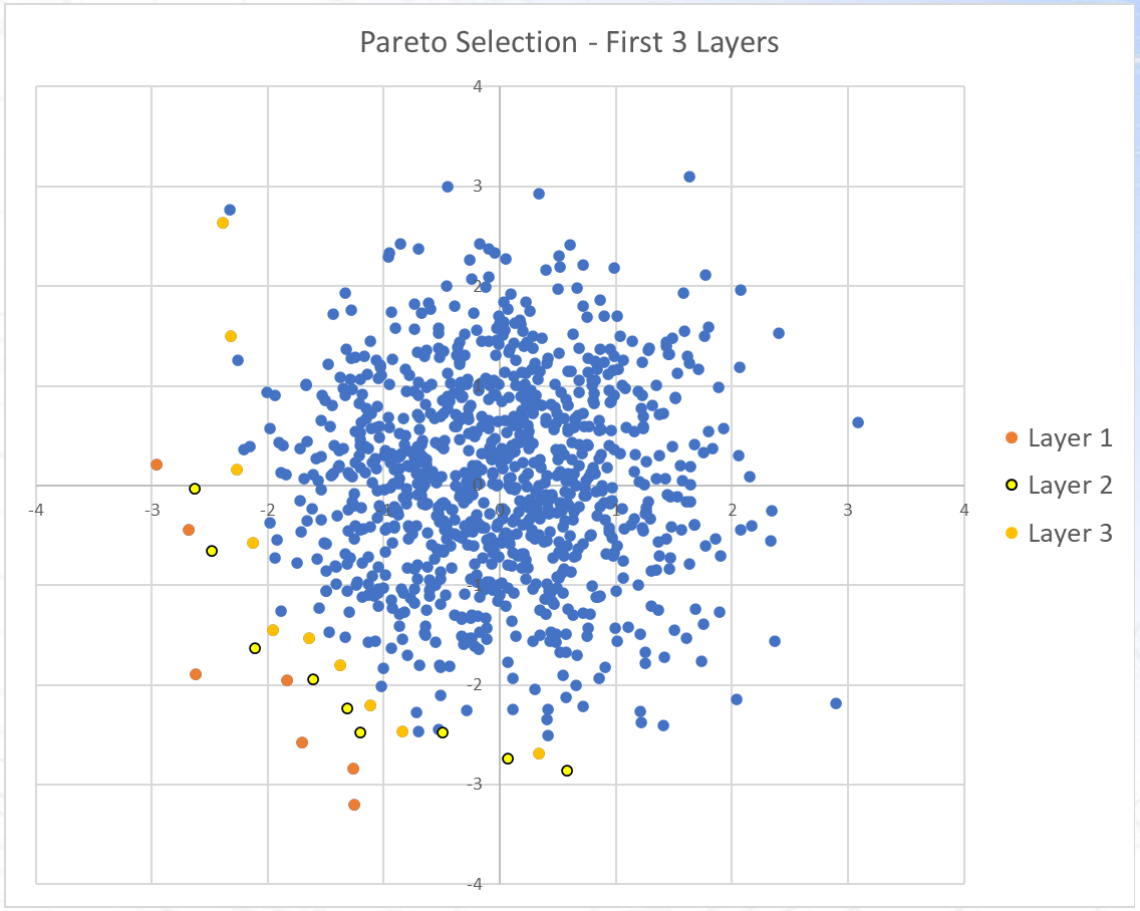
AIDD procedure



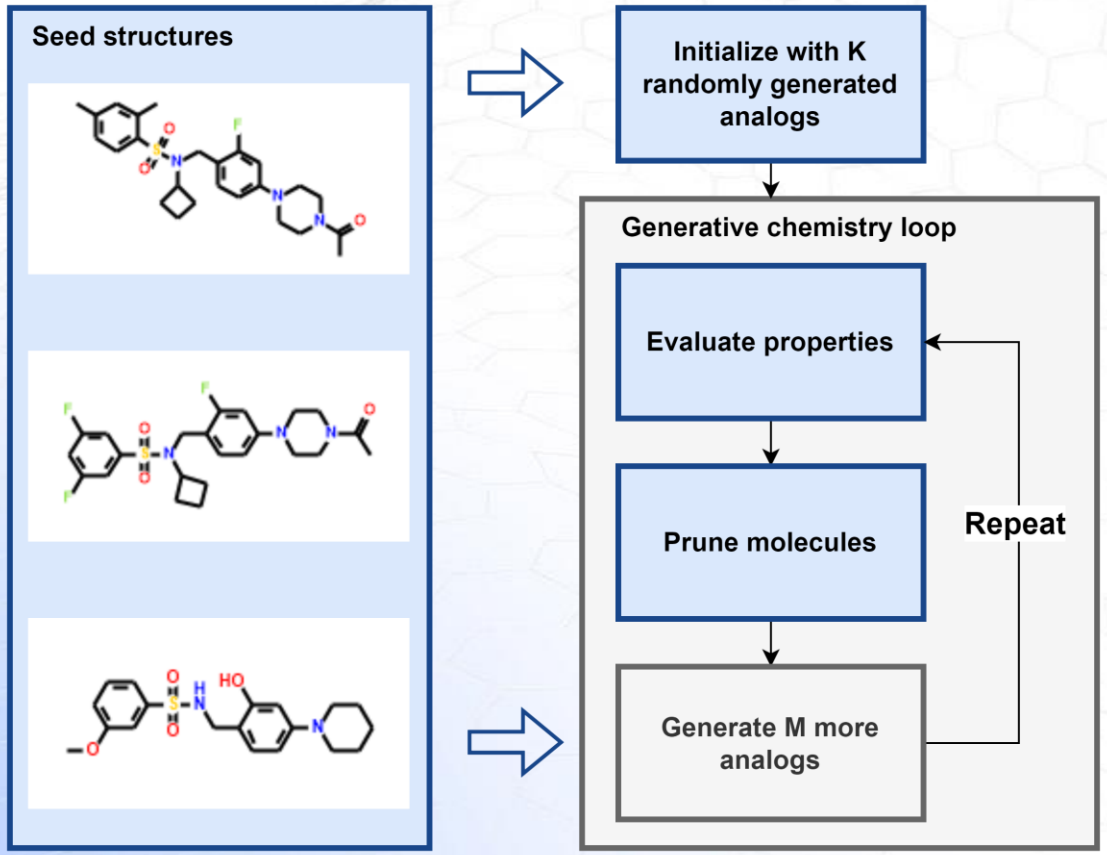
- Population pruning
- Multicriteria optimization
- Pareto selection
- Calculated properties form a space
- Succesive exploration of Pareto fronts
- Pareto-optimal solutions (molecules) form new population
- New population contains the most valuable individuals

AIDD procedure



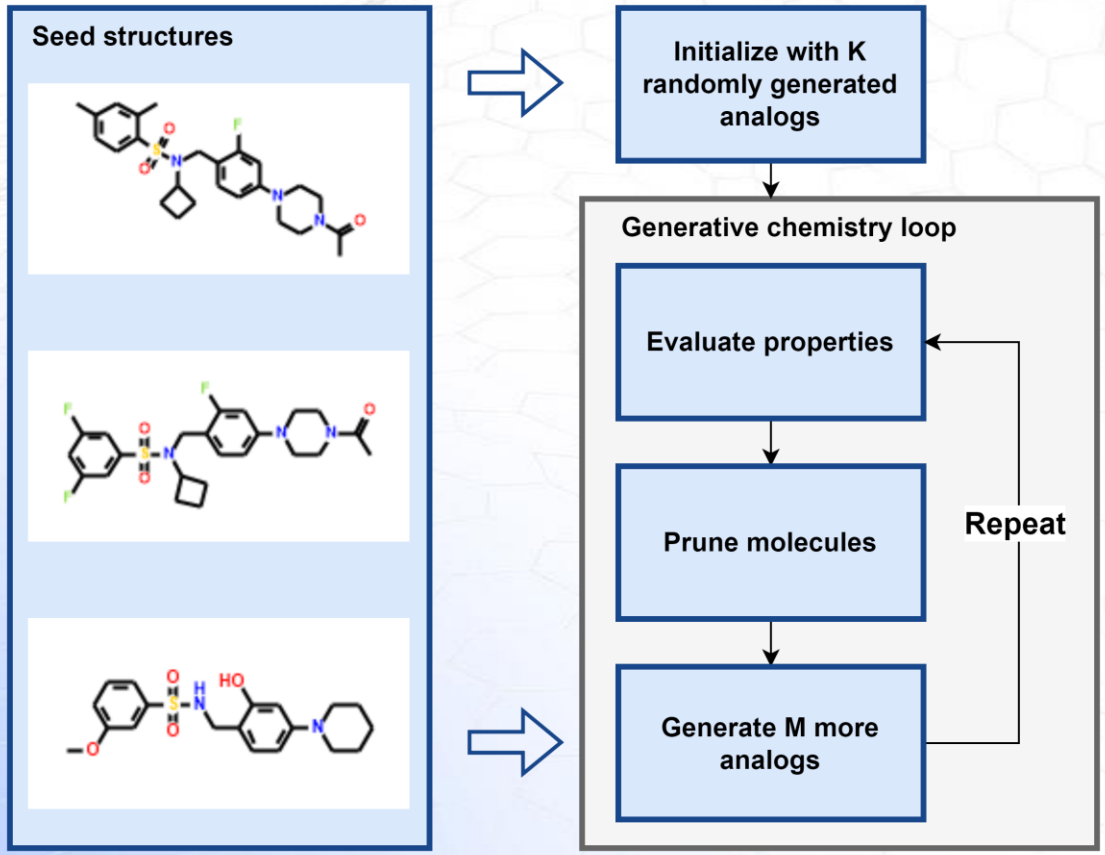


AIDD procedure



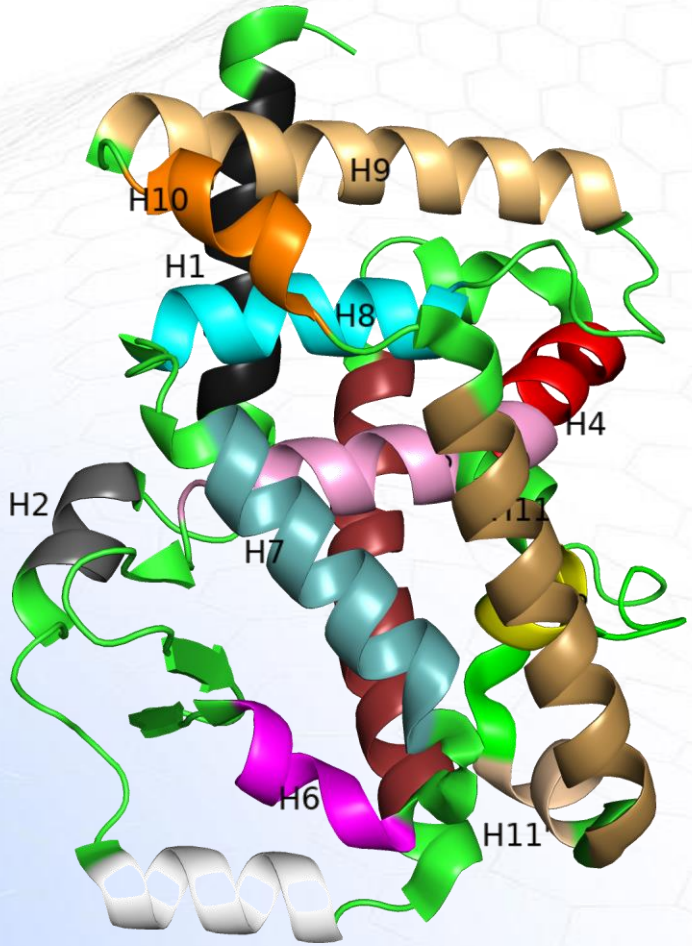
- Population pruning
- Multicriteria optimization
- Pareto selection
- Calculated properties form a space reflecting qualitative aspects of the molecules
- Succesive exploration of Pareto fronts
- Pareto-optimal solutions (molecules) form a new population
- New population contains the most valuable individuals

AIDD procedure



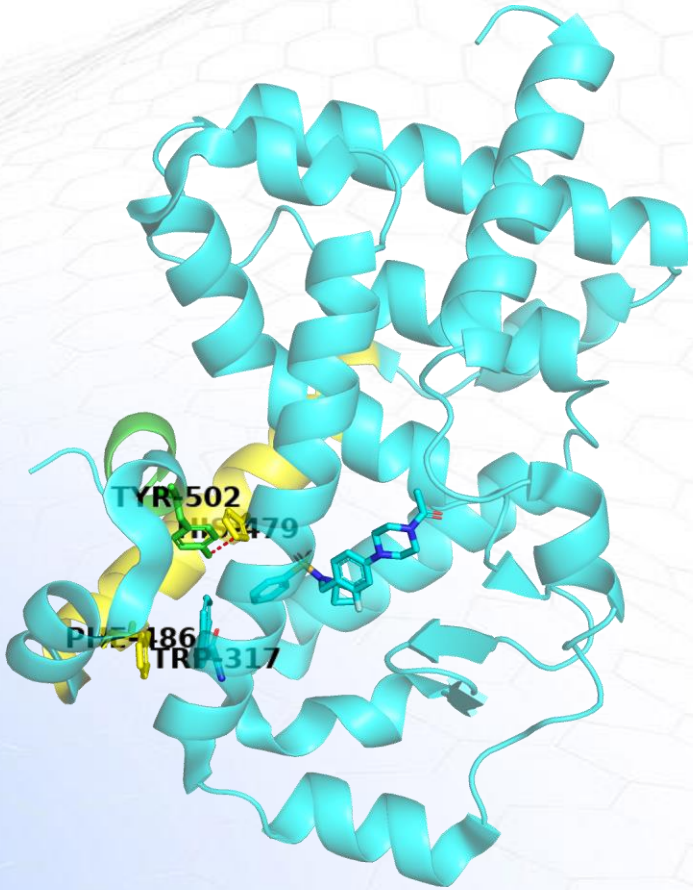
- Molecules generation
- SMIRKS transformation applied again
- Contrary do neural network-based generative approaches: fully interpretable
- Chemical space potentially restricted by predefined molecular scaffold
- Entire cycle repeated desired number of times
- Successive improvements of molecules

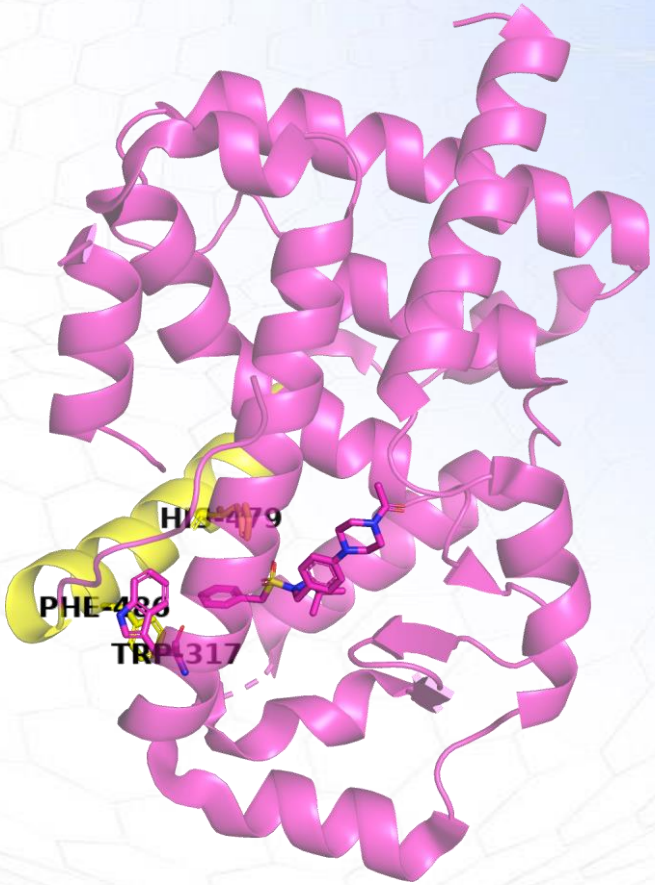
Nuclear receptors



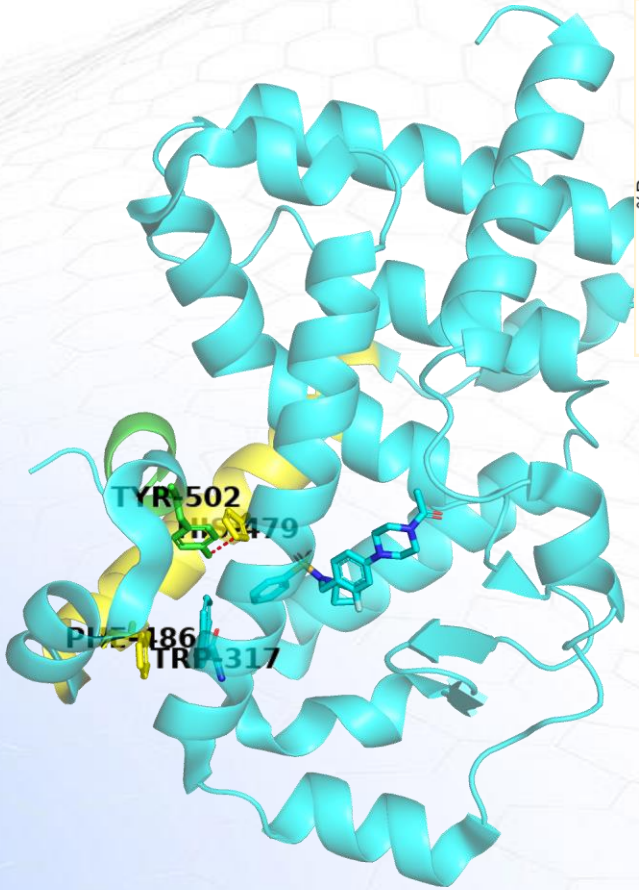
- Family of ligand regulated transcription factors
- RORs: the retinoic acid-related orphan receptors (α, β, γ)
- Many low-weight compounds exert biological activity by binding to NRs
- The activity of NRs depends on the conformation change
- The conformation change can be initiated by the binding of a small molecule to the protein moiety
- Binding of a ligand functions as a switch that induces a conformational switch
- Linked to many human diseases like: atherosclerosis, osteoporosis, autoimmunological disorders, obesity, asthma, and cancer

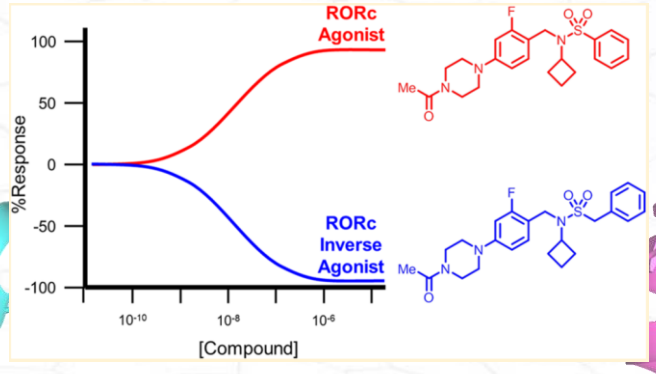
Agonistic vs. Inverse Agonistic activity



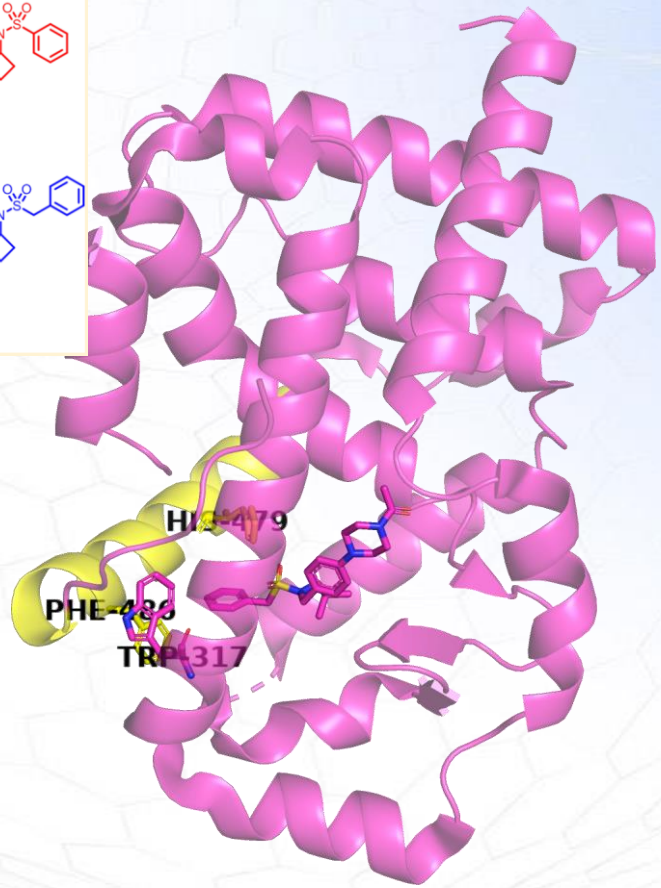


Agonistic vs. Inverse Agonistic activity

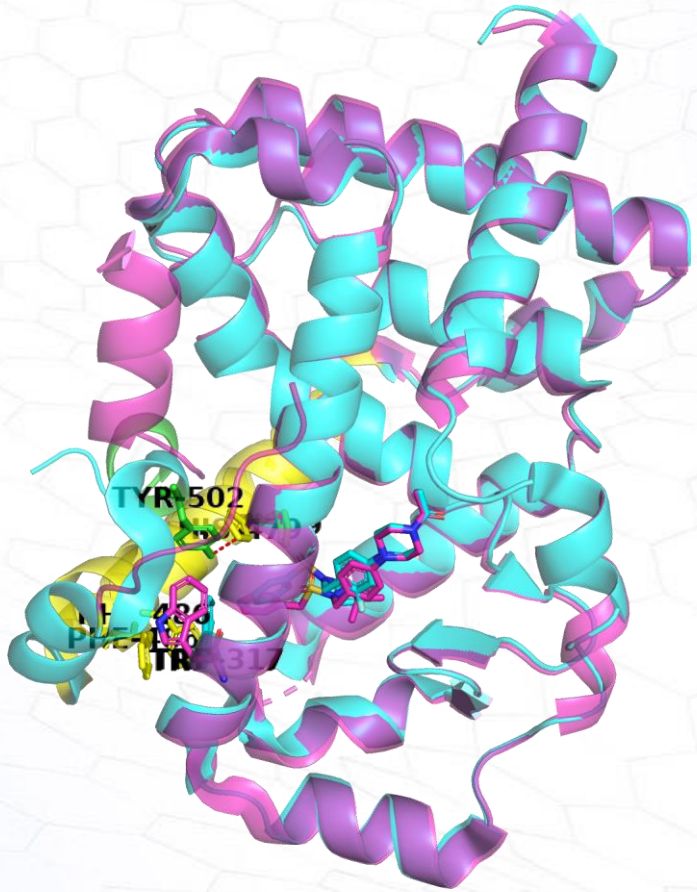




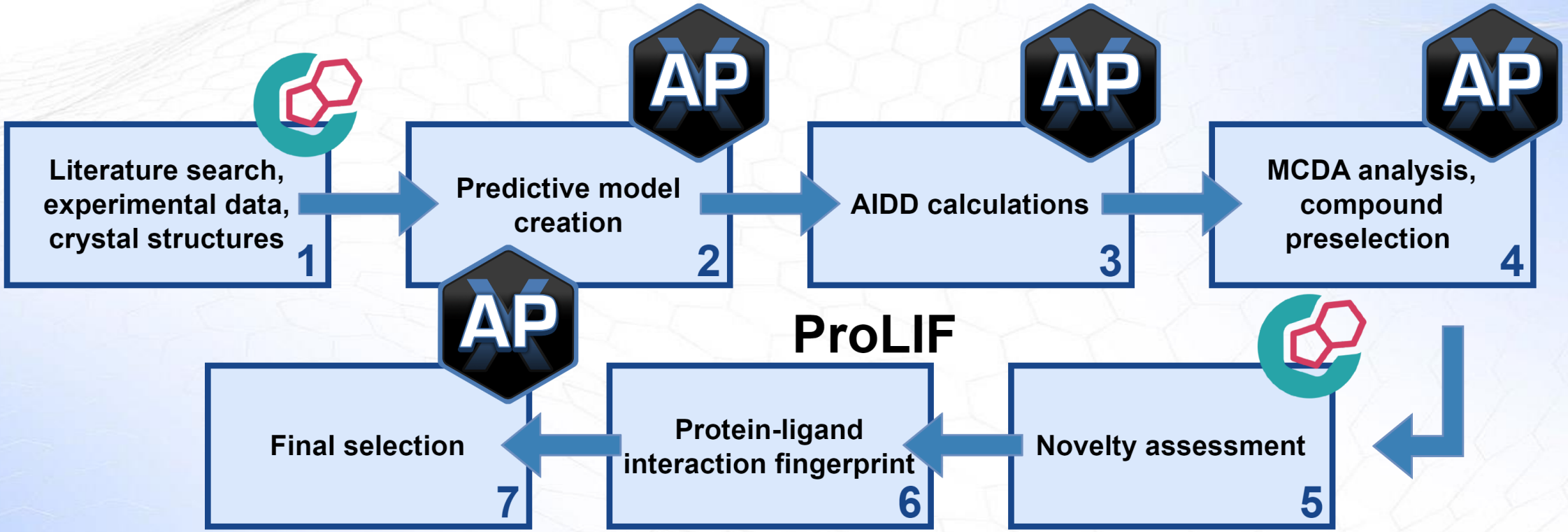
René et al. DOI: 10.1021/ml500420y



Agonistic vs. Inverse Agonistic activity

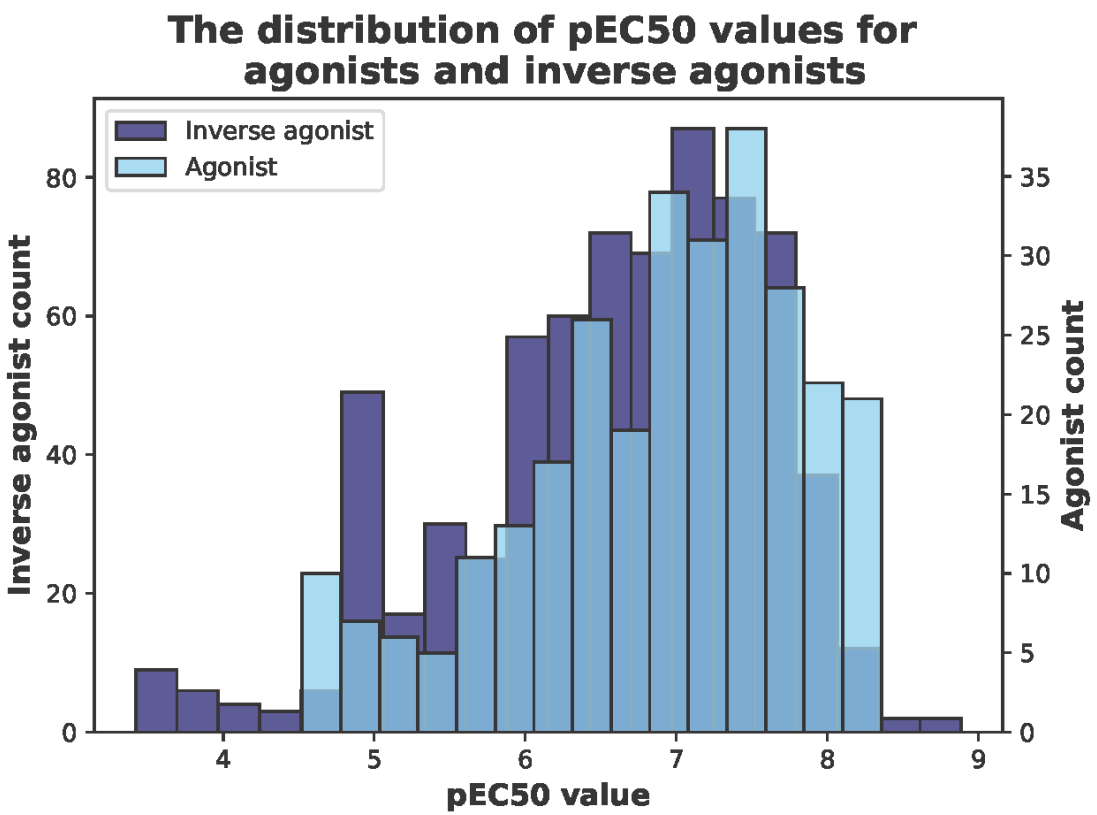


Designing new agonists



Literature search, experimental data, crystal structures 1

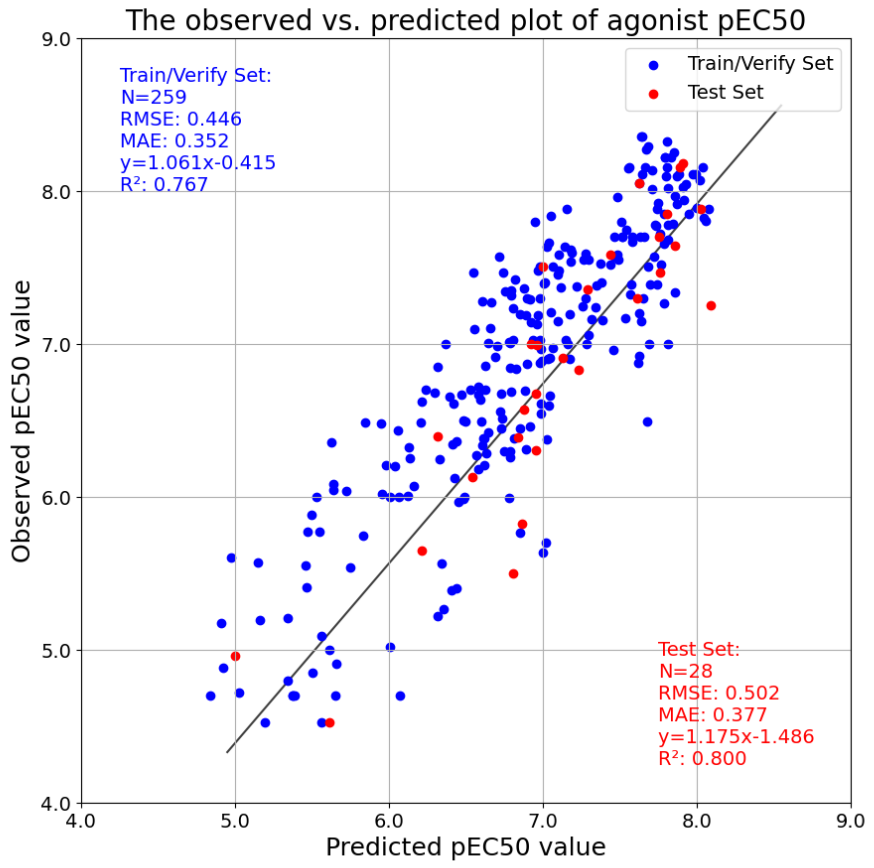
Literature search



- The ChEMBL data base was mainly used as a source of experimental data
- The biological data for ca. 3500 molecules, both IC50 and EC50 data included
- Expert knowledge was used to categorized molecules as agonists, inverse agonists, and antagonists
- The quality of experimental assays was also inspected – some endpoints were disregarded
- Manual cleaning and data curation
- Separate data sets for ROR agonists and inverse agonists

Predictive model creation 2

Biological activity QSAR model



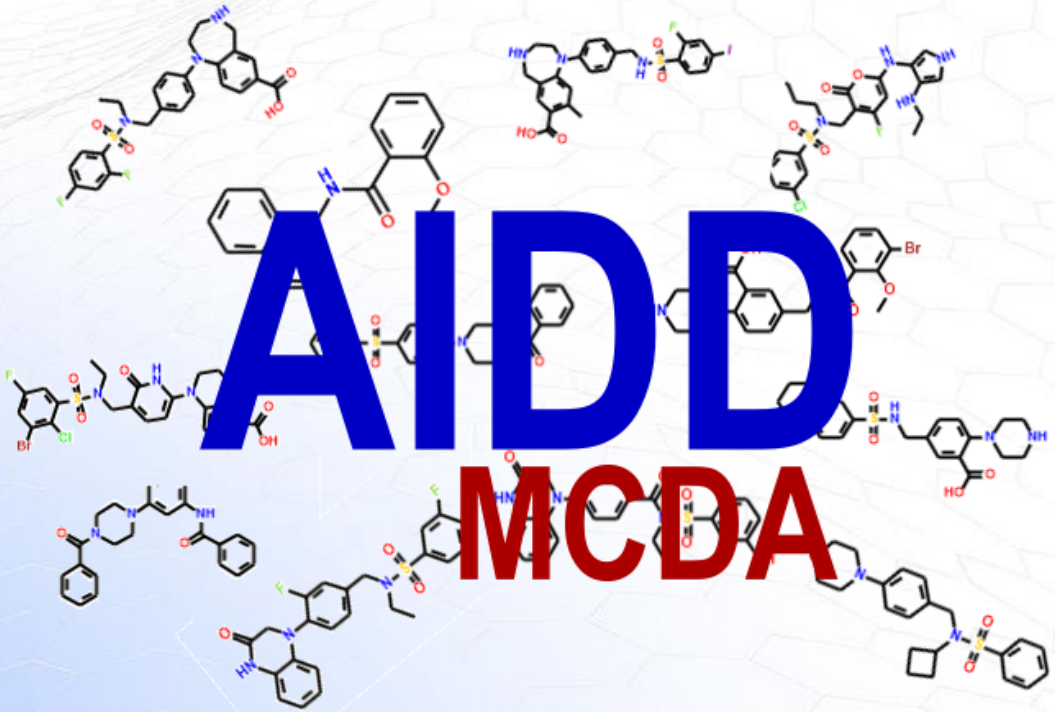
- ADMET Modeler™ applied here
- Artificial neural network approach
- Target quantity: pEC50
- Classification and regression models
- Threshold separating Active/Inactive classes: 1000 nM
- Key objective for the AIDD molecular optimization
- Good performance of the model

AIDD calculation

- Selected objective functions:
  - agonistic activity (predicted pEC50)
  - 3D pharmacophoric similarity
  - absorbtion risk
  - synthetic difficulty
- Multiple runs with different setups
- Various reference crystal structures for 3D similarity
- Molecular scaffold definition as a generative chemistry restriction

MCDA analysis, compound preselection 4

MCDA and novelty assessment

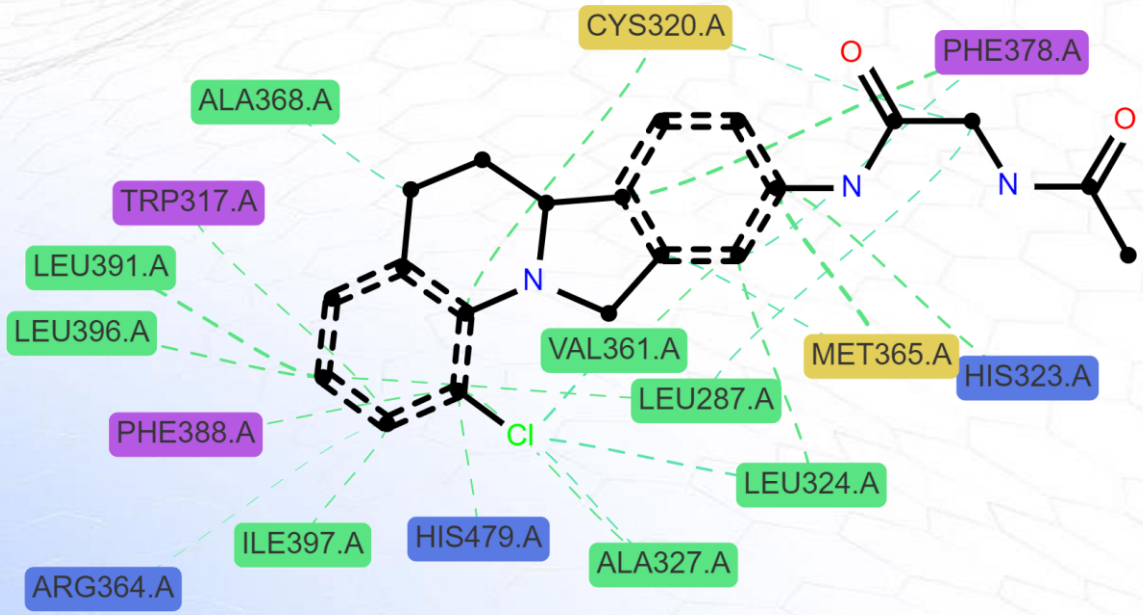


- Carefully inspect candidate molecules
- Multicriteria Decision Analysis techniques employed
  - Vikor
  - Topsis
- All objectives considred simultanously with different weights
- Ranking created
- Each molecule confronted with the SureChEMBL data base
- Patent data
- The exploration of commercially available compounds (https://arthor.docking.org/)

ProLIF

Protein-ligand interaction fingerprint 6

ProLIF: protein ligand fingerprint

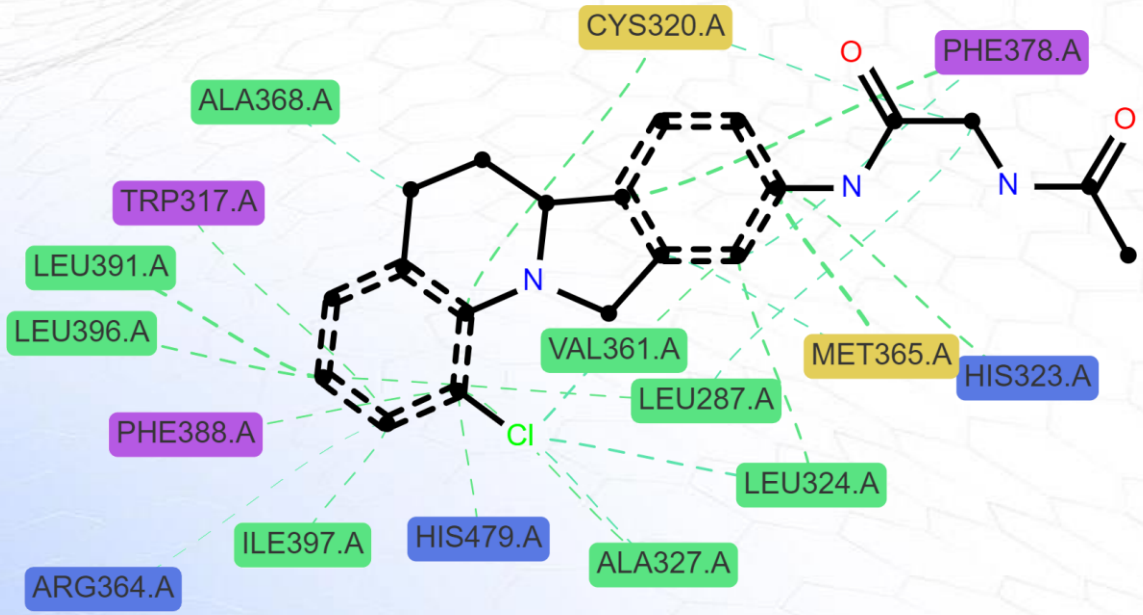


- Systematic analysis of the protein-ligand interaction
- Reference receptor: 4WPF
- Reference interaction with potent agonist
- The interactions are quantified in a vector form
- Analogous representation for AIDD-generated candidates
- Tanimoto similarity calculated
- Additional presumption for final selection

Final selection

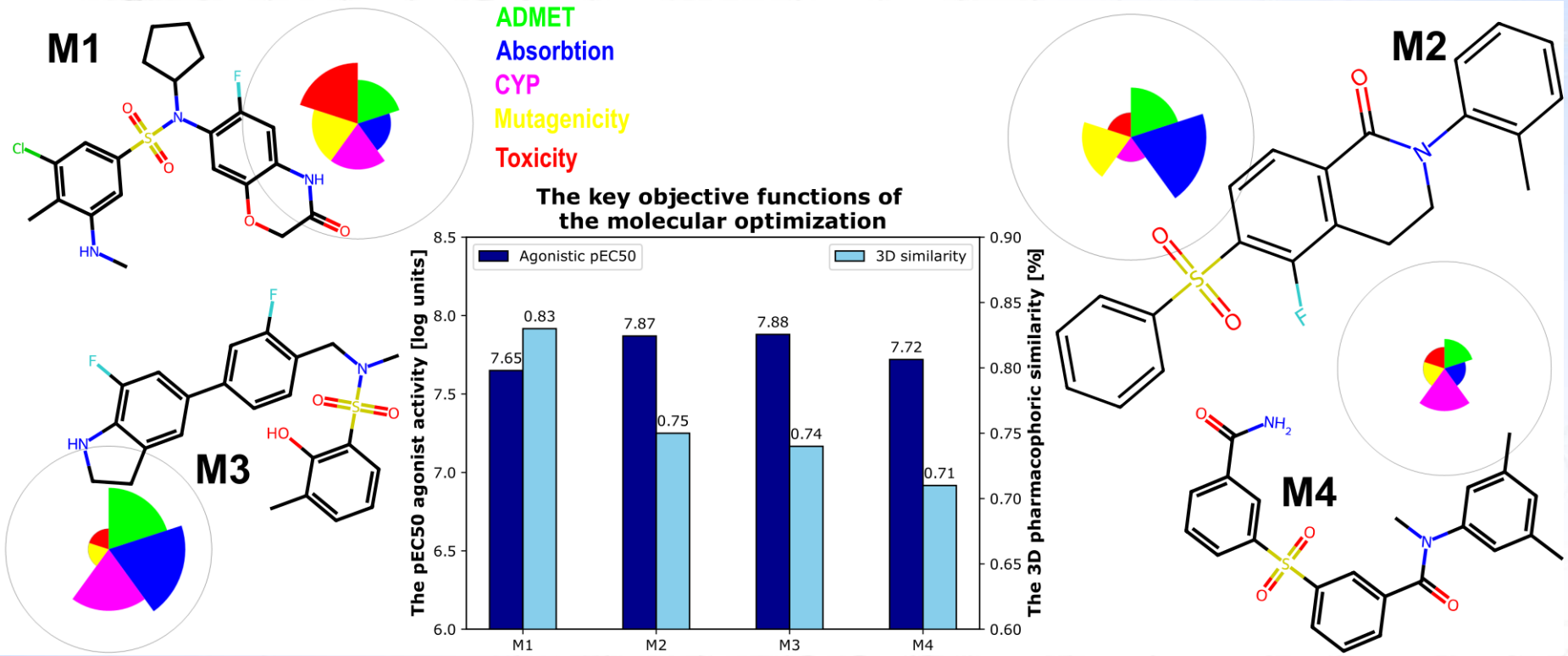
7

Final selection & Results



- Multiple AIDD calculations with different setups
- Majority of known ligand crystal structures taken into account in the 3D similarity calculation
- 65 compounds selected as final choices
- 28 compounds are being experimentally verified
- Institute of Medical Biology of Polish Academy of Sciences

Final selection & Results



Summary

- ADMET Predictor® as a *de novo* drug design environment
- AIDD: multicriteria generative chemistry tool
- Application to nuclear receptors
- Searching potent agonists of RORγ
- A workflow involving multiple techniques oriented on hit discovery
- Promissing candidates are being experimentally verified

SCAN ME

Acknowledgements

- Dr. Michael Lawless, Simulations Plus
- Dr. Jeremy Jones, Simulations Plus
- Dr. Marcin Ratajewski, Institute of Medical Biology of Polish Academy of Sciences

- SLP University+ Program

SCAN ME