ChemPass

AI-assisted lead optimization with SynSpace

Greg Makara
ChemPass Ltd.

Global Engage

Pharma R&D Informatics & AI Congress

October 28, 2019

ChemPass

- Founded 01/2016
- Location: Budapest, Hungary

- Creating a **synthesis-based design concept** and software for medicinal chemists and AI-powered drug discovery
- Built an **AI-powered lead discovery platform**
- SynSpace availability

aws
Microsoft Azure
openstack


Cloud


SaaS

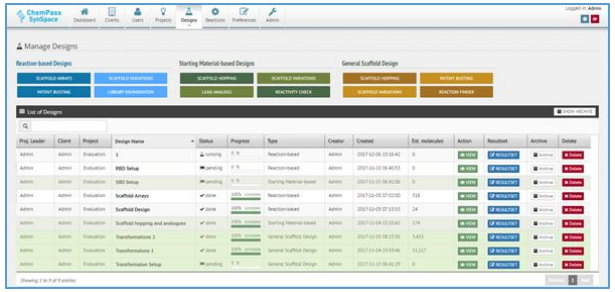

Server

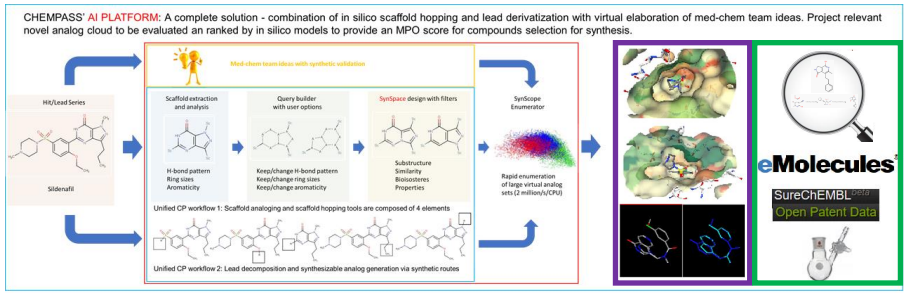

Desktop

- Aims: reduce cycle time, optimization time, expand idea and IP space

Key technologies

- ChemPass has both AI-based <u>software</u> for medicinal or computational chemists AND a complete AI-assisted lead discovery <u>platform</u>

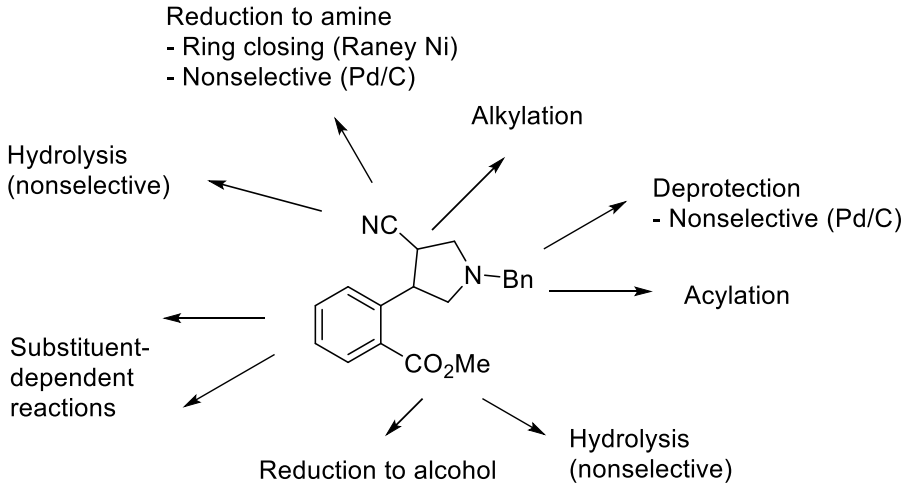


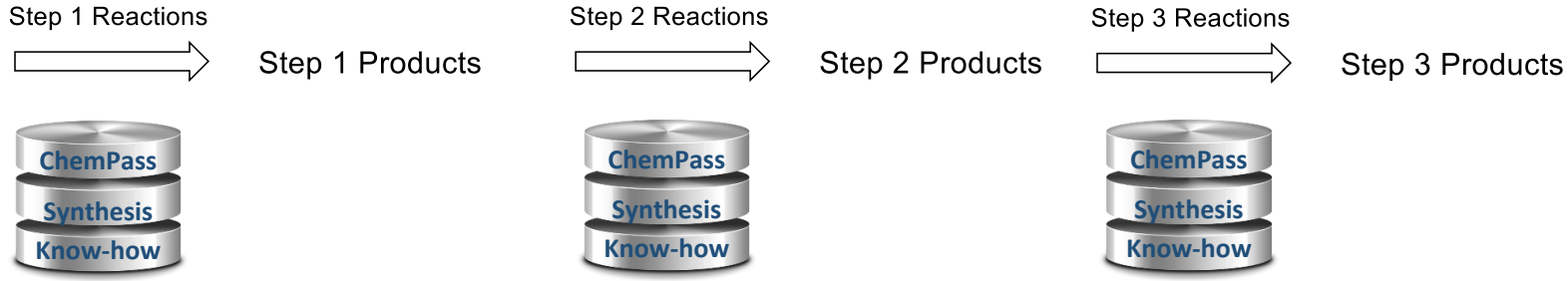


- ChemPass has AI-based technologies and software for both <u>design</u> of novel leads or lead analogs AND <u>retrosynthesis</u>

ChemPass design concept

- ChemPass Design Technology: What can be synthesized from starting materials, intermediates or lead structures?





Reactivity Assessment

Reactivity Assessment

Reactivity Assessment

- Unique proprietary technologies
  - Rule-based AI for forward in silico synthesis
  - Molecule design based on multistep in silico synthesis
  - Control of combinatorial explosion

Comparison to a popular DL design method

Feature	Generative deep learning*	ChemPass SynSpace
Can design focused but diverse novel structures for lead analog design	Yes	Yes
Fully controllable scaffold design	No	Yes
Fully controllable side-chain design	No	Yes
Design can be limited to structural areas	No	Yes
Required prior art	Significant	Little (1-)
Synthetic feasibility considered	No	Yes
Bottleneck – expert synthesis evaluation	Yes	No

Why use methods that incorporate no synthesis information?
ChemPass has the technology to generate lead analog space with synthetic chemistry incorporated!

*Ref: 1. Deep learning enables rapid identification of potent DDR1 kinase inhibitors. Nature Biotechnology. Vol 1038,37,1038–1040, 2019.
2. Generating Focused Molecule Libraries for Drug Discovery with Recurrent Neural Networks. ACS Cent. Sci. 4, 120−131, 2018.
3. Generative Recurrent Networks for De Novo Drug Design. Mol. Inf. 2018, 37, 1700111.

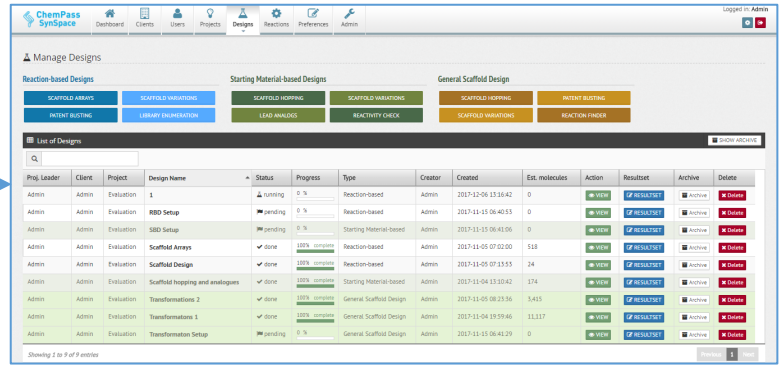
Products/services

- Software license:
  - SynSpace for medicinal chemists and computational chemists
  - SynSpace API for computational chemists
  - SynSpace modules
    - Library design (e.g. DEL)
    - Side-chain design
    - Scaffold and lead analog design modules (new fully automated AI modules)
    - Retrosynthesis module (new)

- Services:
  - Design and problem-solving medicinal chemistry support
  - AI-driven lead optimization platform for lead generation and lead optimization

SynSpace software availability and uses

ChemPass Designer Technology

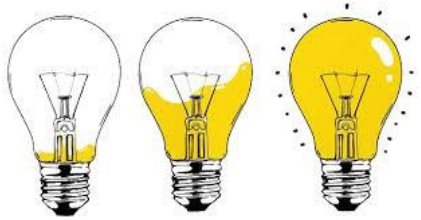


Lead discovery project support

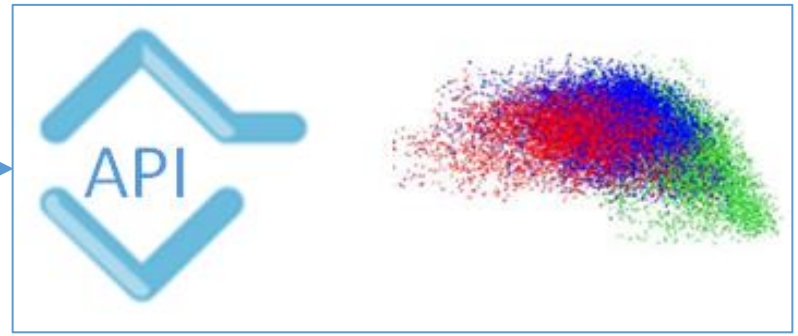
SYNSPACE Software

Scaffold, scaffold analog, lead analog design by chemists and cheminformaticians

Reduce cycle time



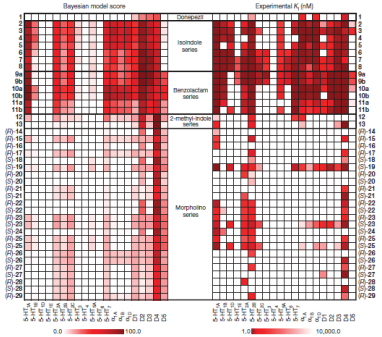
Expand idea space



Machine learning project support

SYNSPACE API

Synthetically enabled scaffold and vast virtual library design from available reagents



Deep learning, AI-powered drug discovery

ChemPass

Lead optimization support: **36 projects in 2018-2019**

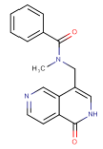
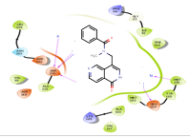
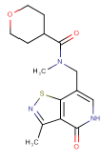
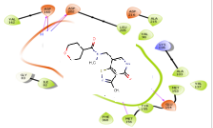
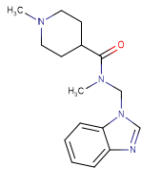
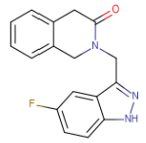
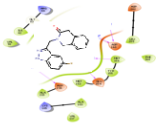
- Scaffold hopping for many different scaffold types
- Side-chain design for specific scaffolds
- Side-chain design in combination with scaffold hopping
- Hit expansion
- Probe molecule to lead
- Patent to new lead (busting)
- ADMET liability removal in late stage lead optimization
- Monocycle to bicycle lead transformation
- Create freedom of operation in crowded IP space against a target
- Create follow-on secondary series from primary series after candidate designation

Partners:
- Big Pharma
- Big Biotech
- Small biotech

Example collaboration report for lead optimization

Key information for rapid decision making:

- Structure
- ID
- Docking pose
- Docking interactions
- Docking scores
- Selectivity assessment
- Synthesis score
- Reagent availability score
- Synthesis scheme
- Lead likeness
- Selected properties (not shown)
- Flags
- Novelty assessment
- Commercial availability assessment

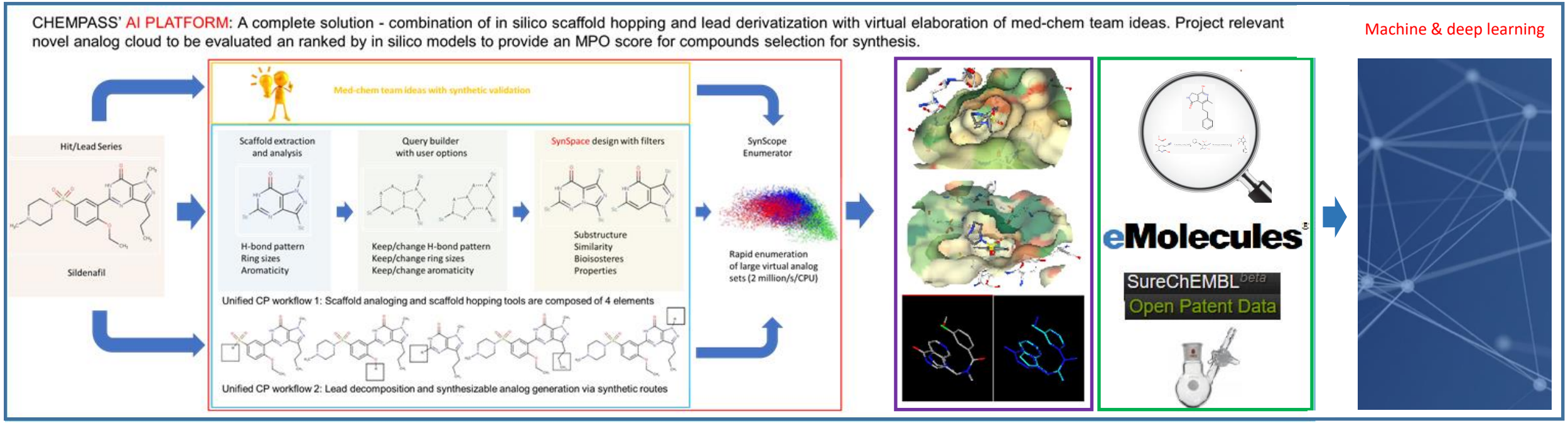
Molecule	Compound_id	Target_docking_pose_interactions	Target_docking score	Off_target_docking score	Docking_selectivity	Synthesis_score_(1-5)	BB_avail_score_(1-5)	Synthesis_route	Lead-likeness	Structural_Flag	Novelty_in_patent	Novelty_in_eMolecules
	"[[X14+#4-C62]-S9]"		-10.05	-7.78	2.27	2	2	Aryl_ketone_synthesis_Rxn62(62), Catalytic_reduction_50oC/30bar_Rxn9(9)	true	1	N	N
	"[#6+X4-C64]"		-10.01	-7.74	2.27	2	2	Ullmann_reaction_Rxn64(64)	true	0	N	N
	"[[#3+X55-C64]-S8]"		-9.98	-6.46	3.52	2	2	Ullmann_reaction_Rxn64(64),Catalytic_reduction_rt/1bar_Rxn8(8)	true	0	N	N
	"[[X14+#4-C62]-S9]"		-9.85	-8.96	0.89	2	2	Aryl_ketone_synthesis_Rxn62(62), Catalytic_reduction_50oC/30bar_Rxn9(9)	true	1	N	N

Extracted from real report with altered structures

Report format and data types are fully customizable

A detailed table is also available with synthesis scheme, reagent structures, ids, cost etc., hundreds of customizable properties

Complete AI-driven lead optimization platform

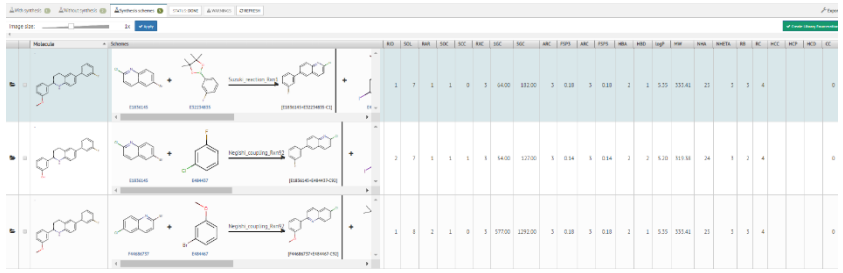


AI-powered ideation: scaffold hopping, substituent designs
Property predictions
Potency, selectivity predictions: semi-automated protein or ligand-based methods

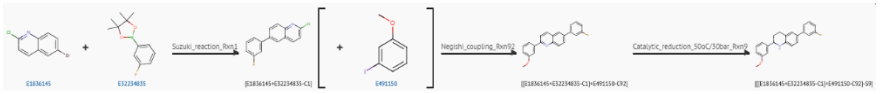
Full synthesis and reagent availability assessment
Novelty assessment
Synthesis outsourcing

MPO scores
Ranking
Selection

Med-chem team ideas



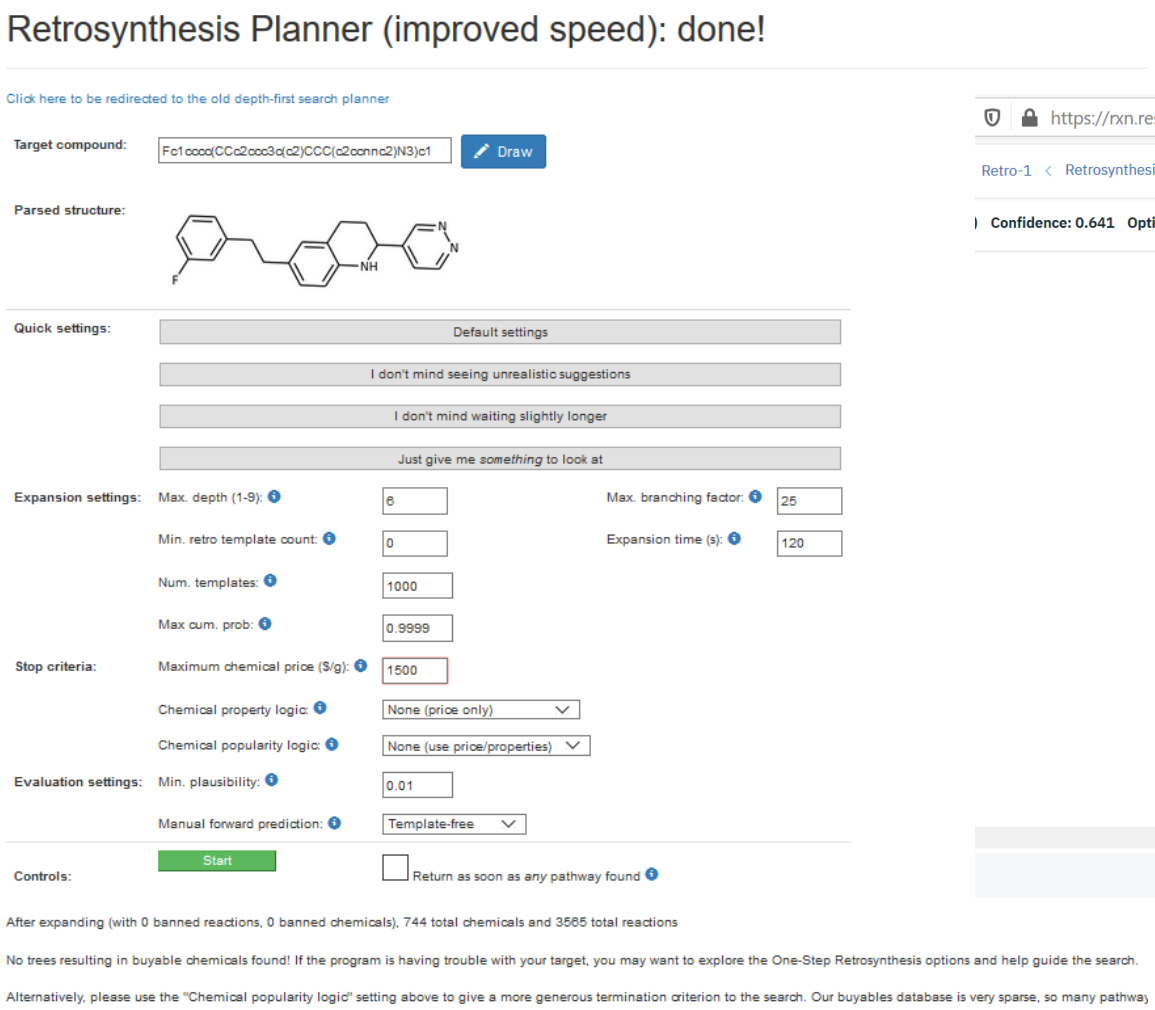
SynSpace retrosynthesis module

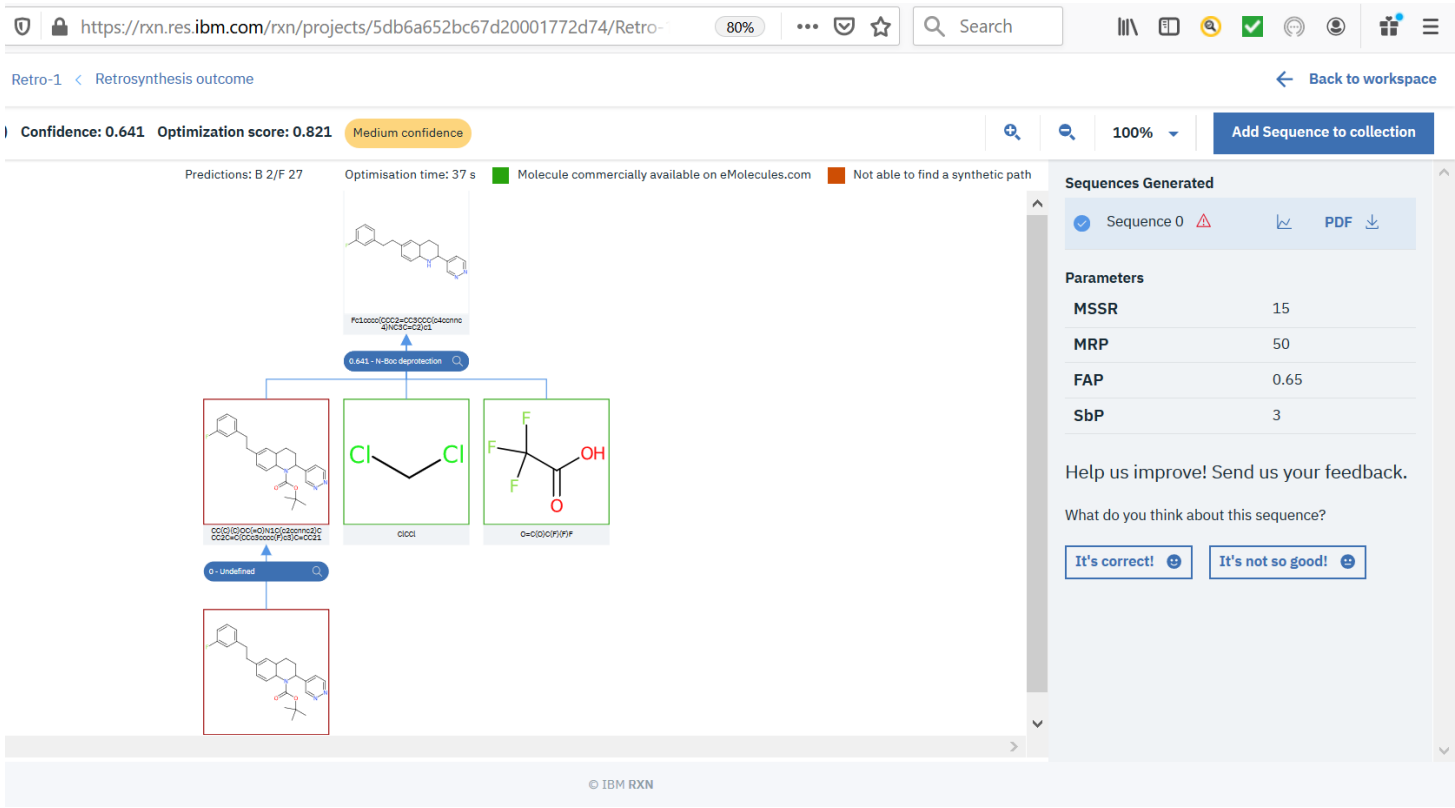


- Synthetic schemes, rankings
- Reagent structures, IDs

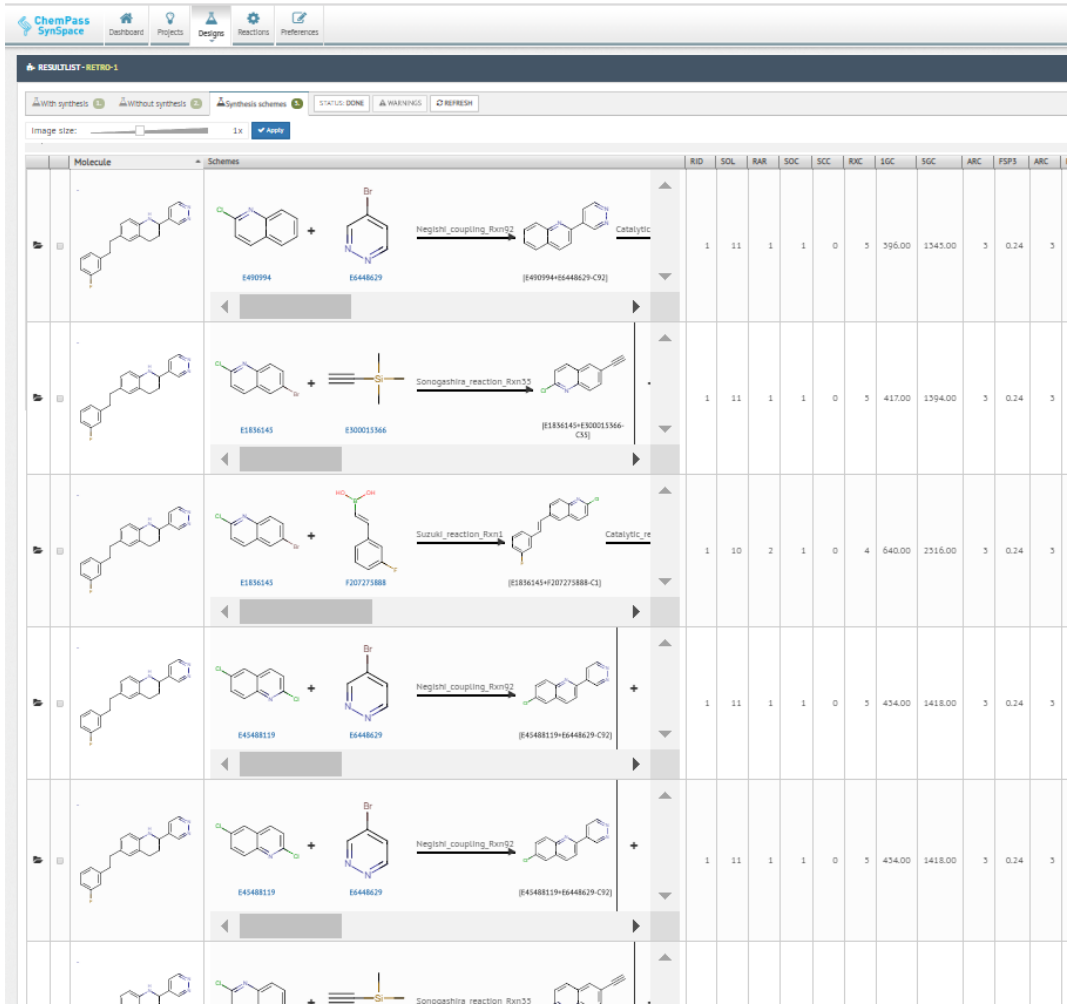
- Retrosynthesis enables human ideas to enter the design workflow

Retrosynthesis example 1

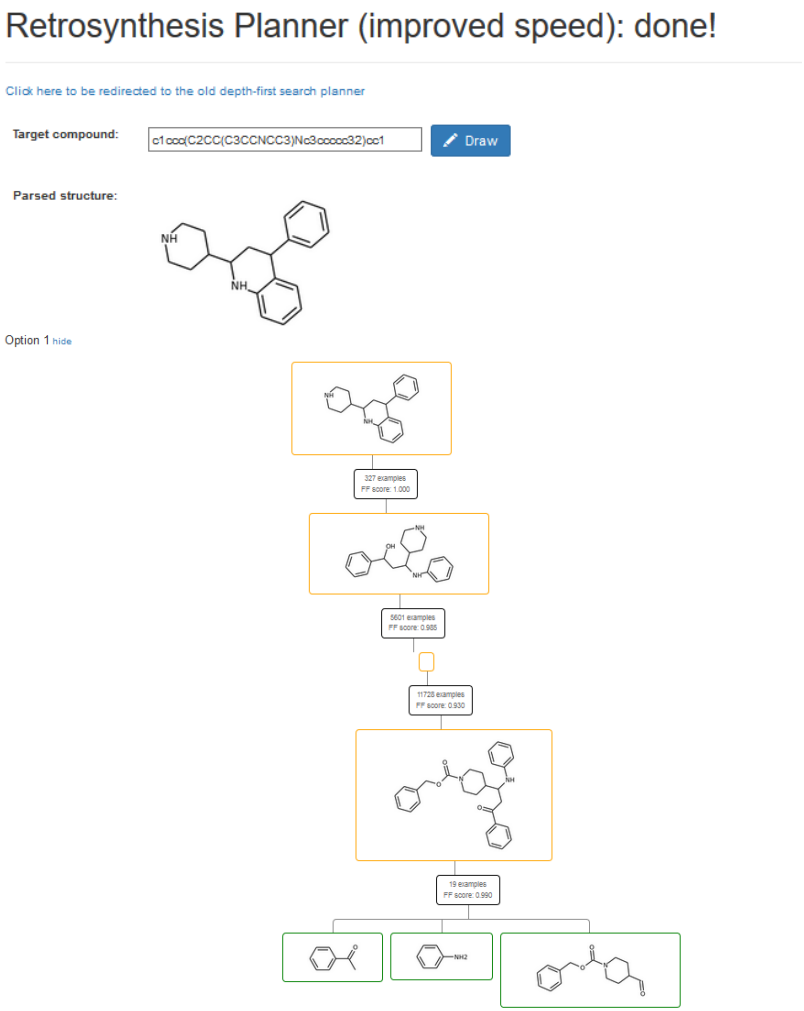


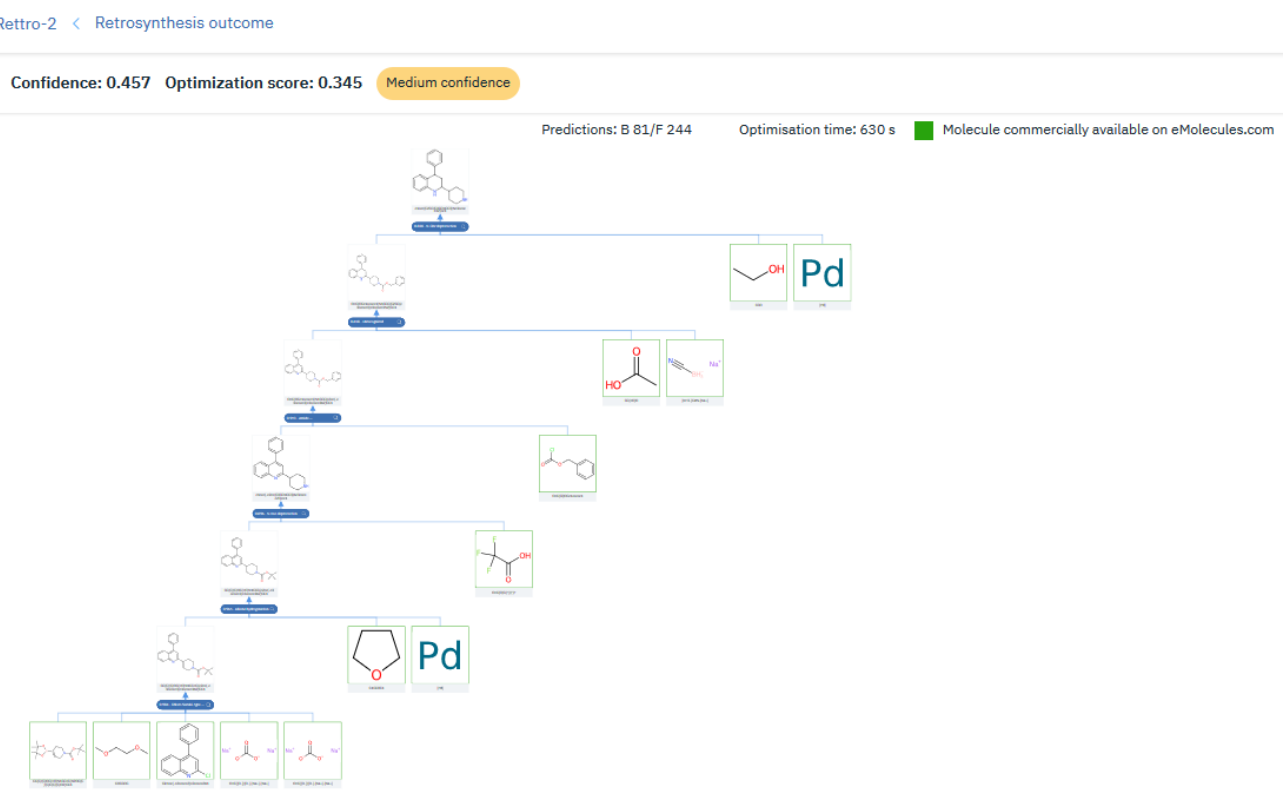


Retrosynthesis example 1

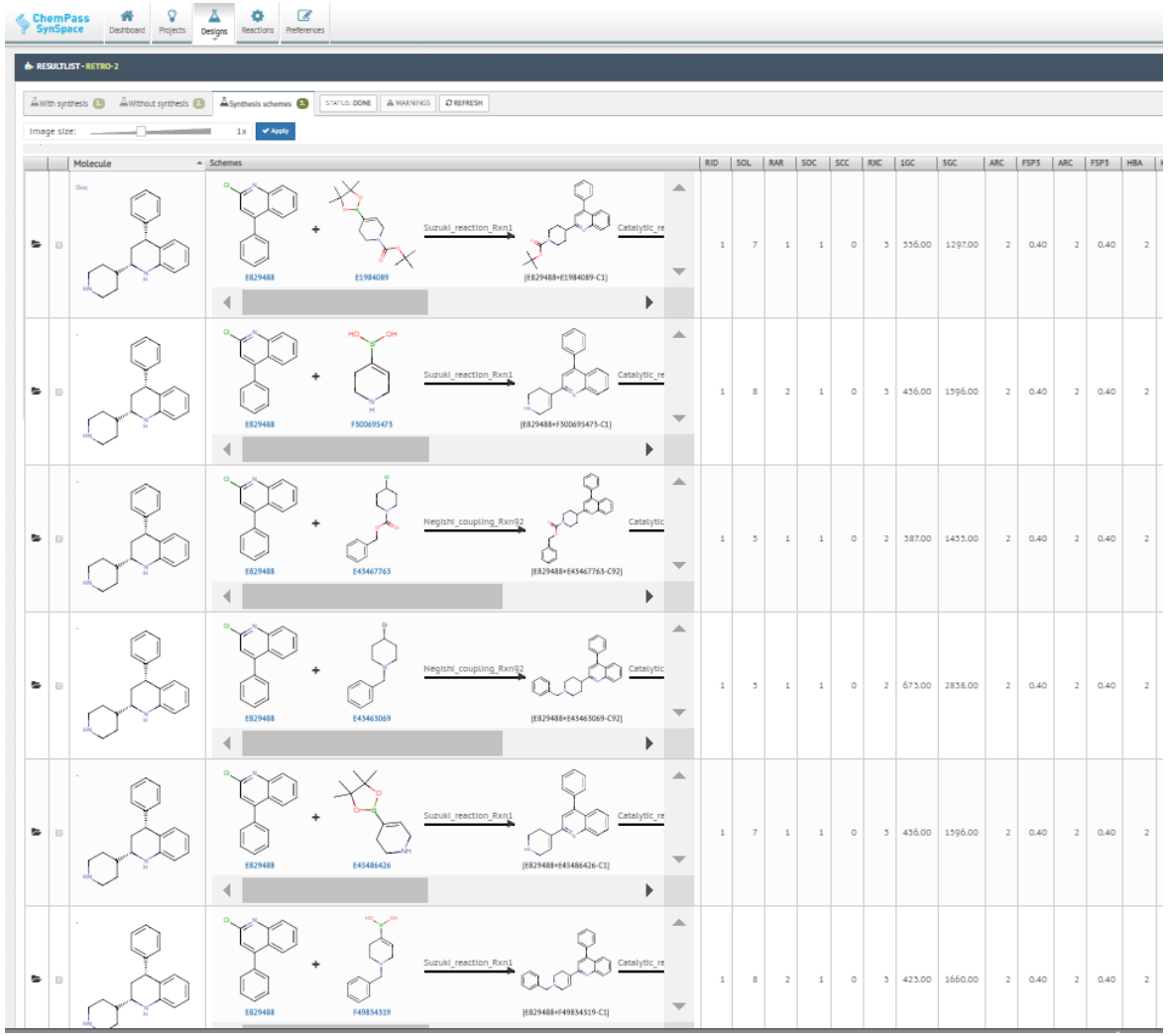


Retrosynthesis example 2

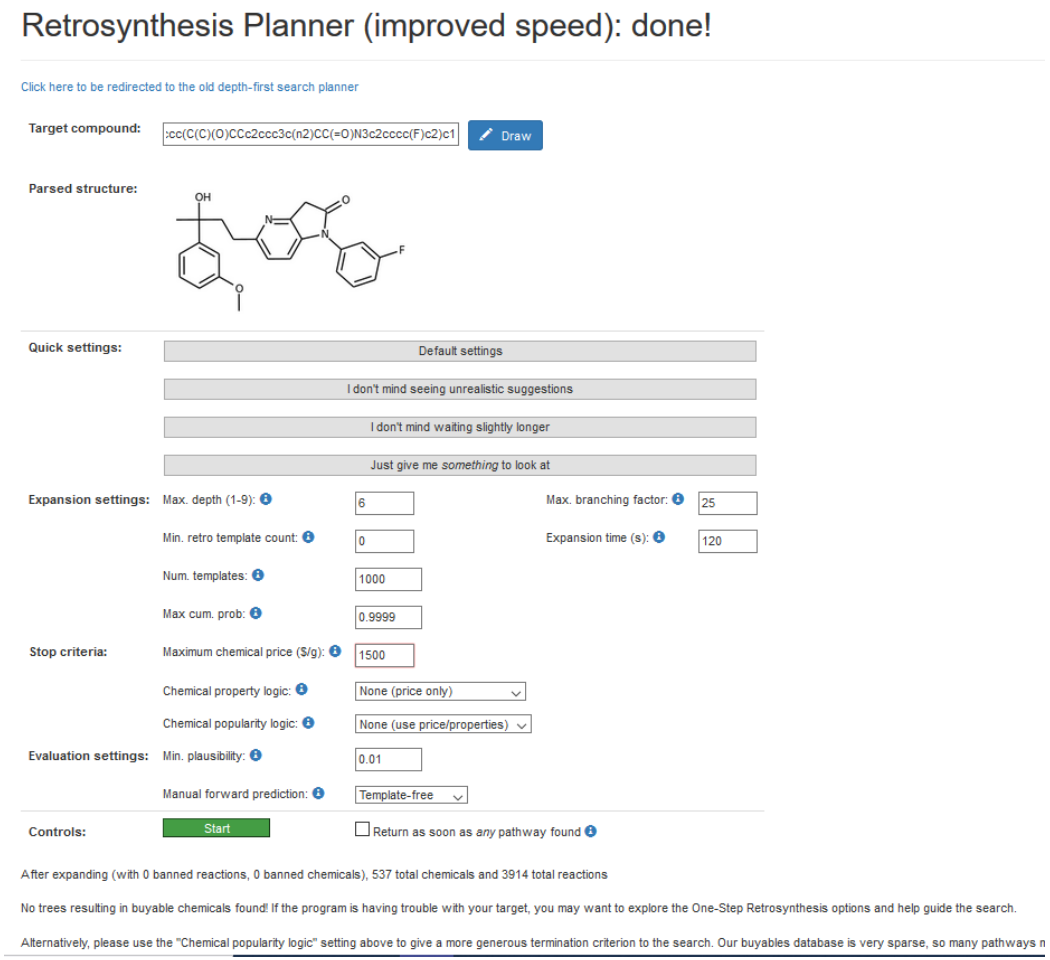


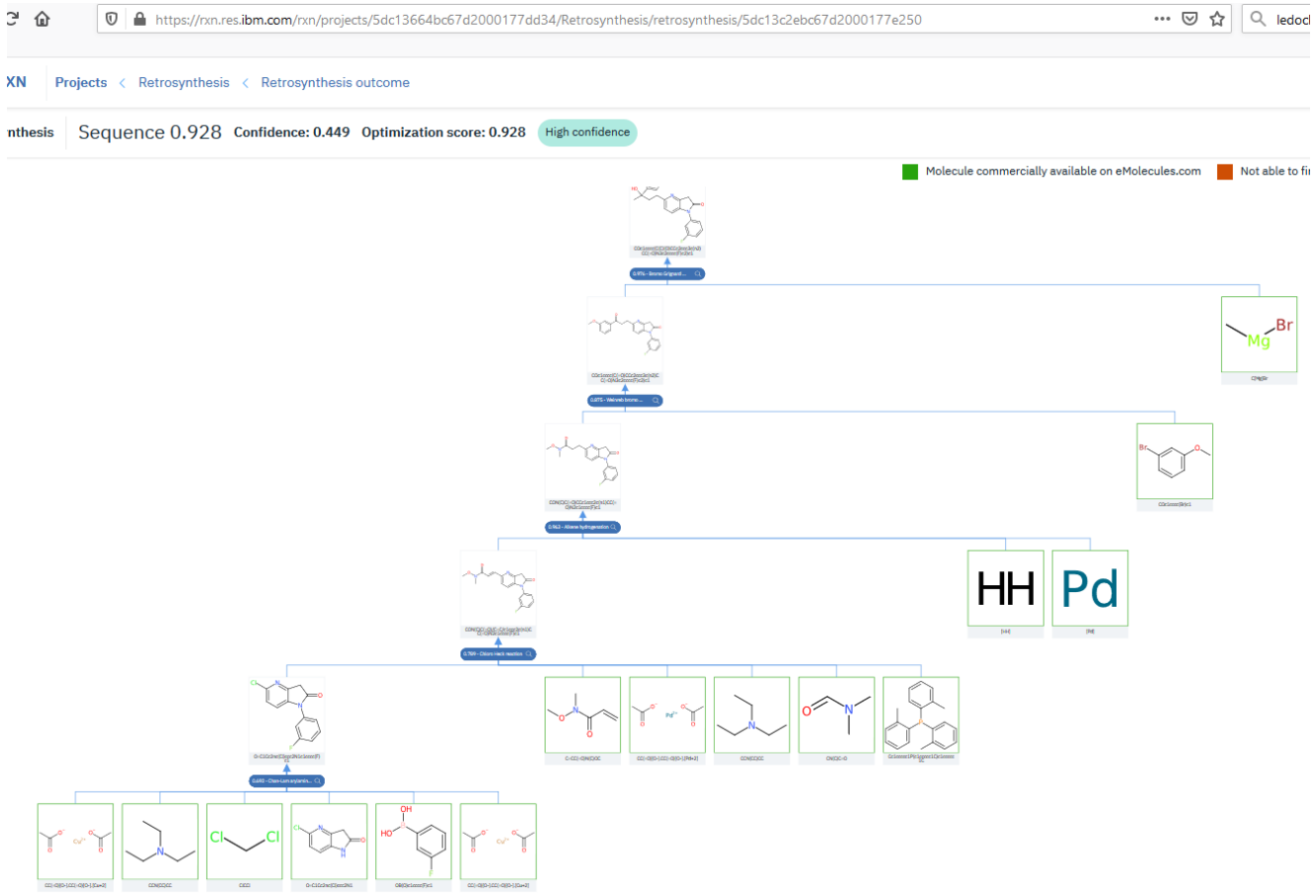


Retrosynthesis example 2

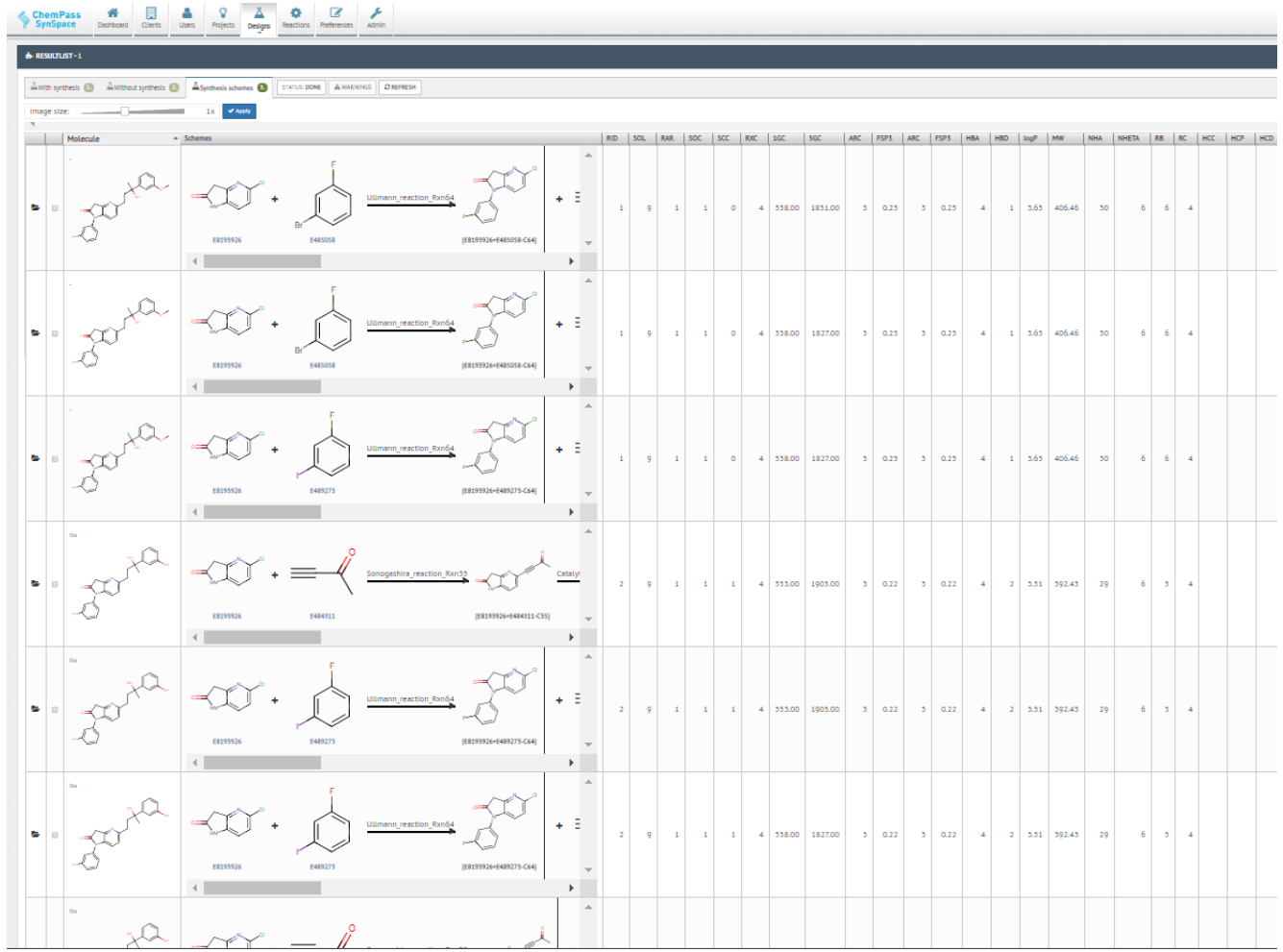


Retrosynthesis example 3

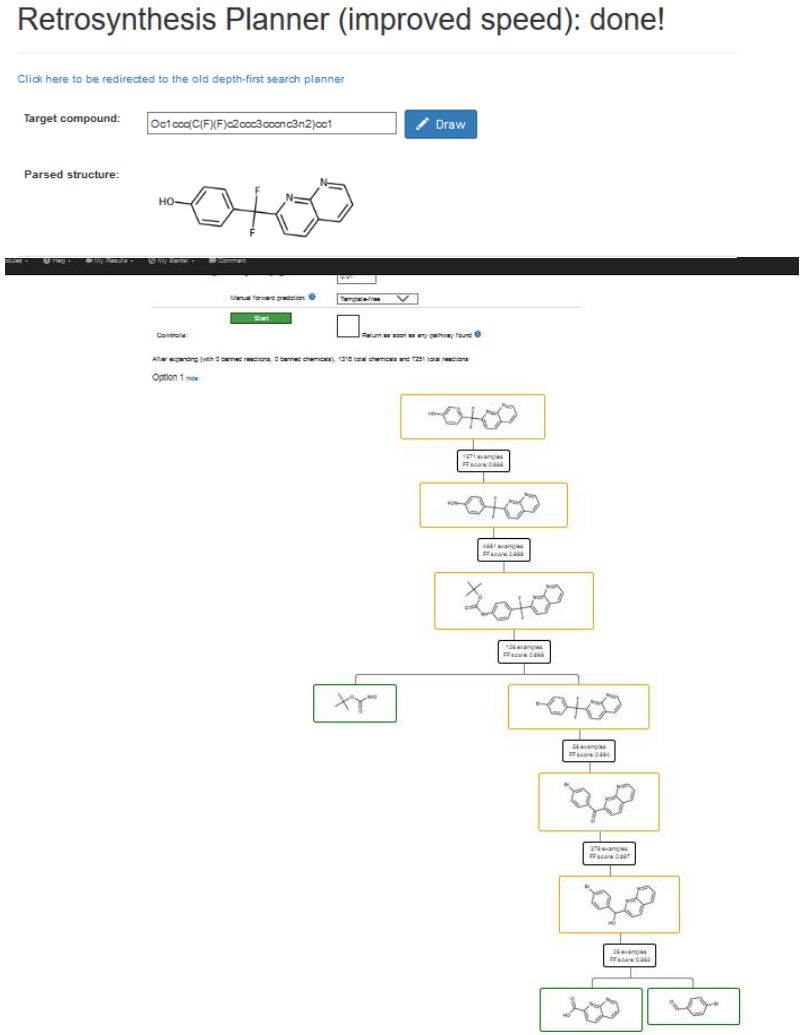


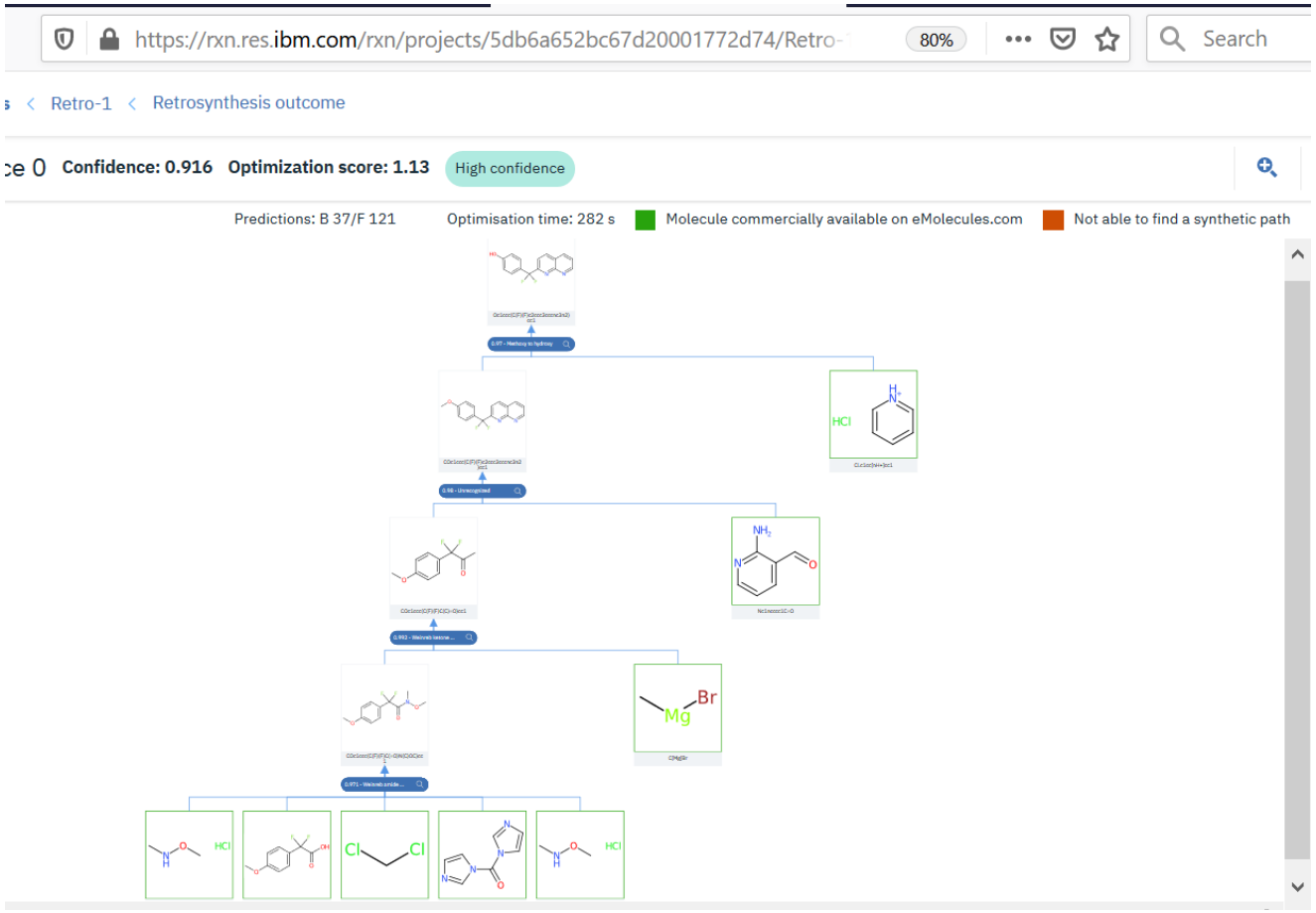


Retrosynthesis example 3

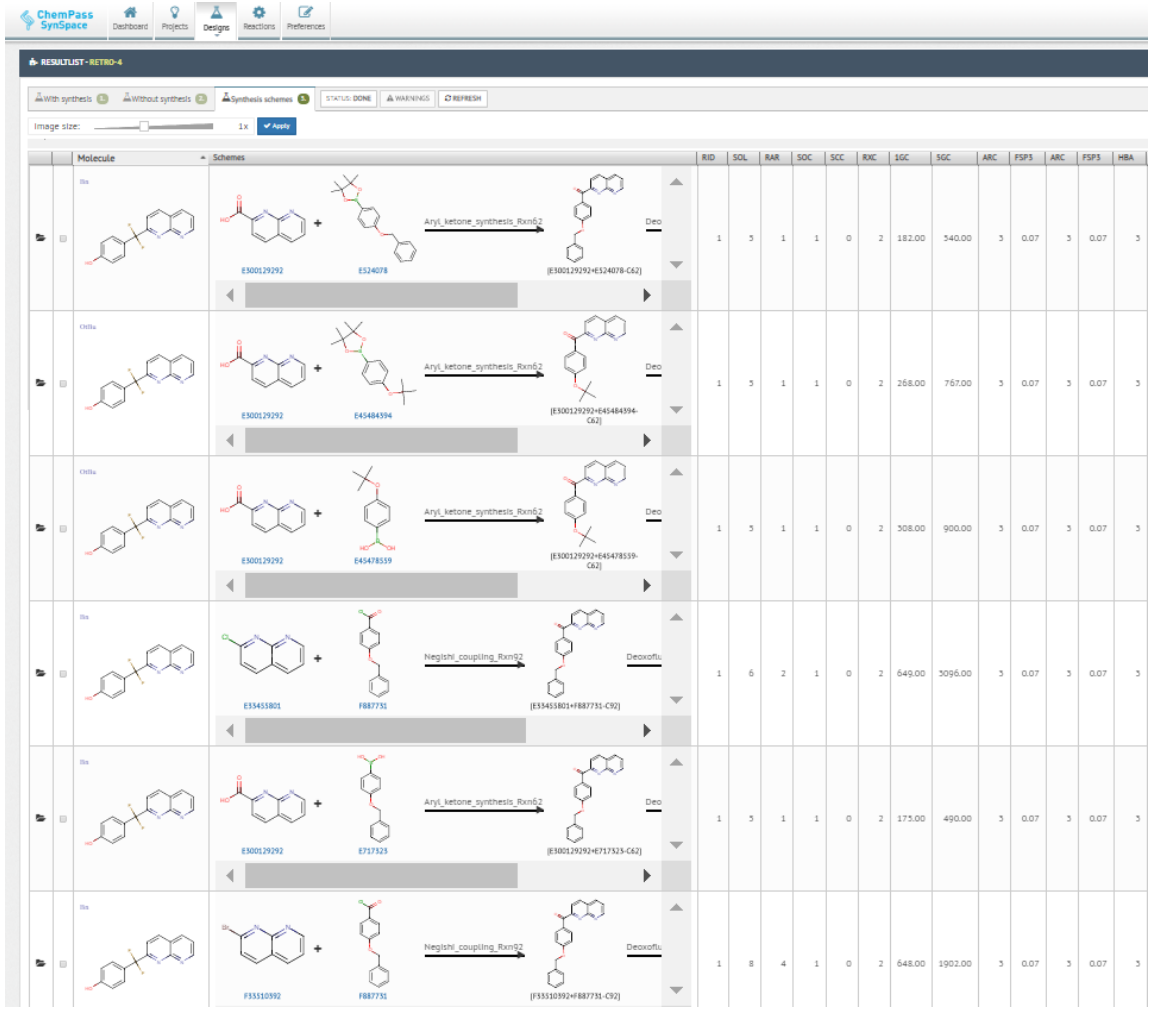


Retrosynthesis example 4

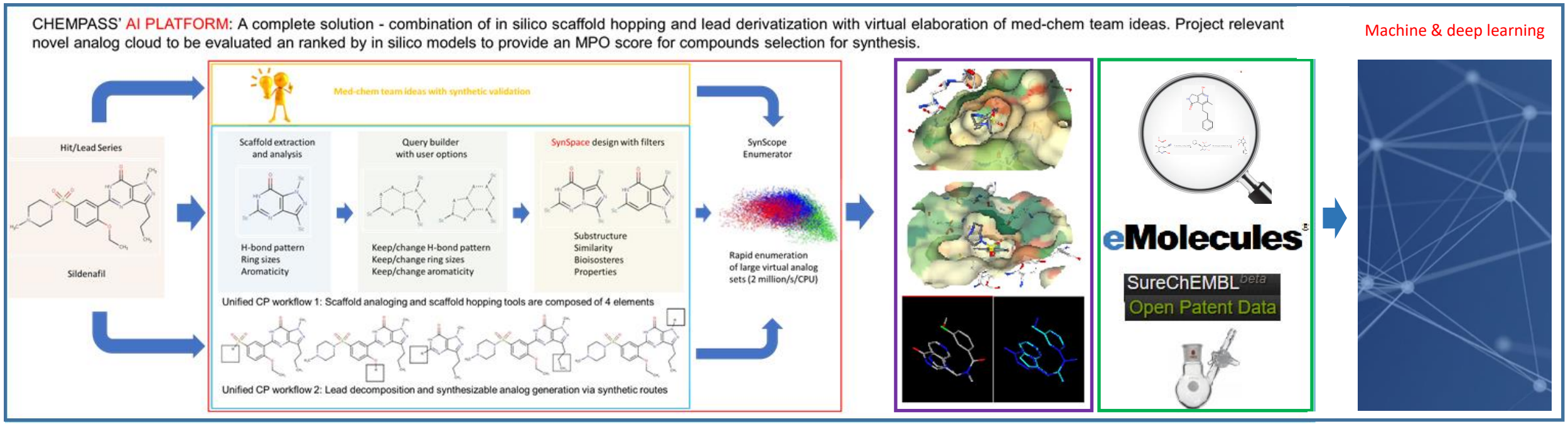




Retrosynthesis example 4



Complete AI-driven lead optimization platform

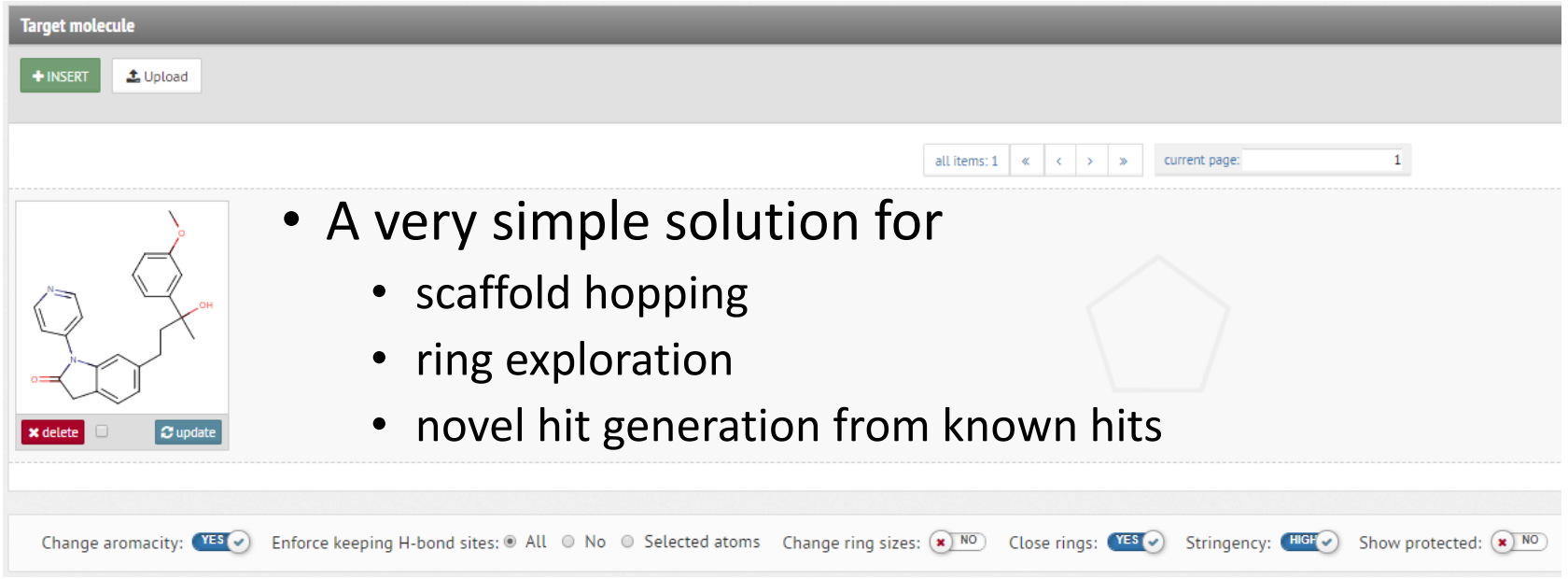


AI-powered ideation: scaffold hopping, substituent designs
Property predictions
Potency, selectivity predictions: semi-automated protein or ligand-based methods

Full synthesis and reagent availability assessment
Novelty assessment
Synthesis outsourcing

MPO scores
Ranking
Selection

Scaffold analog design/hit generation with SynSpace

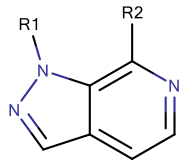


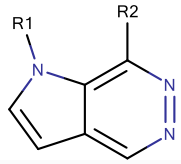
- A very simple solution for
  - scaffold hopping
  - ring exploration
  - novel hit generation from known hits

- Fully automated design of scaffold analogs containing the desired side-chains
- Intelligent ring closing mechanisms included
- Intelligent explorations of alternative ring sizes with same side-chain vectors

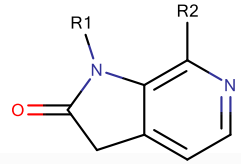
Automated scaffold design options

- Scaffold aromaticity:



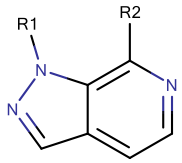


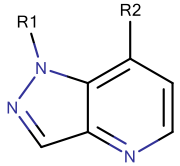
Change aromacity: NO



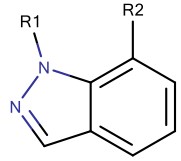
Change aromacity: YES

- Scaffold H-bonding:



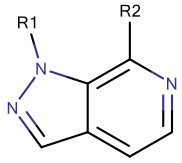


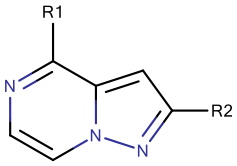
Enforce keeping H-bond sites: All No Selected atoms



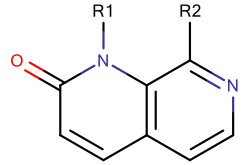
Enforce keeping H-bond sites: All No Selected atoms

- Scaffold ring size:





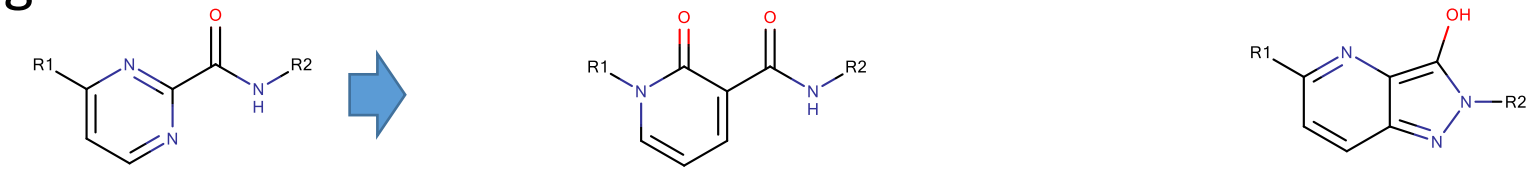
Change ring sizes: NO



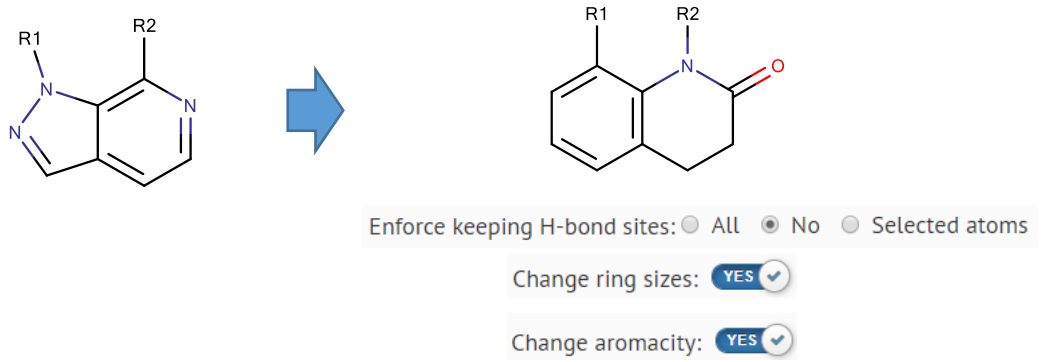
Change ring sizes: YES

Automated scaffold design options

- Close new rings:

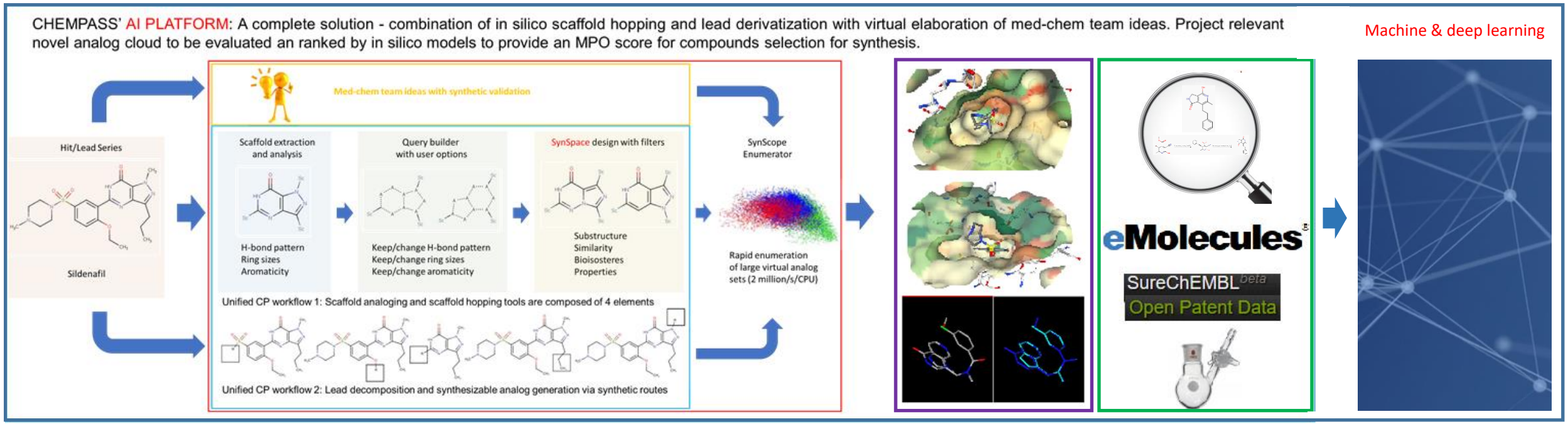


- Major change example: Side-chains and their vector remain similar



Lost both ring H-bond acceptors
Modified ring size
Modified aromaticity

Complete AI-driven lead optimization platform



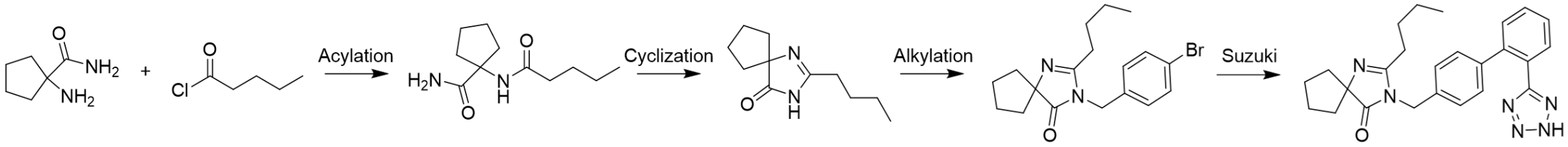
AI-powered ideation: scaffold hopping, substituent designs
Property predictions
Potency, selectivity predictions: semi-automated protein or ligand-based methods

Full synthesis and reagent availability assessment
Novelty assessment
Synthesis outsourcing

MPO scores
Ranking
Selection

Derivatization design: side-chain, connectivity and scaffold design with SynSpace API

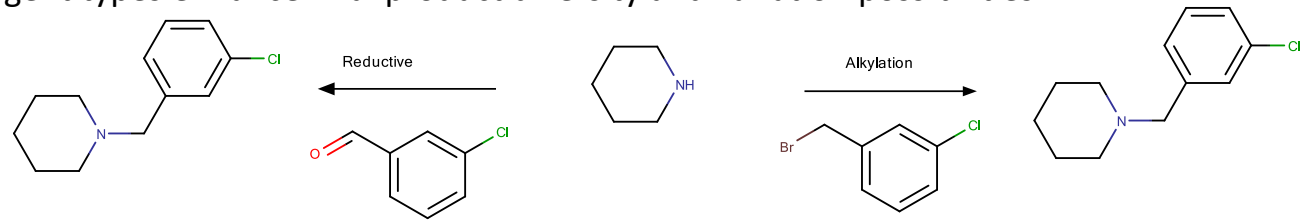
- User settings needed:
  - How many total variations per compound
  - How many total compounds
  - Minimum similarity
- Derivatization of side-chains via synthetic routes: Irbesartan analog design via a 4-step synthetic sequence



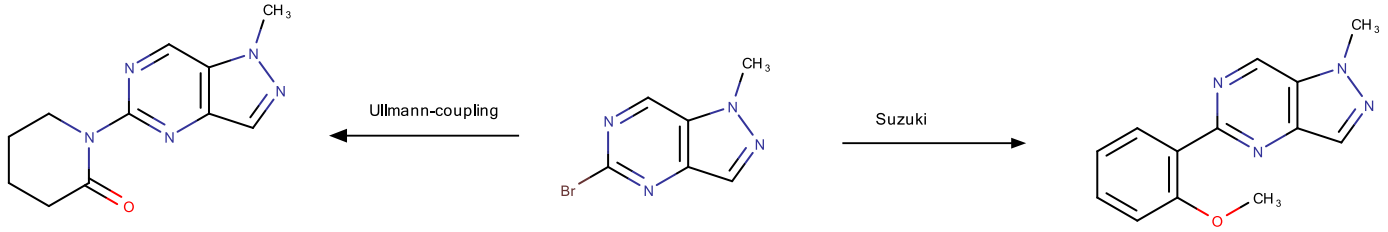
- Analog selection by multiple similarity methods, bioisosteres and substructures:
  - >90K analogs generated before property filtration

Countless variation possibilities: related reactions

- Each synthetic step has 2 types of „related reactions”
  - Different reaction(s) that can lead to the same product type using different reagent classes (i.e. alkylation with halides or reductive amination)
    - Different reagent types enhance final product diversity and variation possibilities

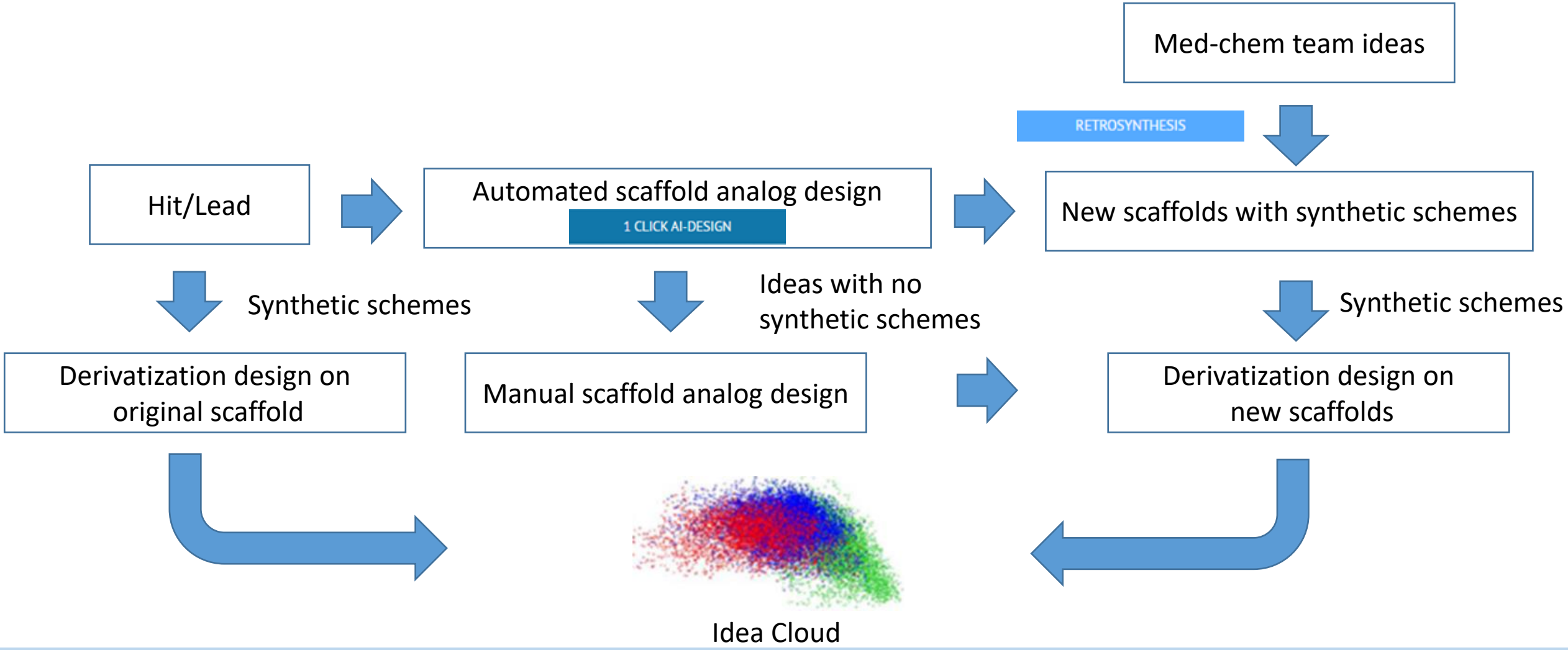


  - Different reaction(s) that use functional groups found at least in one of the reagents and can produce a similar product motif

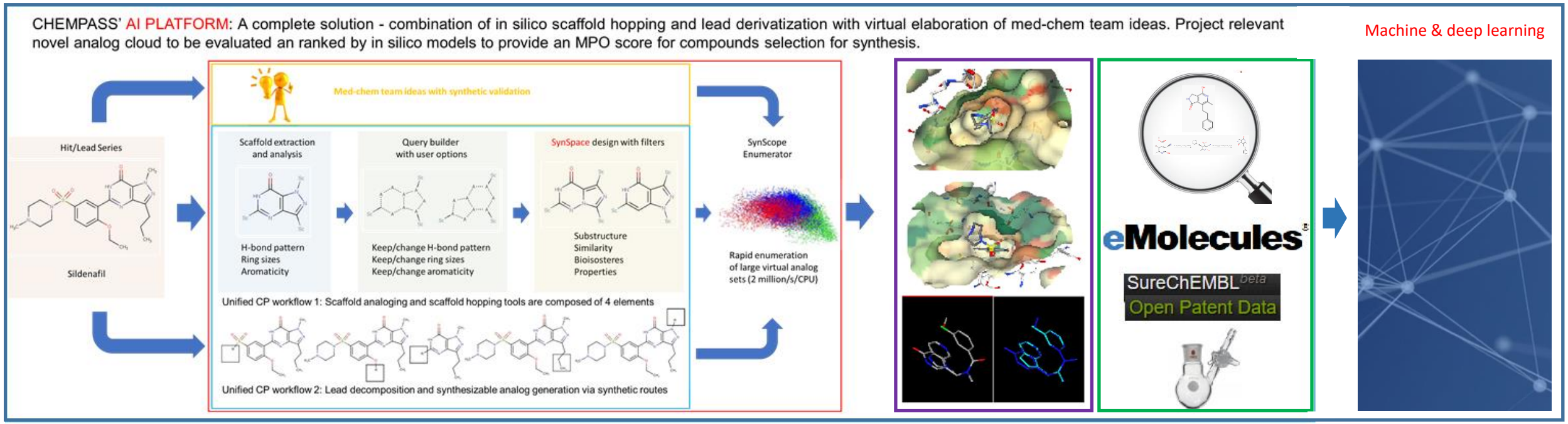


- The above synthetic knowledge is naturally encoded in our synthetic know-how

Complete design cycle



Complete AI-driven lead optimization platform

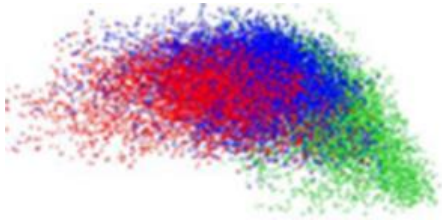


AI-powered ideation: scaffold hopping, substituent designs
Property predictions
Potency, selectivity predictions: semi-automated protein or ligand-based methods

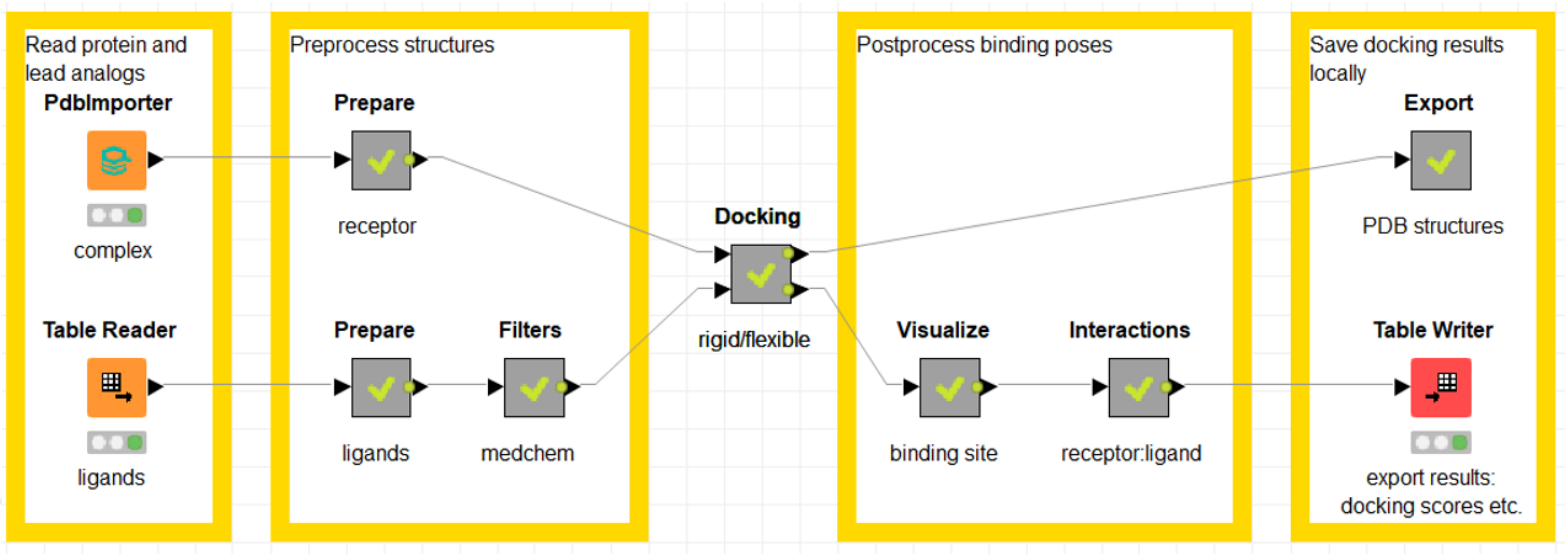
Full synthesis and reagent availability assessment
Novelty assessment
Synthesis outsourcing

MPO scores
Ranking
Selection

Binding predictions



Idea Cloud



Output:

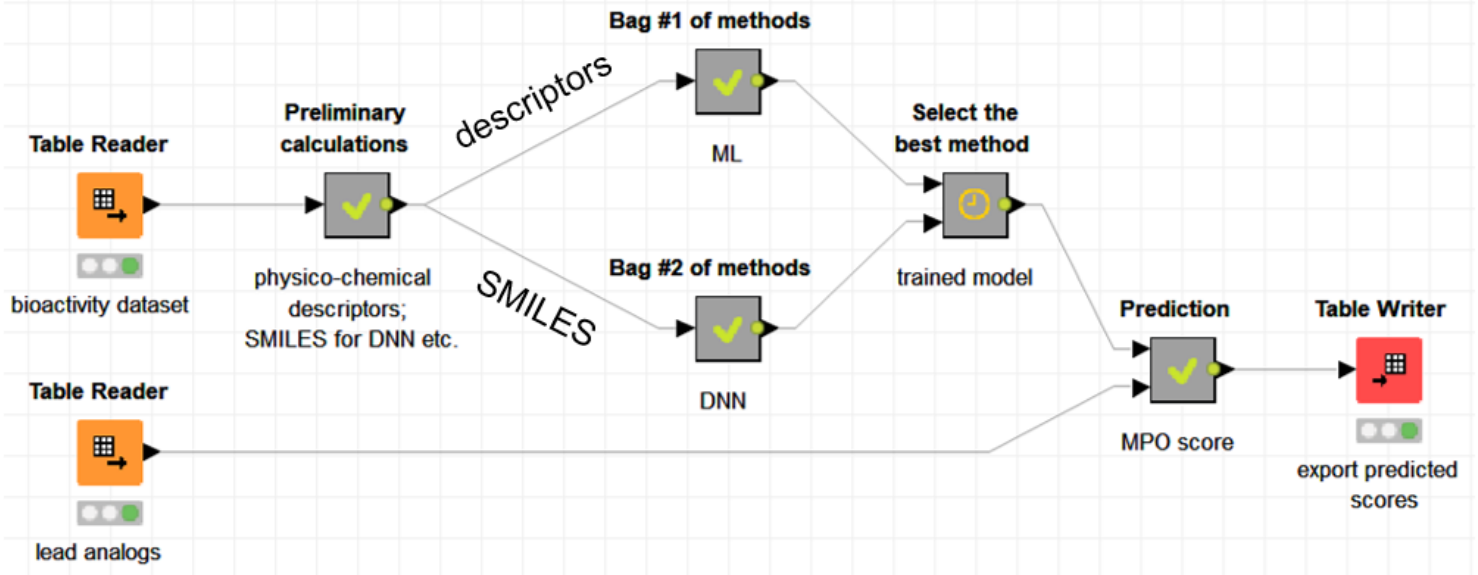
- Docking pose
- Docking interactions
- Docking scores
- Selectivity assessment

Built-in

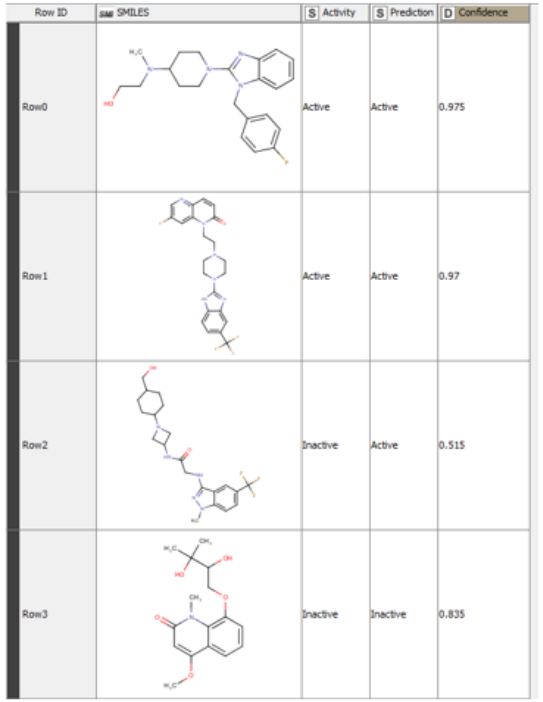
- SuCOS for pose analysis
- Clash analysis
- Interaction analysis

ML and deep learing models for an MPO score

Predictive models trained with various ML and DNN methods using a mix of publicly available and proprietary datasets.



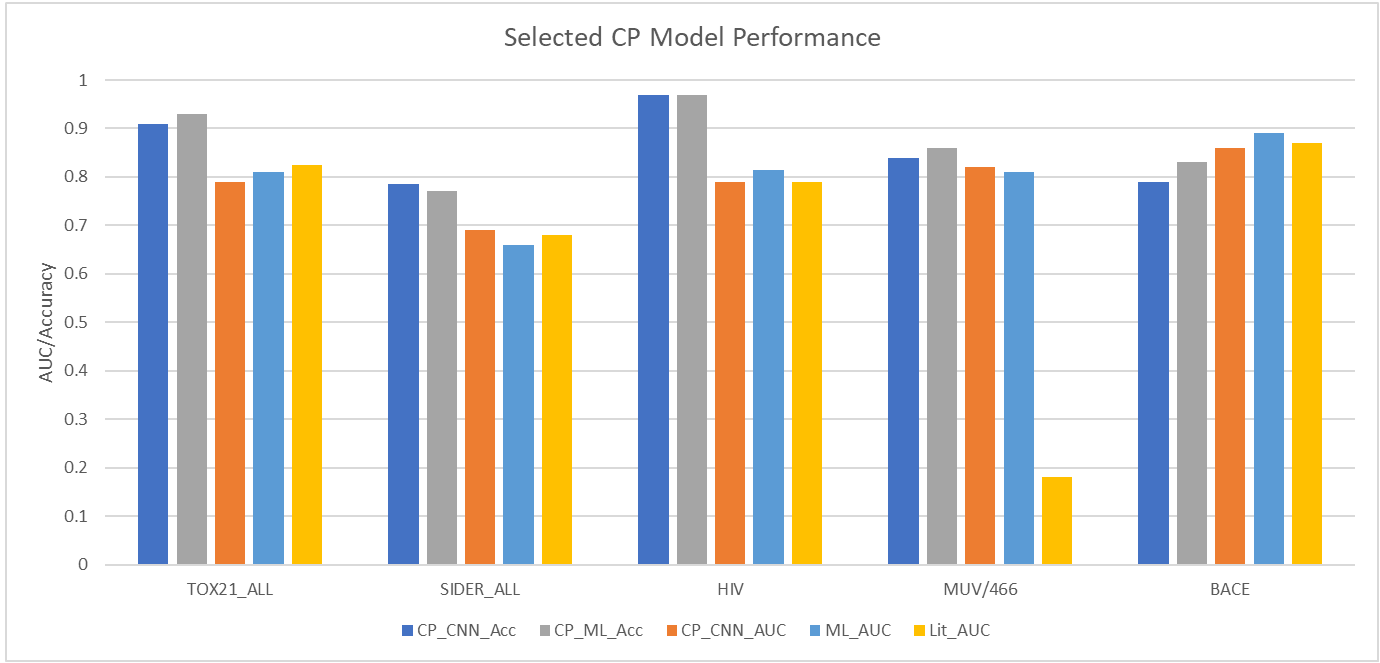
Models are automatically re-trained in each design cycle



Training set data

ChemPass predictive AI and ML models

- Several deep learning and machine learning (ML) methods automatically refined in each optimization cycle
- Deep learning methods with smiles or property inputs
  - ChemPass' smiles-based CNN appears to be string-independent
- Comparison to literature models
  - M. Hirohara et al. BMC Bioinform. 19, 526 (2018)
  - Z. Wu, et al. Chem. Sci. 9, 513 (2018)
  - **CP models perform well**



ML methods tested: Tree Ensemble, Random Forest
ML_acc: obtained in a 5-fold validation with optimal hyperparameters (number of models, maximum number of levels) to maximize the accuracy
Tox21: average of 12 sub-datasets
SIDER: average of 27 sub-datasets

A typical use of the AI-driven lead optimization platform

- ChemPass
  - design, in silico profiling (traditional and ML methods)
  - ranking of newly designed analogs
  - management of chemistry and biology CROs (if requested)

- Joint selection with project owner partner of compounds for synthesis
- Objectives:
  - H2L in 6 months or
  - Hit generation and H2L in 8 months
  - Lead to advanced lead in 8-12 months

ChemPass

Current lead optimization tasks in the design-make process

Ideation

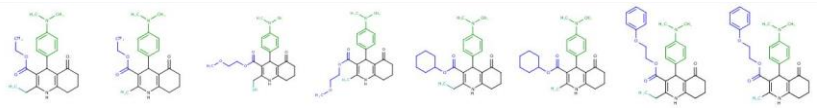
SciFinder® A CAS SOLUTION

Literature check for prior art

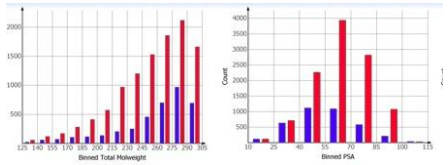
WIPO PATENTSCOPE

Idea sharing

Focused library enumeration

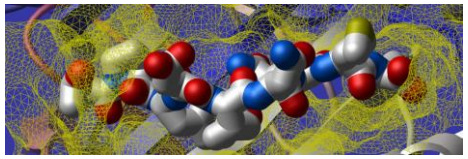


Property calculations



Predictions

Modelling, comp-chem



Selection

Synthesis planning

Reaxys®

Improved synthetic planning with Reaxys

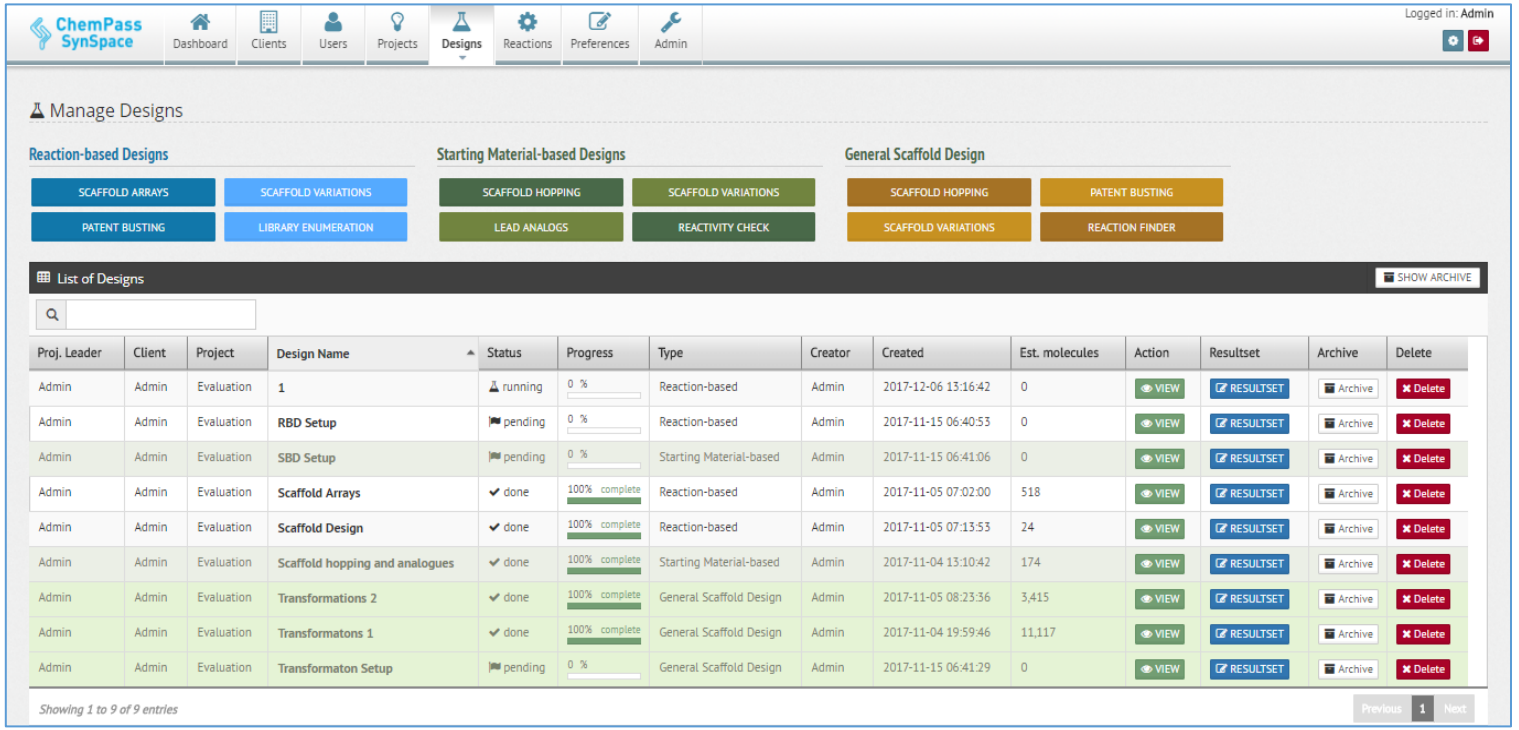
Reagent search

eMolecules®

Lead optimization with SynSpace AI-assisted platform

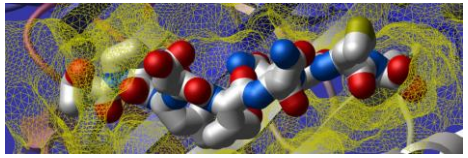
ChemPass SynSpace

Ideation Literature check for prior art
Idea sharing Focused library enumeration
Property calculations Synthesis planning Reagent search



ChemPass AI Platform

Modelling, comp-chem



Predictions

Selection

ChemPass

Thank you for your attention!