

Al-assisted lead optimization with SynSpace Greg Makara ChemPass Ltd.



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 Alpowered drug discovery
- Built an Al-powered lead discovery platform
- SynSpace availability



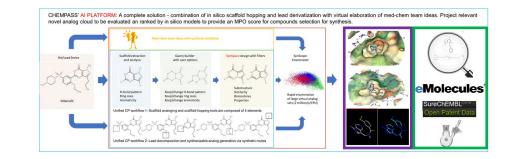
• Aims: reduce cycle time, optimization time, expand idea and IP space



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>





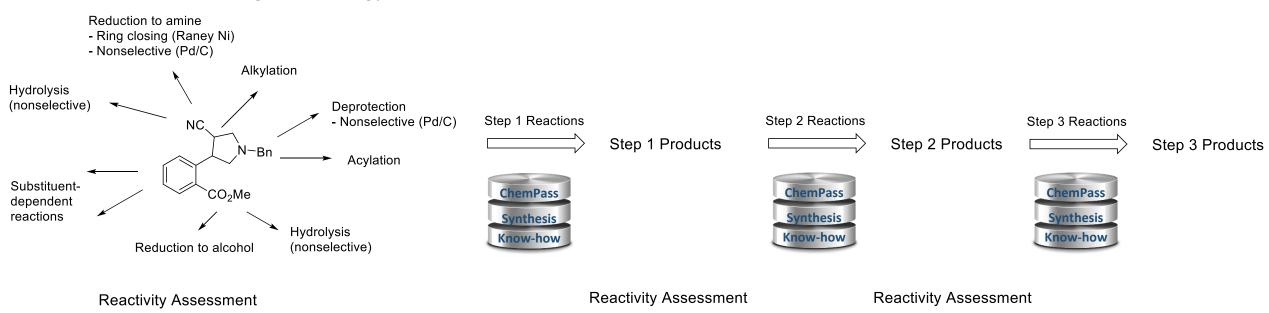
 ChemPass has AI-based technologies and software for both <u>design</u> of novel leads or lead analogs AND <u>retrosynthesis</u>

A Manage Designs					
Reaction-based Designs	Starting Material-based Designs	Enumerations	General Scaffold Design	Automated Modules	
SCAFFOLD ARRAYS	SCAFFOLD HOPPING	LIBRARY ENUMERATION	ENHANCED GSD	1-CLICK SCAFFOLD DESIGN	
PATENT BUSTING	LEAD AND SIDE-CHAIN ANALOGS		SCAFFOLD HOPPING	RETROSYNTHESIS	



ChemPass design concept

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



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

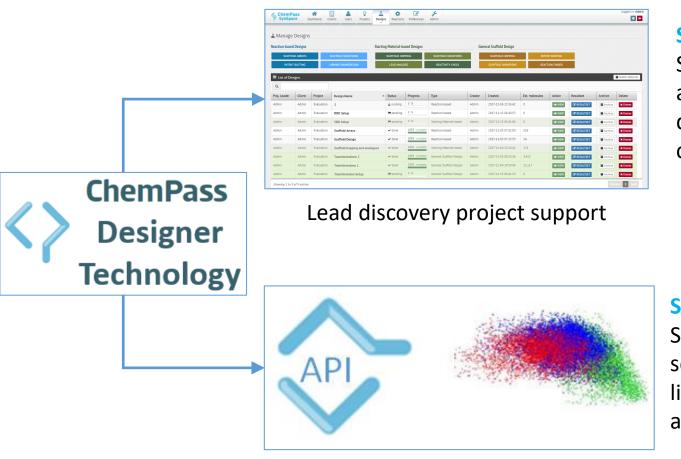
- Software license:
 - SynSpace for medicinal chemists and computational chemists
 - SynSpace API for computational chemists
 - SynSpace modules
 - Library design (e.g. DEL)
 - Side-chain design
 - Scaffold and lead analog design modules (new fully automated AI modules)
 - Retrosynthesis module (new)

Reaction-based Designs	Starting Material-based Designs	Enumerations	General Scaffold Design	Automated Modules
SCAFFOLD ARRAYS	SCAFFOLD HOPPING	LIBRARY ENUMERATION	ENHANCED GSD	1-CLICK SCAFFOLD DESIGN
PATENT BUSTING	LEAD AND SIDE-CHAIN ANALOGS		SCAFFOLD HOPPING	RETROSYNTHESIS

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



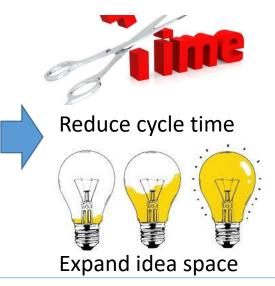
SynSpace software availability and uses



Machine learning project support

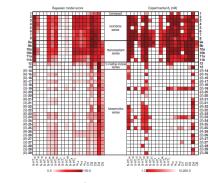
SYNSPACE Software

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



SYNSPACE API

Synthetically enabled scaffold and vast virtual library design from available reagents



Deep learning, AIpowered drug discovery



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

Pa	artners:
•	Big Pharma
•	Big Biotech

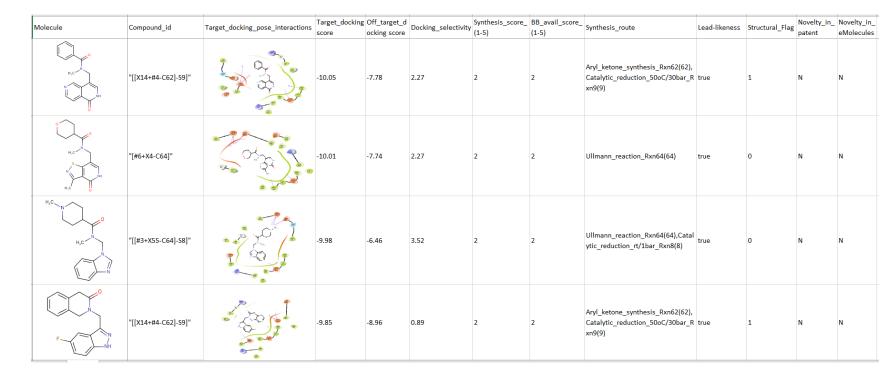
• Small biotech



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



Extracted from real report with altered structures

Report format and data types are fully customizable

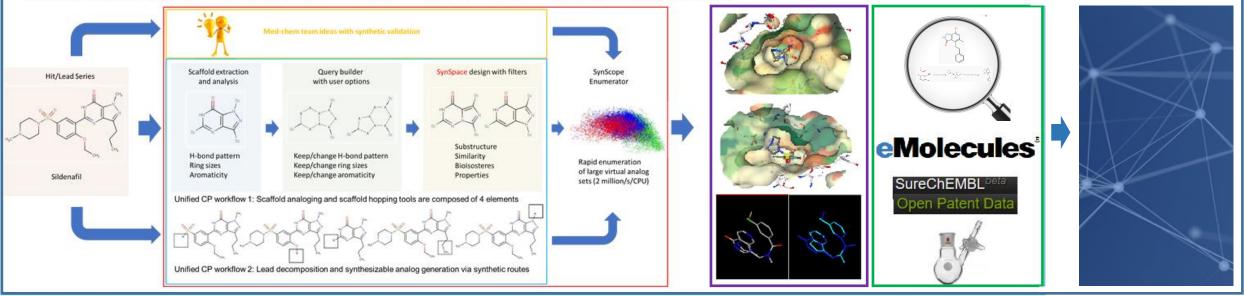
A detailed table is also available with synthesis scheme, reagent structures, ids, cost etc., hundreds of customizable properties



Machine & deep learning

Complete Al-driven lead optimization platform

CHEMPASS' AI PLATFORM: A complete solution - combination of in silico scaffold hopping and lead derivatization with virtual elaboration of med-chem team ideas. Project relevant novel analog cloud to be evaluated an ranked by in silico models to provide an MPO score for compounds selection for synthesis.



Al-powered ideation: scaffold hopping, substituent designs Property predictions

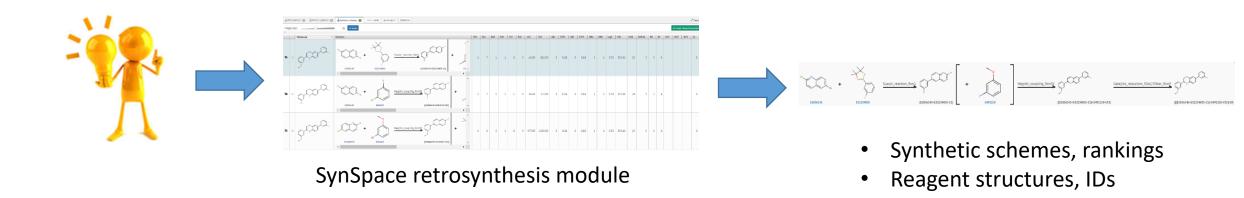
Potency, selectivity predictions: semi-automated protein or ligand-based methods

Full synthesis and reagent availability assessment Novelty assessment Synthesis outsourcing

MPO scores Ranking Selection



Med-chem team ideas



• Retrosynthesis enables human ideas to enter the design workflow



Retrosynthesis Planner (improved speed): done!

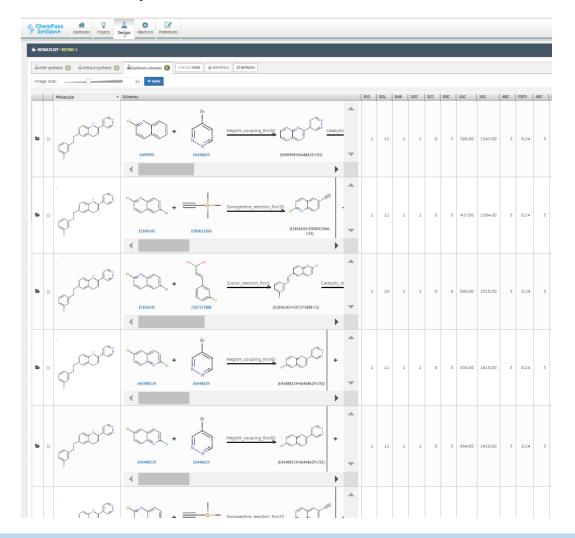
	ted to the old depth-first search plar		https://rxn.res.ibm.com/rxn/proj	ects/5db6a652bc67d20001772d	74/Retro- 80% •••	☑ ☆ Q Search	III\ 🗊 🧕	
Target compound:	Fc1cccc(CCc2ccc3c(c2)CCC(c2cc	nnc2)N3)c1 / Draw	Retro-1 < Retrosynthesis outcome					← Back to workspace
Parsed structure:) Confidence: 0.641 Optimization score: 0.821	Medium confidence		Q	Q 100% -	Add Sequence to collection
	₽ _		Predictions: B 2/F 27	Optimisation time: 37 s Molecule cor	nmercially available on eMolecules.com	Not able to find a synthetic path	Sequences Generated	^
Quick settings:		Default settings		âm			📀 Sequence 0 🛕	M PDF 🔽
		I don't mind seeing unrealistic suggestions					Parameters	
		I don't mind waiting slightly longer		Fc1cccc(CCC2=CC3CCC(c4ccnnc 4)NC3C=C2)c1			MSSR	15
		Just give me <i>something</i> to look at		0.641 - N-Boc deprotection Q			MRP	50
							FAP	0.65
Expansion settings:	Max. depth (1-9): 🔮	6 Max. branching factor: 0 25	Â	I F			SbP	3
	Min. retro template count: 🌖	0 Expansion time (s):			н		Help us improve!	Send us your feedback.
	Num. templates: 0	1000	od o	F O			What do you think abou	-
	Max cum. prob: 6	0.9999	CC(C)(C)DD(=0))N1C(c2conne2)C CC2C+0(CCCCcccco(P)=3)C+CC21	CICCI 0=C(0)C(F)(F)F				
Stop criteria:	Maximum chemical price (\$/g): 6	1500	a - Undefined				It's correct! 🙂	It's not so good! 🙂
	Chemical property logic: 🜖	None (price only)						
	Chemical popularity logic: 🜖	None (use price/properties)						
Evaluation settings:	Min. plausibility: 🜖	0.01	TO STO					
	Manual forward prediction: ()	Template-free 🗸				>		
Controls:	Start	Return as soon as any pathway found 🕥			© IBM RXN			

After expanding (with 0 banned reactions, 0 banned chemicals), 744 total chemicals and 3565 total reactions

No trees resulting in buyable chemicals found! If the program is having trouble with your target, you may want to explore the One-Step Retrosynthesis options and help guide the search.

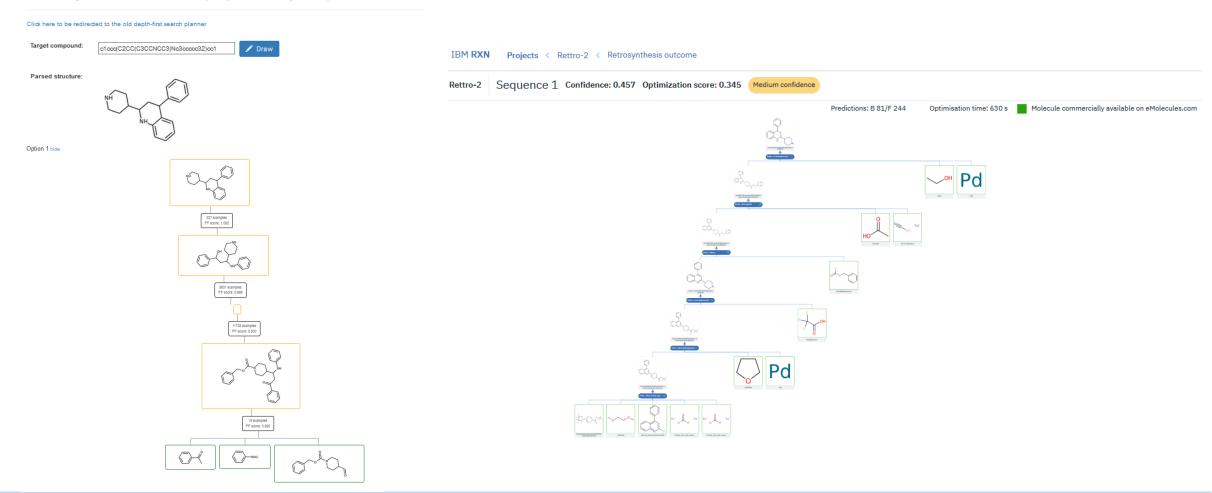
Alternatively, please use the "Chemical popularity logic" setting above to give a more generous termination criterion to the search. Our buyables database is very sparse, so many pathway



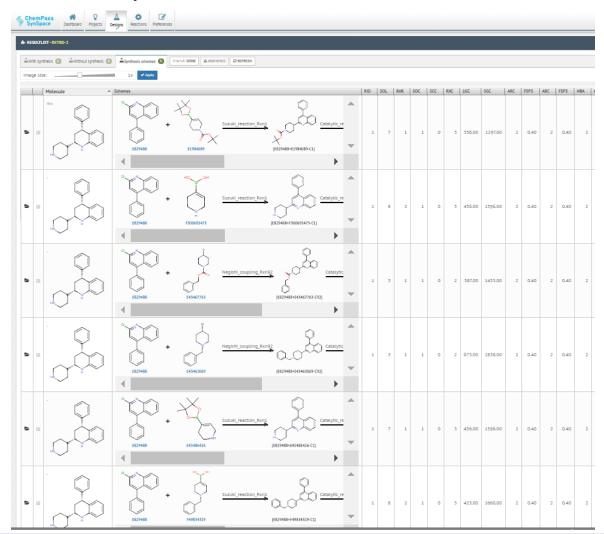




Retrosynthesis Planner (improved speed): done!





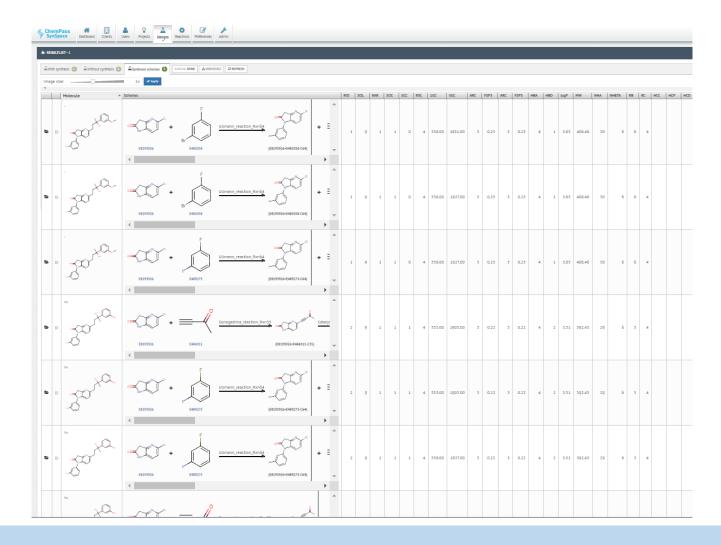




Retrosynthesis Planner (improved speed): done!

Click here to be redirecte	d to the old depth-first search plan	ner	C 🕜 🕒 https://rxn.res.ibm.com/rxn/projects/5dc13664bc67d2000177dd34/Retrosynthesis/retrosynthesis/5dc13c2ebc67d2000177e250 … 🗵 🏠 🔍 ledocl
Target compound:	cc(C(C)(0)CCc2ccc3c(n2)CC(=	D)N3c2cccc(F)c2)c1	XN Projects < Retrosynthesis < Retrosynthesis outcome
Parsed structure:	ОН	0	nthesis Sequence 0.928 Confidence: 0.449 Optimization score: 0.928 High confidence
		F F	Molecule commercially available on eMolecules.com Not able to fin
Quick settings:		Default settings	
		I don't mind seeing unrealistic suggestions	Mg ^{Br}
		I don't mind waiting slightly longer	
		Just give me something to look at	(Left investigation
Expansion settings:	Max. depth (1-9): 3	6 Max. branching factor: 3 25	
	Min. retro template count: 🟮	0 Expansion time (s): 0 120	
	Num. templates: 3	1000	Differences
	Max cum. prob: 😉	0.9999	
Stop criteria:	Maximum chemical price (\$/g): 3	1500	HH Pd
	Chemical property logic: 🕚	None (price only)	
	Chemical popularity logic: 🕄	None (use price/properties) 🗸	
Evaluation settings:	Min. plausibility: 🟮	0.01	
	Manual forward prediction: 4	Template-free 🧹	
Controls:	Start	Return as soon as any pathway found 3	
After expanding (with 0	banned reactions, 0 banned chemi	cals), 537 total chemicals and 3914 total reactions	
No trees resulting in buy	able chemicals found! If the program	n is having trouble with your target, you may want to explore the One-Step Retrosynthesis options and help guide the search.	
Alternatively, please use	the "Chemical popularity logic" sett	ng above to give a more generous termination criterion to the search. Our buyables database is very sparse, so many pathways n	



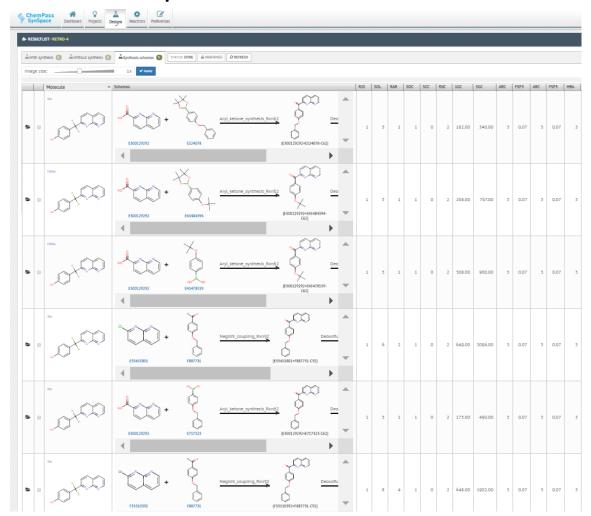




Retrosynthesis Planner (improved speed): done!

Click here to be redirected to the old depth-first search planner Target compound: Oct ccc(C(F)(F)c2ccc3cccnc3n2)cc1	Image: Solution of the second state of the second stat	Q Search
	Ce () Confidence: 0.916 Optimization score: 1.13 High confidence	Q
	Predictions: B 37/F 121 Optimisation time: 282 s Molecule commercially available on eMolecules.com Not able t	o find a synthetic path
XY ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~		
		~



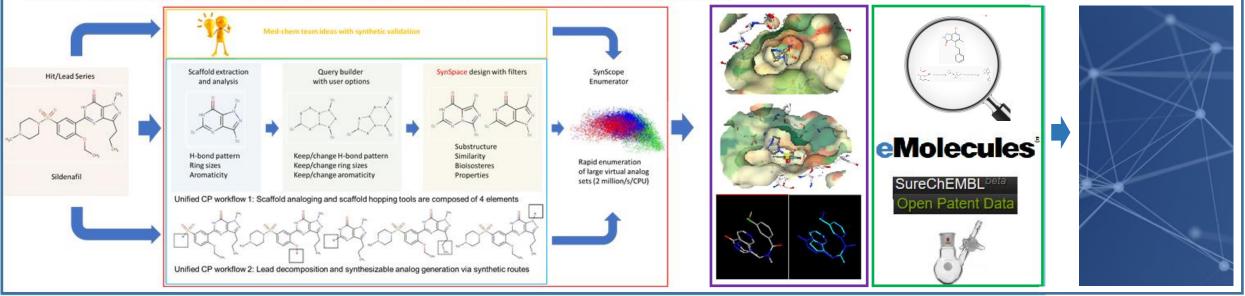




Machine & deep learning

Complete Al-driven lead optimization platform

CHEMPASS' AI PLATFORM: A complete solution - combination of in silico scaffold hopping and lead derivatization with virtual elaboration of med-chem team ideas. Project relevant novel analog cloud to be evaluated an ranked by in silico models to provide an MPO score for compounds selection for synthesis.



Al-powered ideation: scaffold hopping, substituent designs Property predictions

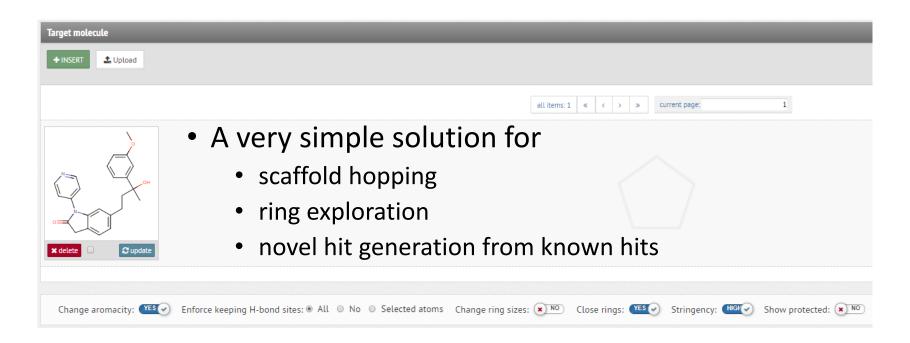
Potency, selectivity predictions: semi-automated protein or ligand-based methods

Full synthesis and reagent availability assessment Novelty assessment Synthesis outsourcing

MPO scores Ranking Selection



Scaffold analog design/hit generation with SynSpace

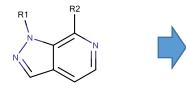


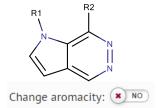
- Fully automated design of scaffold analogs containing the desired side-chains
- Intelligent ring closing mechanisms included
- Intelligent explorations of alternative ring sizes with same side-chain vectors

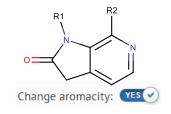


Automated scaffold design options

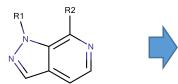
• Scaffold aromaticity:

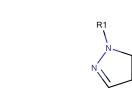


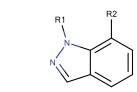




• Scaffold H-bonding:

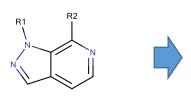


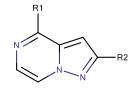




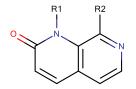
Enforce keeping H-bond sites: O All O No O Selected atoms Enforce keeping H-bond sites: O All O No O Selected atoms

• Scaffold ring size:





Change ring sizes: 💌 🔊



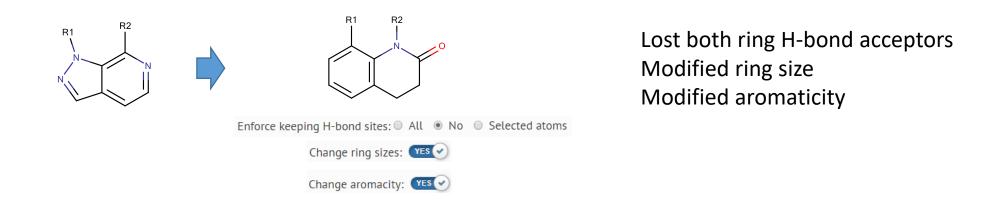




Automated scaffold design options



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

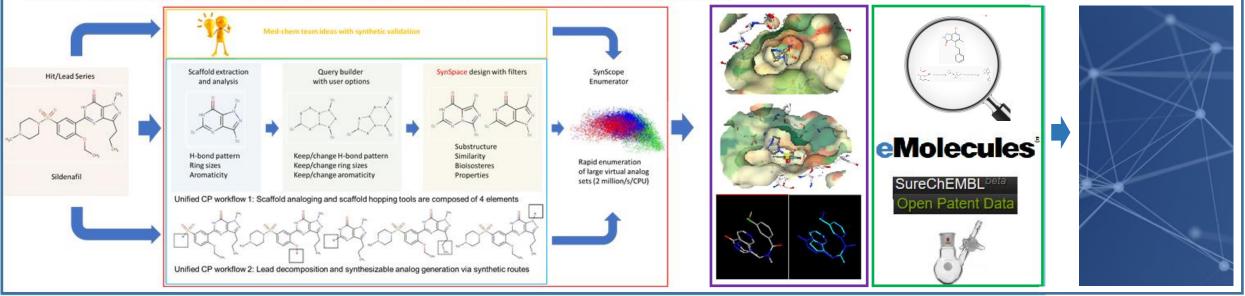




Machine & deep learning

Complete Al-driven lead optimization platform

CHEMPASS' AI PLATFORM: A complete solution - combination of in silico scaffold hopping and lead derivatization with virtual elaboration of med-chem team ideas. Project relevant novel analog cloud to be evaluated an ranked by in silico models to provide an MPO score for compounds selection for synthesis.



Al-powered ideation: scaffold hopping, substituent designs Property predictions

Potency, selectivity predictions: semi-automated protein or ligand-based methods

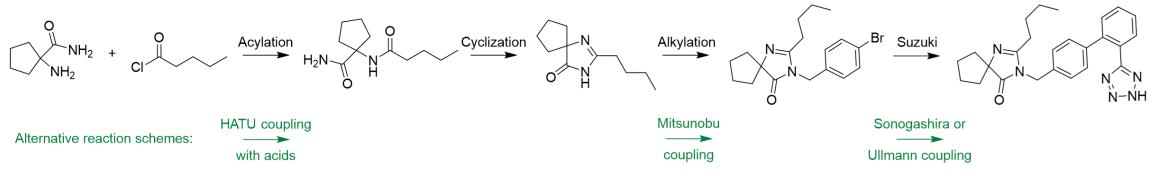
Full synthesis and reagent availability assessment Novelty assessment Synthesis outsourcing

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

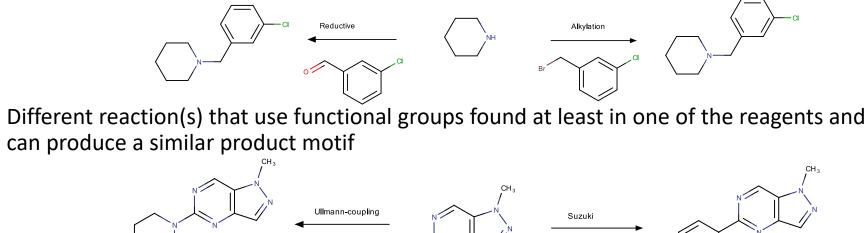


- Analog selection by multiple similarity methods, bioisosteres and substructures:
 - >90K analogs generated before property filtration



Countless variation possibilities: related reactions

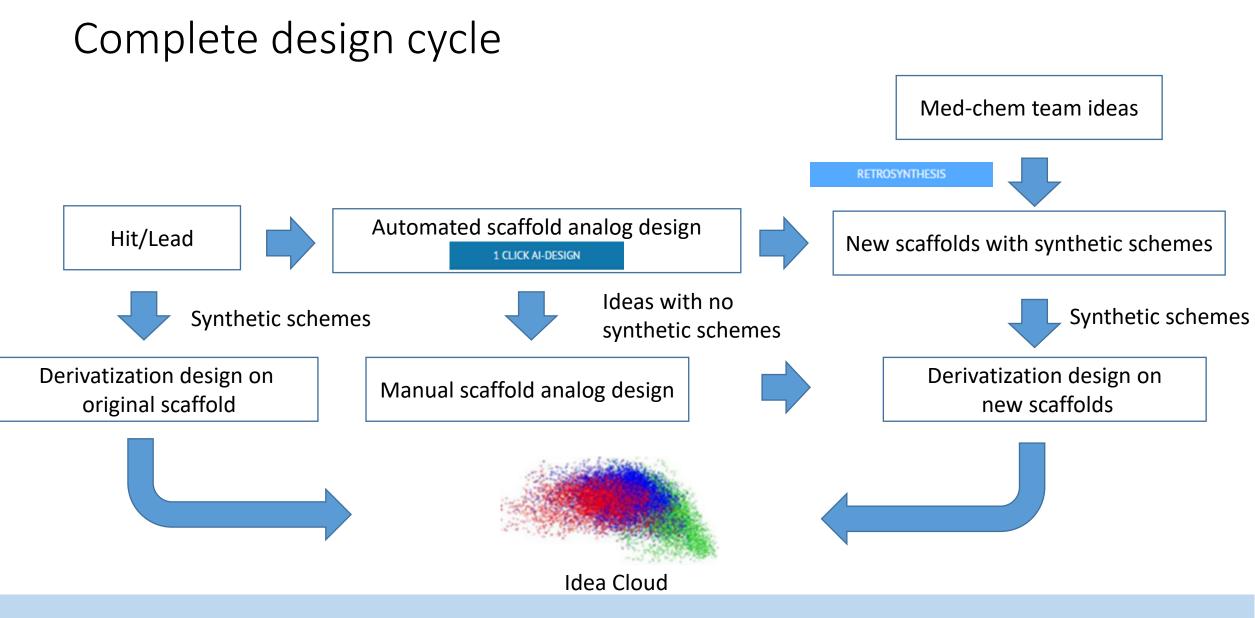
- Each synthetic step has 2 types of "related reactions"
 - Different reaction(s) that can lead to the same product type using different reagent classes (i.e. alkylation with halides or reductive amination)
 - Different reagent types enhance final product diversity and variation possibilities



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

•



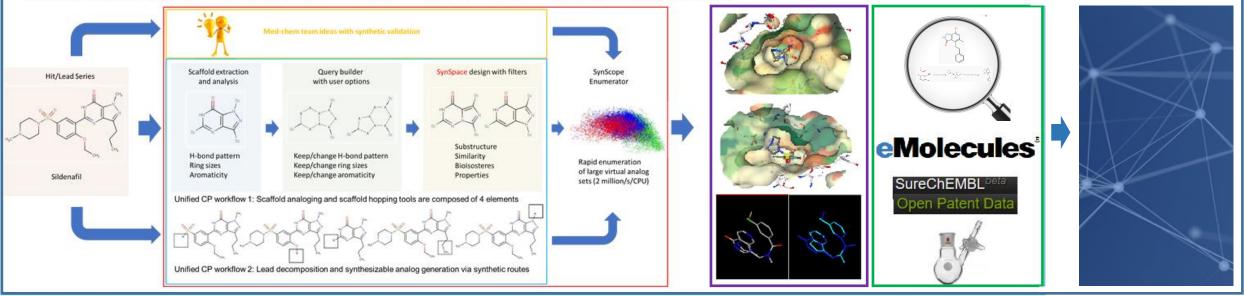




Machine & deep learning

Complete Al-driven lead optimization platform

CHEMPASS' AI PLATFORM: A complete solution - combination of in silico scaffold hopping and lead derivatization with virtual elaboration of med-chem team ideas. Project relevant novel analog cloud to be evaluated an ranked by in silico models to provide an MPO score for compounds selection for synthesis.



Al-powered ideation: scaffold hopping, substituent designs Property predictions

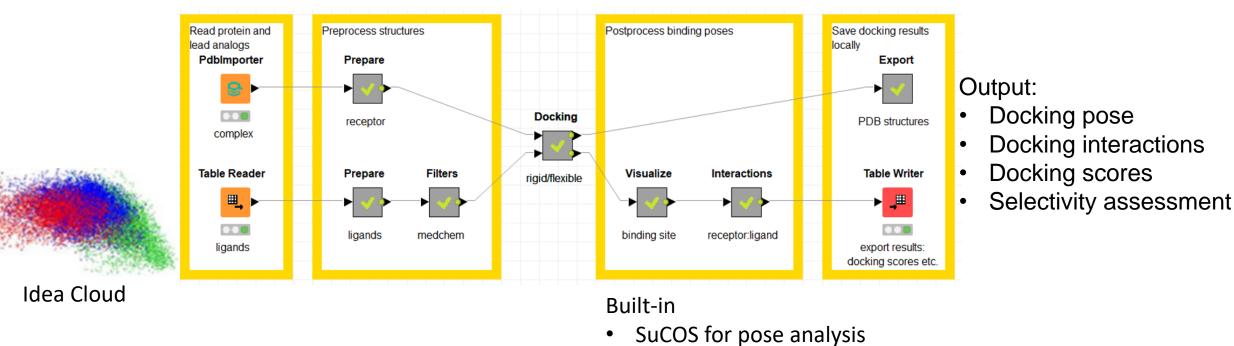
Potency, selectivity predictions: semi-automated protein or ligand-based methods

Full synthesis and reagent availability assessment Novelty assessment Synthesis outsourcing

MPO scores Ranking Selection



Binding predictions



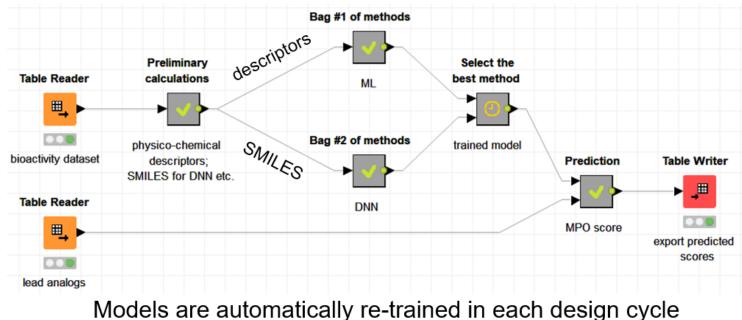
- Clash analysis
- Interaction analysis

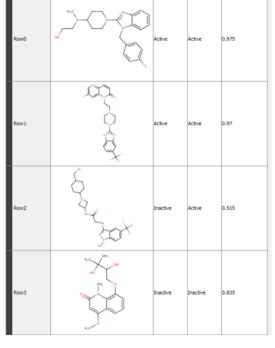
•



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.





S Activity S Prediction D Confidence

Row ID

SMI SMILES

Training set data



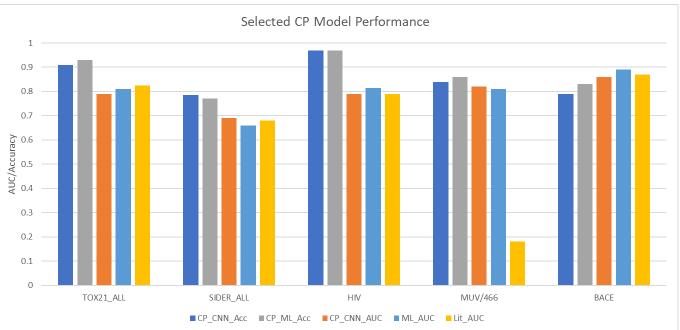
ChemPass predictive AI and ML models

- Several <u>deep learning and machine learning (ML)</u> methods automatically <u>refined in each</u> 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)
 - <u>CP models perform well</u>

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



Current lead optimization tasks in the design-make process

Ideation



Literature check for prior art



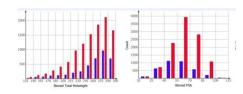
Idea sharing



Focused library enumeration

教教~教~教教教教

Property calculations



Predictions



Modelling, comp-chem





Synthesis planning

Reaxys®

Reagent search eMolecules[®]

Improved synthetic planning with Reaxys



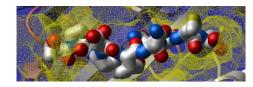
Lead optimization with SynSpace AI-assisted platform

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

Manage	Docigne												
action-based		>	Si	tarting Material-I	ased Designs		Ger	neral Scaffold Design					
SCAFFO	LD ARRAYS		SCAFFOLD VARIATIONS	SCAFFOLD H	OPPING	SCAFFOLD VARIATIONS		SCAFFOLD HOPPING	PATE	IT BUSTING			
PATENT	BUSTING		IBRARY ENUMERATION	LEAD ANAI	.0GS	REACTIVITY CHECK		SCAFFOLD VARIATIONS	REACT	ION FINDER			
List of Des	signs												SHOW ARCH
Q	agna												
Proj. Leader	Client	Project	Design Name	▲ Status	Progress	Туре	Creator	Created	Est. molecules	Action	Resultset	Archive	Delete
Admin	Admin	Evaluation	1	∆ running	0 %	Reaction-based	Admin	2017-12-06 13:16:42	0	VIEW	RESULTSET	Archive	¥ Delete
Admin	Admin	Evaluation	RBD Setup	Pending	0 %	Reaction-based	Admin	2017-11-15 06:40:53	0	VIEW	RESULTSET	Archive	¥ Delete
Admin	Admin	Evaluation	SBD Setup	Pending	0 %	Starting Material-based	Admin	2017-11-15 06:41:06	0	👁 VIEW	RESULTSET	Archive	X Delete
Admin	Admin	Evaluation	Scaffold Arrays	🗸 done	100% complete	Reaction-based	Admin	2017-11-05 07:02:00	518	👁 VIEW	RESULTSET	Archive	× Delete
Admin	Admin	Evaluation	Scaffold Design	🗸 done	100% complete	Reaction-based	Admin	2017-11-05 07:13:53	24	👁 VIEW	RESULTSET	Archive	X Delete
Admin	Admin	Evaluation	Scaffold hopping and analogue	es 🗸 done	100% complete	Starting Material-based	Admin	2017-11-04 13:10:42	174	👁 VIEW	RESULTSET	Archive	X Delete
Admin	Admin	Evaluation	Transformations 2	🗸 done	100% complete	General Scaffold Design	Admin	2017-11-05 08:23:36	3,415	I VIEW	RESULTSET	Archive	X Delete
Admin	Admin	Evaluation	Transformatons 1	🗸 done	100% complete	General Scaffold Design	Admin	2017-11-04 19:59:46	11,117	I VIEW	RESULTSET	Archive	X Delete
Admin	Admin	Evaluation	Transformaton Setup	🔎 pending	0 %	General Scaffold Design	Admin	2017-11-15 06:41:29	0	I VIEW	RESULTSET	Archive	X Delete



Modelling, comp-chem



Predictions







Thank you for your attention!