Escape from ‘availability bias’ in compound design

ANDRÁS STRÁCZ, ÁKOS TARCSAY, IVÁN SOLT, GÁBOR IMRE

ABSTRACT Small molecule design is an information demanding activity, since all relevant knowledge is to be accessible within a single space and requires synchronized application of computational models to assist decision making on synthesis candidates. Our study aims to evaluate a software platform coping with this complexity (Marvin Live). The tool provides central management of innovative ideas and a framework that helps triage these. Decision support is based on predicted properties that span phys-chem descriptors, combined metrics like CNS MPO score, 3D overlay and modelling results conducted with KNIME on the one hand, and cross-checking with available knowledge collected from a variety of sources to do rapid freedom to operate analysis and SAR by catalog by ultra-fast searching of patent and compound catalog databases, respectively on the other. This hypothetic study shows the potential evolution of a compound idea from the reference compound to a synthesis candidate through an example discovery project of 5-hydroxytryptamine receptor 6 (HTR6) ligands.

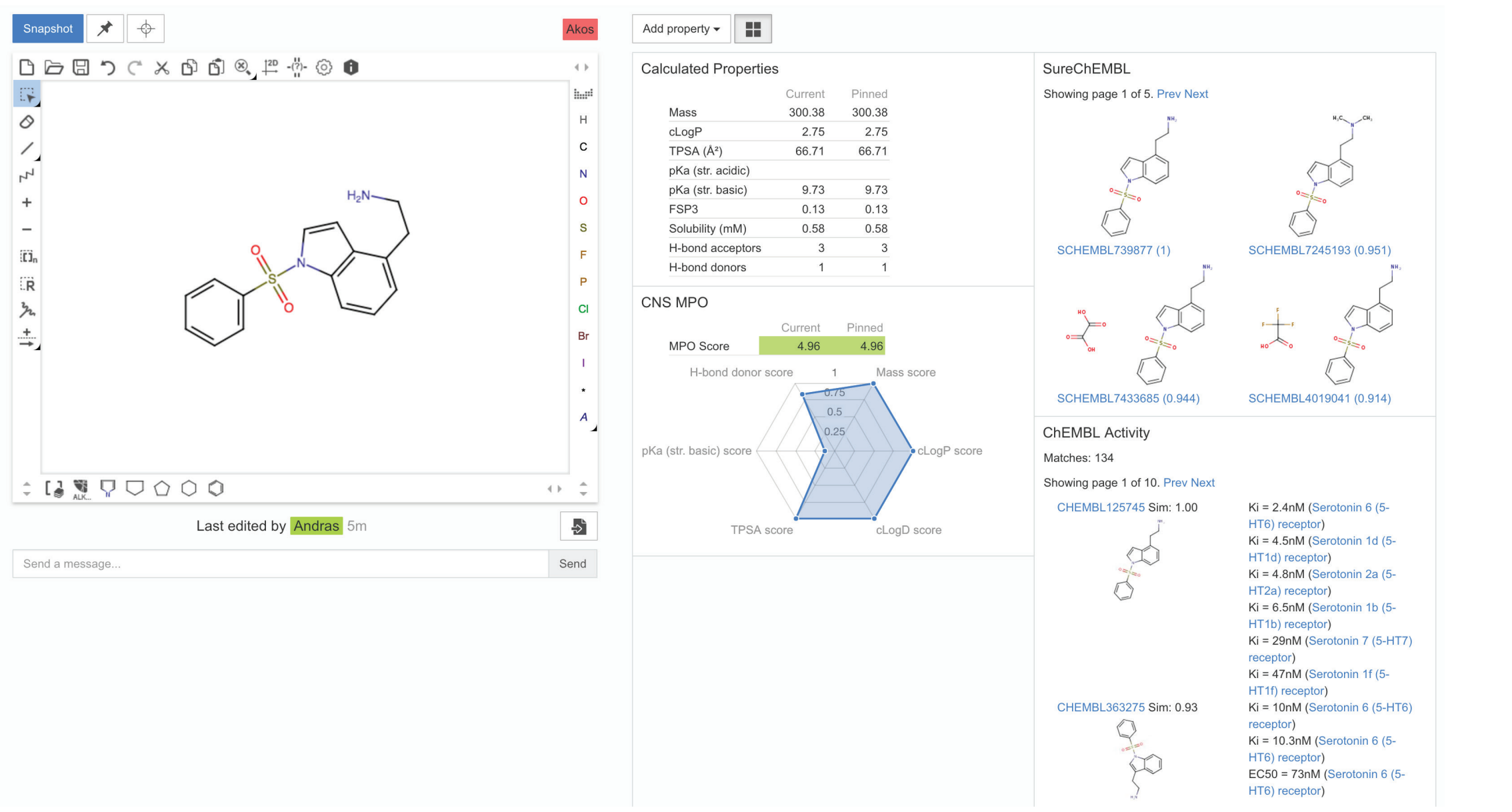
1. Characterization of the reference compound

Pros:

- Balanced phys-chem properties (**SCHEME 1.** Calculated properties)
- pKa limits the CNS MPO[1] score (**SCH. 1.** CNS MPO)
- Body of public activity data available (**SCH 1.** ChEMBL Activity)
- High affinity (2.4 nM) on the primer target (HTR6) (**SCH. 1.** ChEMBL Activity)
- SAR can be deduced from literature data
- 134 ChEMBL analogs are available within 0.8 Tanimoto similarity threshold

Cons:

- Patented (**SCH. 1.** SureChEMBL[2])
- Selectivity might be an issue



Last edited by Andras 5m

Send a message... Send

Calculated Properties

	Current	Pinned
Mass	300.38	300.38
cLogP	2.75	2.75
TPSA (Å²)	66.71	66.71
pKa (str. acidic)		
pKa (str. basic)	9.73	9.73
FSP3	0.13	0.13
Solubility (mM)	0.58	0.58
H-bond acceptors	3	3
H-bond donors	1	1

CNS MPO

	Current	Pinned
MPO Score	4.96	4.96

SureChEMBL

Showing page 1 of 5. Prev Next

SCHEMBL739877 (1)
SCHEMBL7245193 (0.951)
SCHEMBL7433685 (0.944)
SCHEMBL4019041 (0.914)

ChEMBL Activity

Matches: 134

Showing page 1 of 10. Prev Next

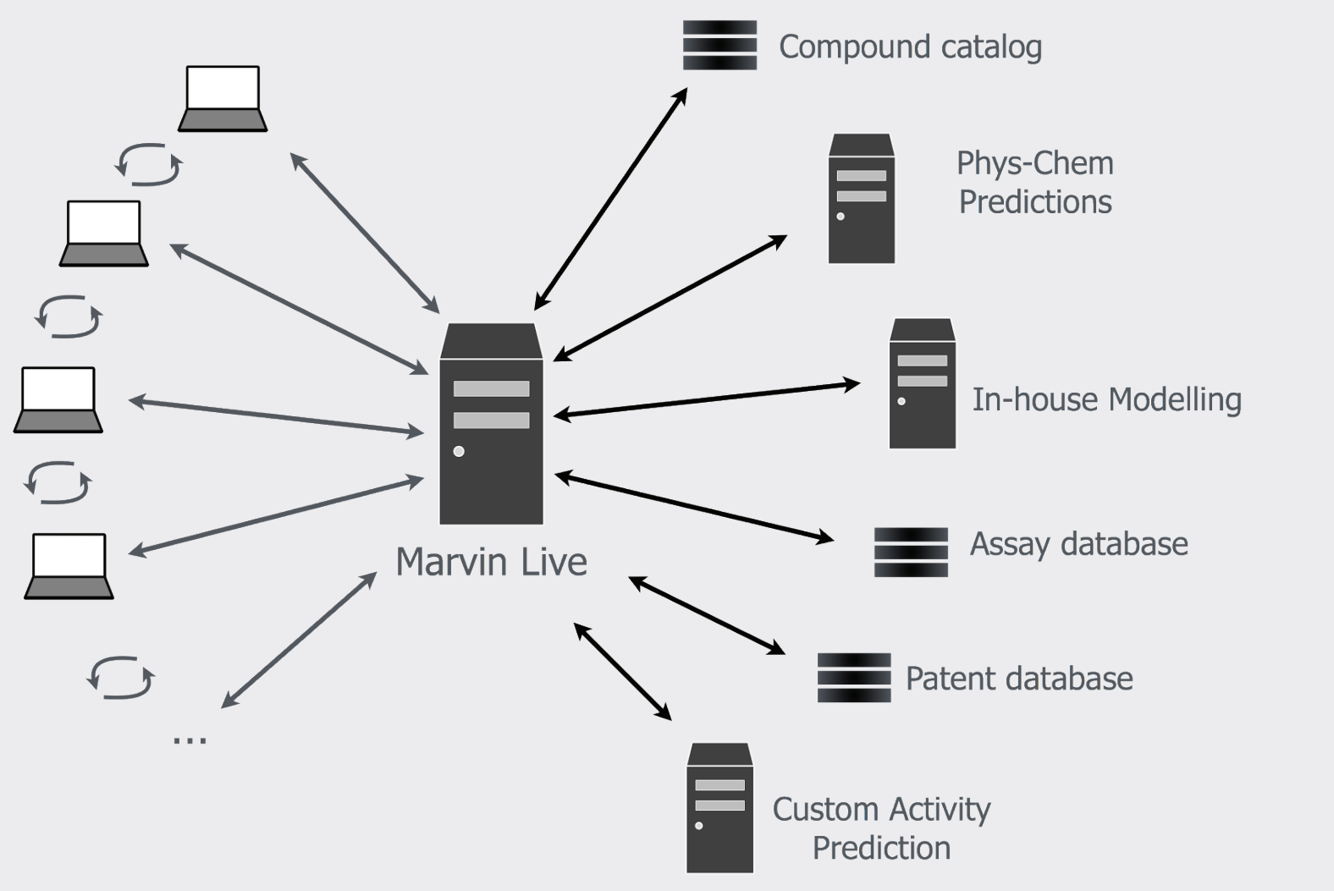
CHEMBL125745 Sim: 1.00
CHEMBL363275 Sim: 0.93

Ki = 2.4nM (Serotonin 6 (5-HT6) receptor)
Ki = 4.5nM (Serotonin 1d (5-HT1d) receptor)
Ki = 4.8nM (Serotonin 2a (5-HT2a) receptor)
Ki = 6.5nM (Serotonin 1b (5-HT1b) receptor)
Ki = 29nM (Serotonin 7 (5-HT7) receptor)
Ki = 47nM (Serotonin 1f (5-HT1f) receptor)
Ki = 10nM (Serotonin 6 (5-HT6) receptor)
Ki = 10.3nM (Serotonin 6 (5-HT6) receptor)
EC50 = 73nM (Serotonin 6 (5-HT6) receptor)

SCHEME 1. Interface to create chemical structures and to display results of plugins. Calculated phys-chem properties, most similar hits from SureChEMBL, and most similar compound from ChEMBL with public assay data are shown. Active links allow drill down for more information from the original source. ▶

Marvin Live

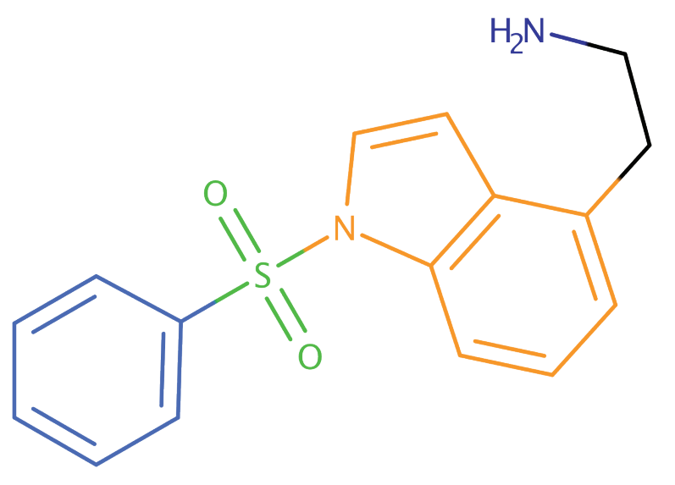
A NodeJS[3] based plugin system was created to communicate with available services. The plugin code sets up the connection to databases, web services, etc., handles the conversion of the molecule to the desired chemical structure format and formats the query to match the provided APIs. When the request completes rich HTML and reportable key-value pairs are sent to the connected browsers. This way Marvin Live and its plugin system acts as a central communication channel (**SCHEME 2.**) through which any internal or public resource is available to chemists for ideation and optimization.



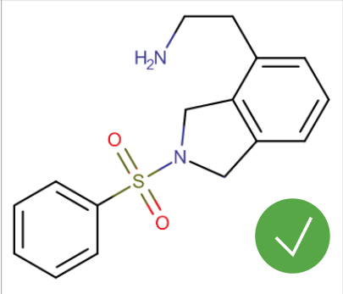
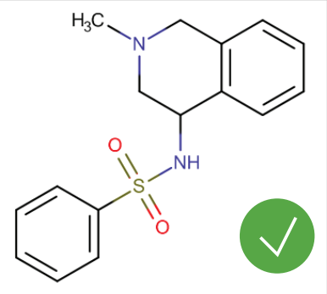
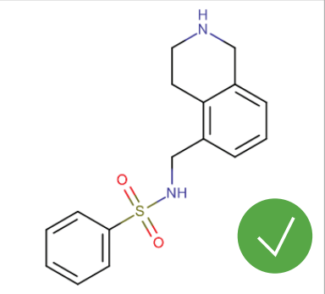
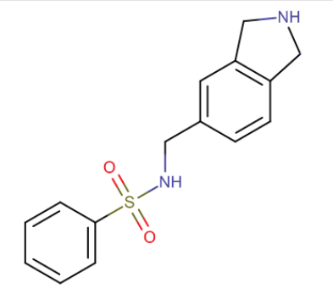
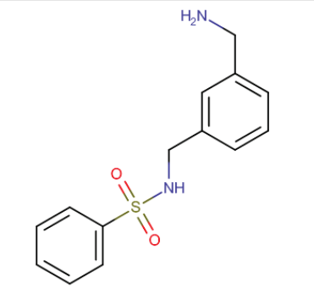
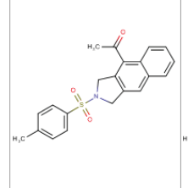
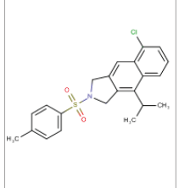
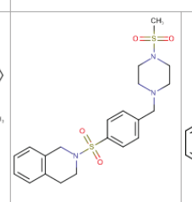
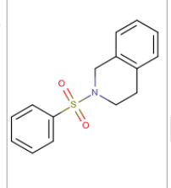
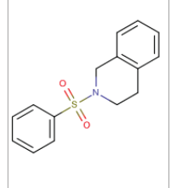
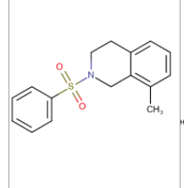
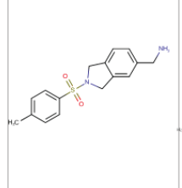
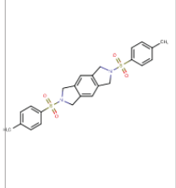
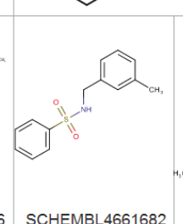
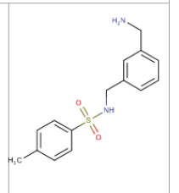
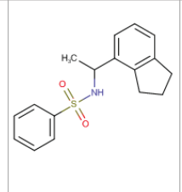
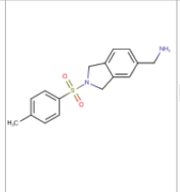
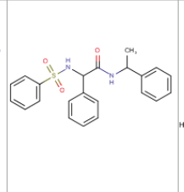
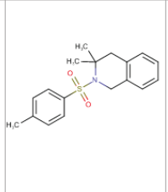
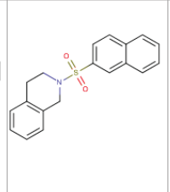
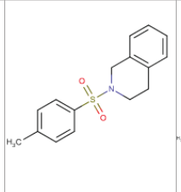
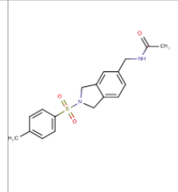
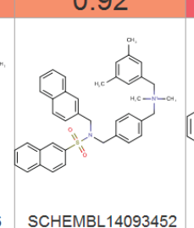
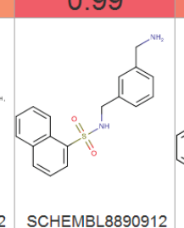
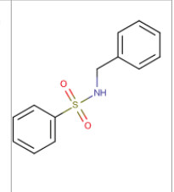
SCHEME 2. A schematic representation of how Marvin Live acts as an interface between chemists and diverse set of cheminformatics and modelling capabilities.

2. Scaffold hopping to IP-free space

- Generation of ideas within the space defined by the pharmacophore elements (**SCHEME 3**)
- Monitor freedom to operate (FTO) by instant similarity search against SureChEMBL database using MadFast Similarity Search[4]
- Preferred structures are located in the chemical space below 0.8 Tanimoto similarity threshold from exemplified structures collected from patent literature
- Top 4 excerpt of the closest analogs are shown in **TABLE 1**.
- 3 out of 5 exemplified structures can be considered in the next round of idea triage (**TABLE 1.**)
- Please note that this rapid similarity search does not include search in Markush structure, a more detailed IP analysis can be achieved by consulting with IP-experts
- Comparison of abundantly (HTR6 H2L - 9) and less represented (HTR6 H2L - 8) structures with regard to IP space are shown on **SCHEME 4.**



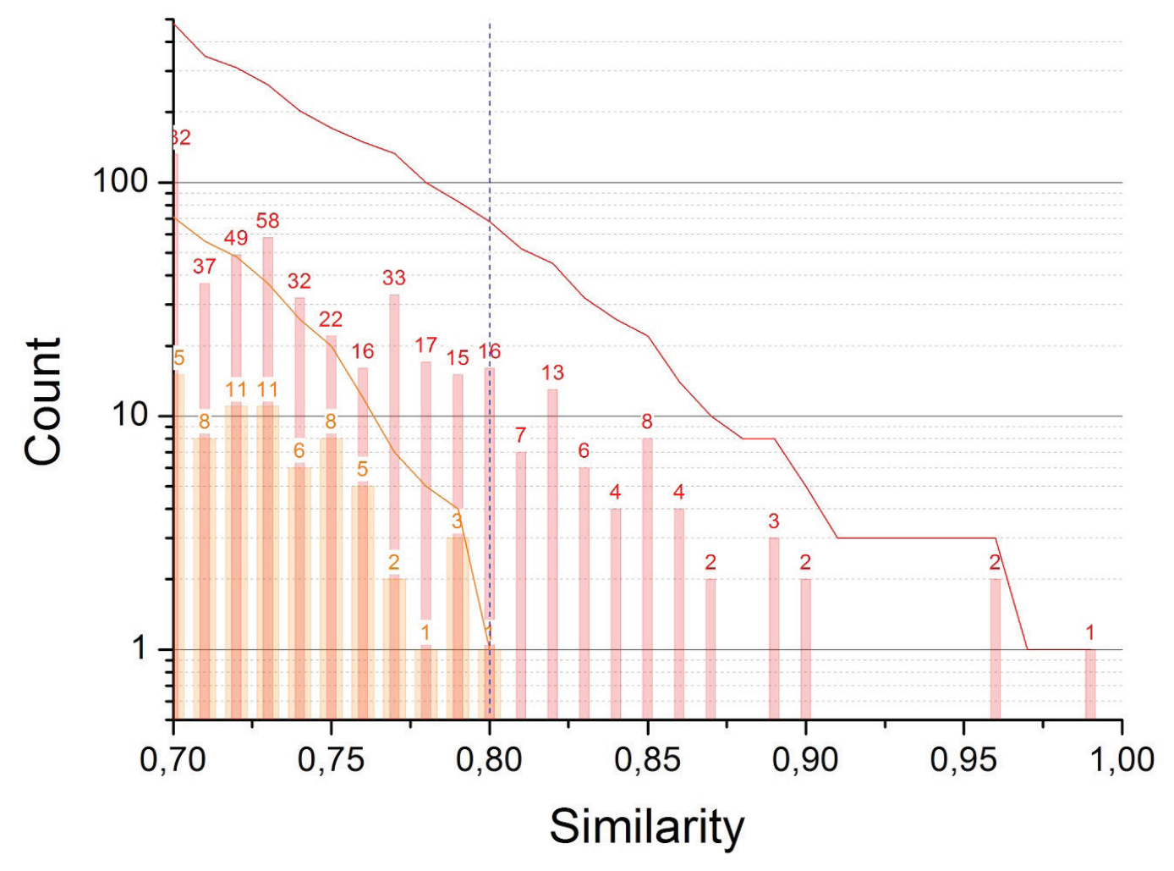
SCHEME 3. Ar–Sulfon–Central Ar-linker and positive center

ID	HTR6 H2L - 4		HTR6 H2L - 6		HTR6 H2L - 8		HTR6 H2L - 7		HTR6 H2L - 9	
Structure	 ✓		 ✓		 ✓					
Sure ChEMBL analogue	 SCHEMBL15388014	 SCHEMBL15397043	 SCHEMBL14025490	 SCHEMBL693455	 SCHEMBL693455	 SCHEMBL14591320	 SCHEMBL9757230	 SCHEMBL11773546	 SCHEMBL4661682	 SCHEMBL4923132
Similarity	0.79	0.78	0.77	0.76	0.8	0.79	0.93	0.92	0.99	0.95
Sure ChEMBL analogue	 SCHEMBL6874843	 SCHEMBL9757230	 SCHEMBL7026805	 SCHEMBL6826623	 SCHEMBL8589869	 SCHEMBL13302259	 SCHEMBL9757496	 SCHEMBL14093452	 SCHEMBL8890912	 SCHEMBL407414
Similarity	0.74	0.74	0.75	0.74	0.79	0.78	0.86	0.84	0.95	0.9
Time (s)	0.15		0.17		0.16		0.18		0.14	

◀ **TABLE 1.** Overlap analysis of the designed analogs versus the chemical space of exemplified structures from patents (SureChEMBL). Heat coloring of the Tanimoto similarity values reveals the IP risk.

SCHEME 4. Similarity based distribution of SureChEMBL analogs. The number of analogs are shown with binning size of 0.01 Tanimoto similarity values. ▶

▨ HTR6 H2L - 9
▨ HTR6 H2L - 8



3. Scaffold availability

- Assess feasibility by similarity based exploration of the availability in Enamine REAL database[5, 6]
- Top 4 excerpt of the closest analogs are shown in **TABLE 2.**
- Check opportunity to explore early SAR based on most similar analogs present in Enamine REAL. Example of well represented and purly represented structures are shown on **SCHEME 5**
- Two derivatives (HTR6 H2L - 6 and - 8) are available in the vendor DB

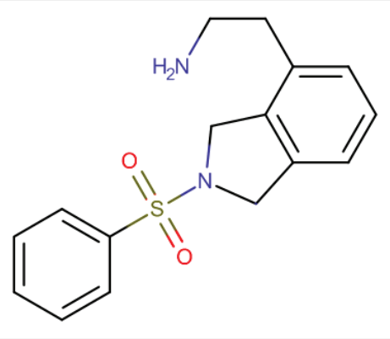
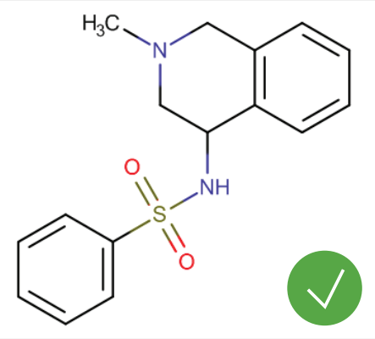
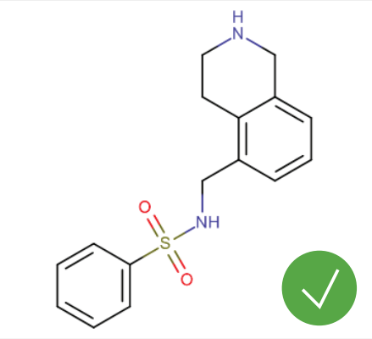
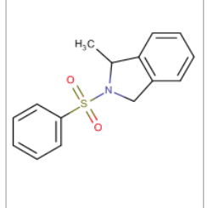
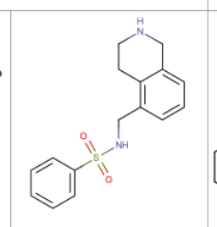
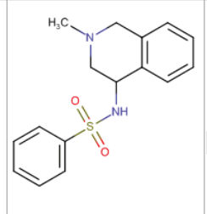
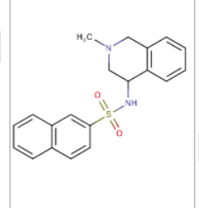
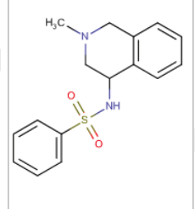
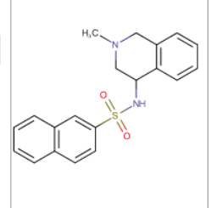
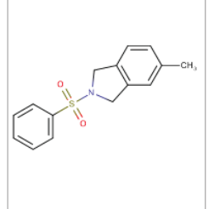
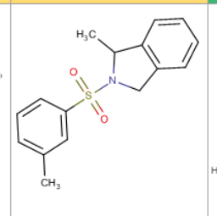
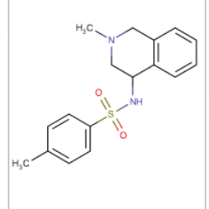
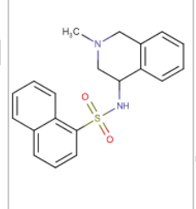
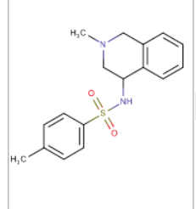
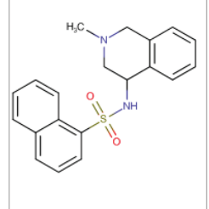
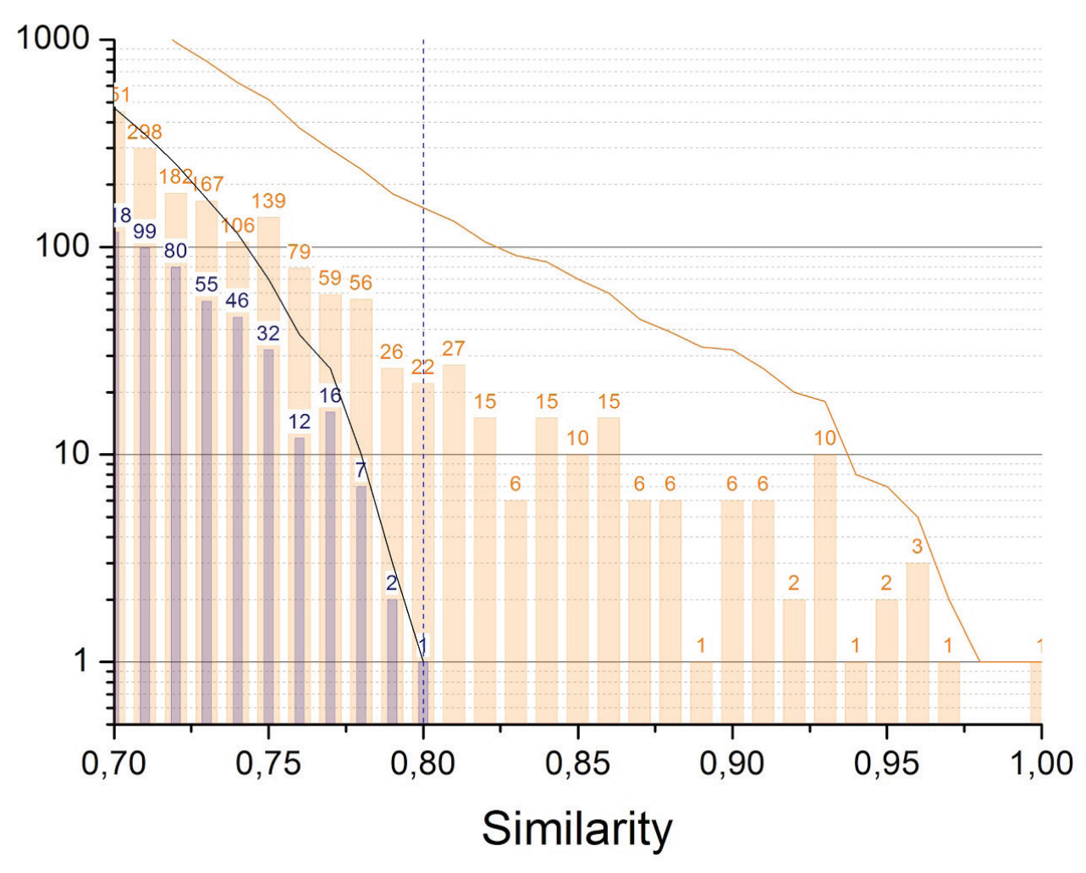
ID	HTR6 H2L - 4		HTR6 H2L - 6		HTR6 H2L - 8	
Structure			 ✓		 ✓	
Enamine REAL analogue	 Z1324206775	 Z2773225784	 Z1521981896	 Z1521980917	 Z1521981896	 Z1521980917
Similarity	0.8	0.79	1	0.98	1	0.97
Enamine REAL analogue	 Z1959132213	 Z1324206741	 Z1521981091	 Z1521981653	 Z1521981091	 Z1521981653
Similarity	0.8	0.79	0.98	0.98	0.95	0.95
Time (s)	1.1		1.1		1.1	

TABLE 2. Overlap analysis of the top ranked scaffolds with Enamine REAL ▶



◀ **SCHEME 5.** Similarity based distribution of Enamine REAL analogs. The number of analogs are shown with binning size of 0.01 Tanimoto similarity values.

▨ HTR6 H2L - 8
▨ HTR6 H2L - 4

MadFast Similarity Search

The Enamine REAL and SureChEMBL databases searched using 1024 bit Chemical Hashed Fingerprints[7] for the 171.5 and 17 million molecules, respectively. These were loaded for in-memory similarity search into a single virtual machine having 24 cores and 64 GB RAM. The high performance search engine, MadFast Similarity Search was accessed through a Marvin Live plugin, to automatically run similarity searches on the selected molecule with a maximum hit count of 20 sorted by similarity. The measured search times are displayed in **TABLE 1** and **TABLE 2.**

4. 3D alignment of the scaffolds to the reference

- Evaluation of steric complementarity with the pharmacophore elements of the reference compound revealed that HTR6 H2L - 8 fits, while HTR6 H2L - 6 displays larger deviations.

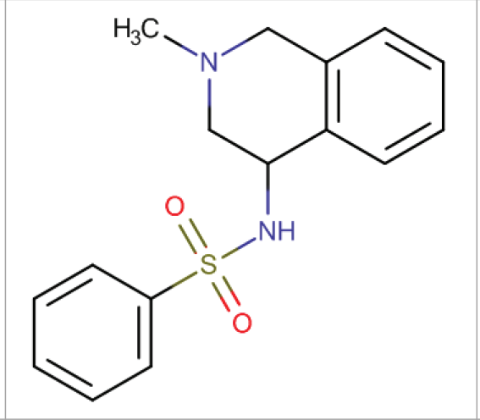
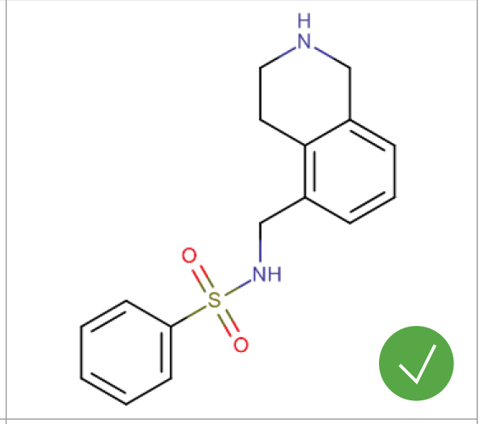
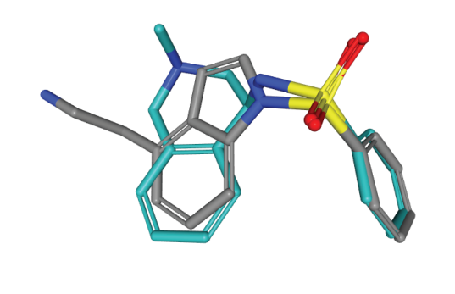
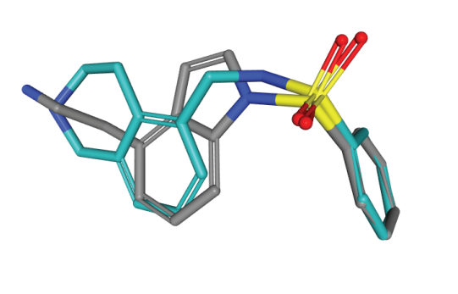
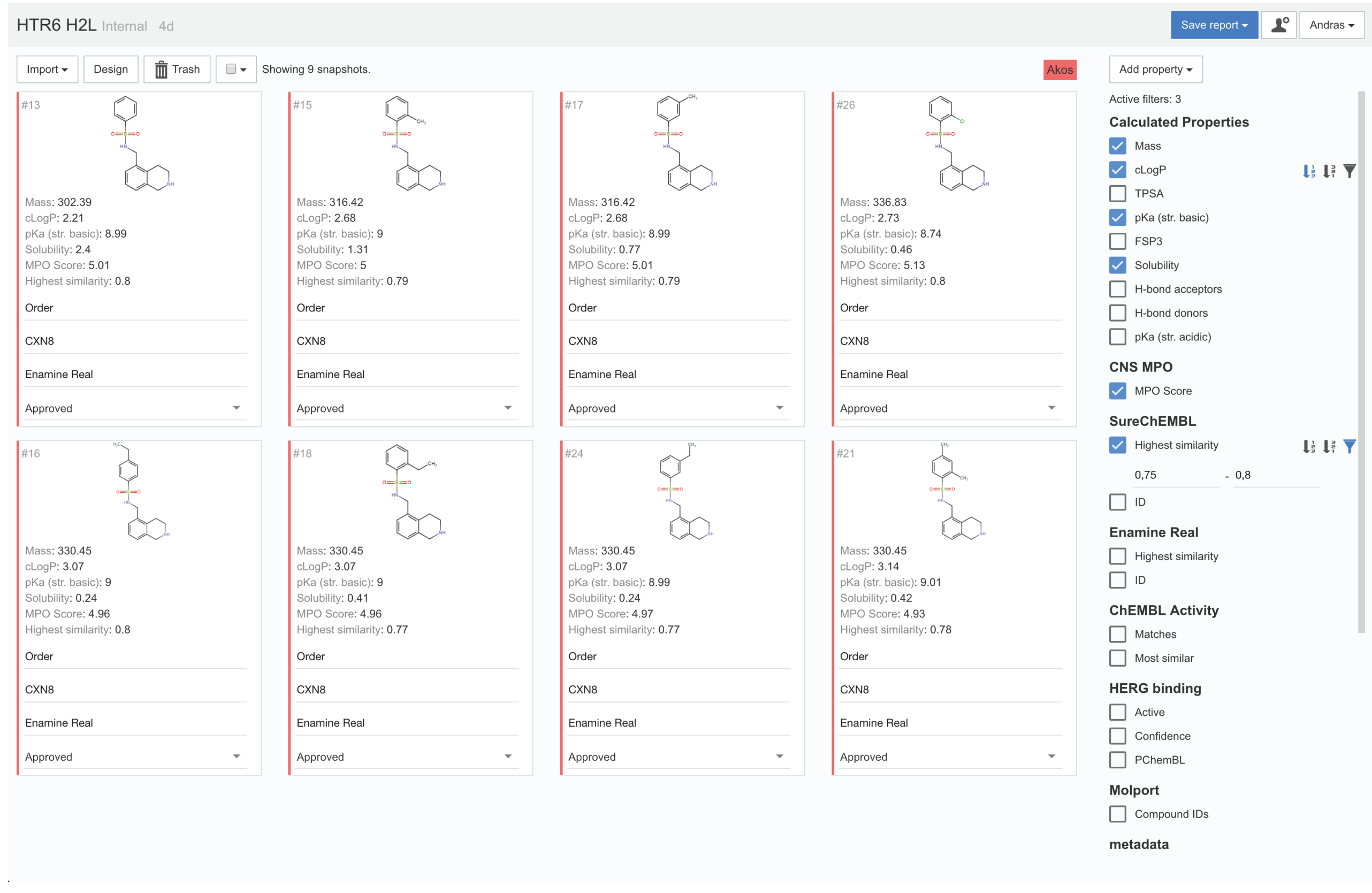
ID	HTR6 H2L - 6	HTR6 H2L - 8
Structure		 ✓
Alignment		

TABLE 3. MCS based 3D alignment of the two selected compounds. Alignment was done by 3D align tool from ChemAxon.[8]

5. SAR by catalog

- The *design mode* was used to find close structural analogs in the Enamine REAL catalog
- Halo- and methyl scan on the phenyl ring were both available from catalog as well as ethyl substituted derivatives
- The collection of close analogs was filtered based on similarity to patented structures from SureChEMBL in the *overview mode* (**SCHEME 6**)
- Sorting on values provided by plugins, like *Mass*, *cLogP* or *Solubility* aids early triage and decision making based on compound quality
- Status and task assignment facilitates the creation of a cherry picked library as a distinguished subgroup.



Showing 9 snapshots.

#13
Mass: 302.39
cLogP: 2.21
pKa (str. basic): 8.99
Solubility: 2.4
MPO Score: 5.01
Highest similarity: 0.8
Order
CXN8
Enamine Real
Approved

#15
Mass: 316.42
cLogP: 2.68
pKa (str. basic): 9
Solubility: 1.31
MPO Score: 5
Highest similarity: 0.79
Order
CXN8
Enamine Real
Approved

#17
Mass: 316.42
cLogP: 2.68
pKa (str. basic): 8.99
Solubility: 0.77
MPO Score: 5.01
Highest similarity: 0.79
Order
CXN8
Enamine Real
Approved

#26
Mass: 336.83
cLogP: 2.73
pKa (str. basic): 8.74
Solubility: 0.46
MPO Score: 5.13
Highest similarity: 0.8
Order
CXN8
Enamine Real
Approved

#16
Mass: 330.45
cLogP: 3.07
pKa (str. basic): 9
Solubility: 0.24
MPO Score: 4.96
Highest similarity: 0.8
Order
CXN8
Enamine Real
Approved

#18
Mass: 330.45
cLogP: 3.07
pKa (str. basic): 9
Solubility: 0.41
MPO Score: 4.96
Highest similarity: 0.77
Order
CXN8
Enamine Real
Approved

#24
Mass: 330.45
cLogP: 3.07
pKa (str. basic): 8.99
Solubility: 0.24
MPO Score: 4.97
Highest similarity: 0.77
Order
CXN8
Enamine Real
Approved

#21
Mass: 330.45
cLogP: 3.14
pKa (str. basic): 9.01
Solubility: 0.42
MPO Score: 4.93
Highest similarity: 0.78
Order
CXN8
Enamine Real
Approved

Active filters: 3

Calculated Properties
☑ Mass
☑ cLogP
☐ TPSA
☑ pKa (str. basic)
☐ FSP3
☑ Solubility
☐ H-bond acceptors
☐ H-bond donors
☐ pKa (str. acidic)

CNS MPO
☑ MPO Score

SureChEMBL
☑ Highest similarity
0,75 - 0,8
☐ ID

Enamine Real
☐ Highest similarity
☐ ID

ChEMBL Activity
☐ Matches
☐ Most similar

HERG binding
☐ Active
☐ Confidence
☐ PChemBL

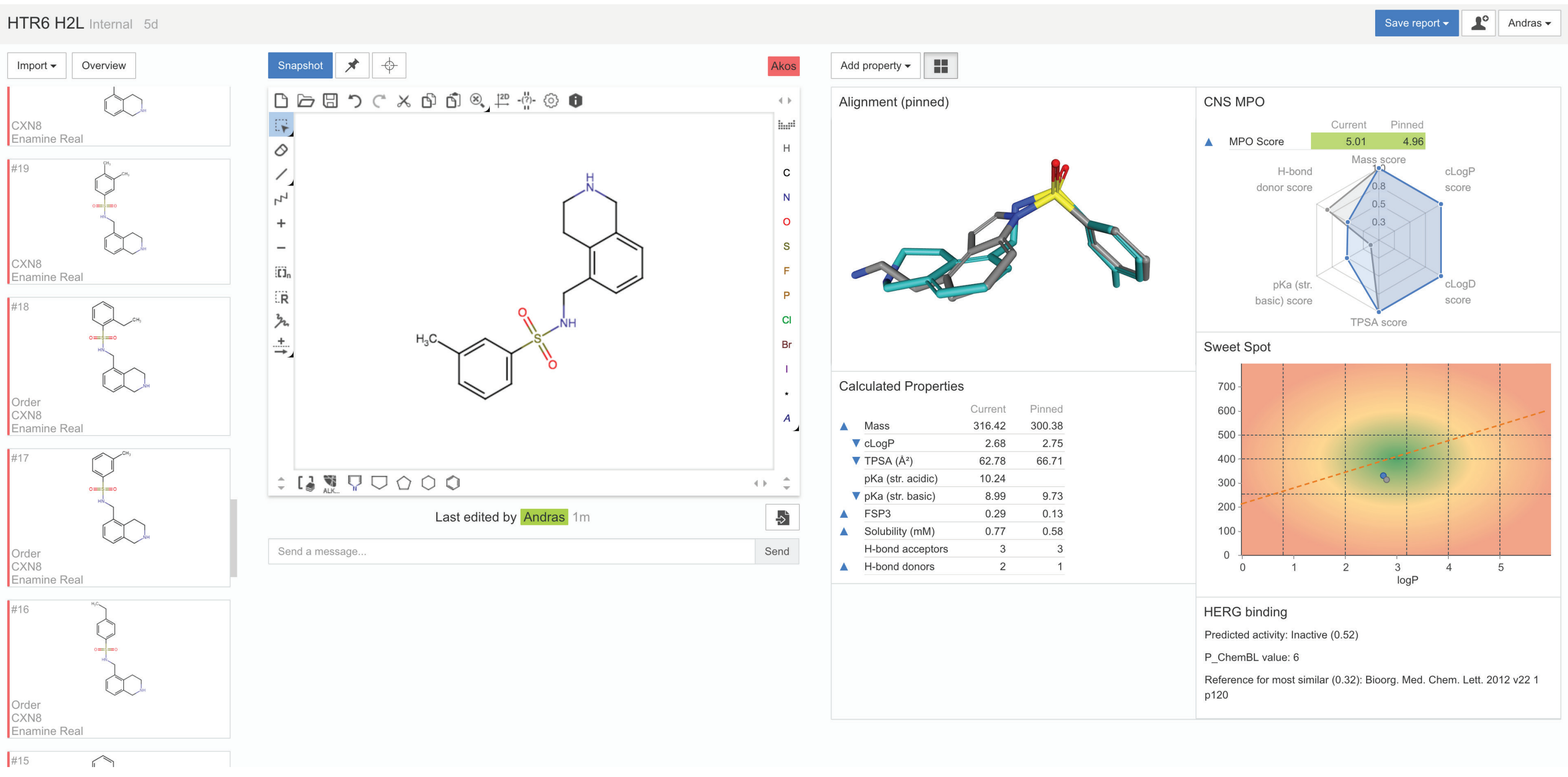
Molport
☐ Compound IDs

metadata

SCHEME 6. Overview mode of the finally selected custom library around HTR6 H2L -8 with selected calculated properties. ▶

6. Comprehend one-by-one

- Assessment of calculated properties compared to the reference compound (Pinned structure)
- Radar chart with visualization of CNS MPO increments for detailed comparison
- Cross-check using 3D alignment plugin
- Predict hERG liability by a utilizing complex KNIME workflow as a plugin



CXN8
Enamine Real

#19
CXN8
Enamine Real

#18
Order
CXN8
Enamine Real

#17
Order
CXN8
Enamine Real

#16
Order
CXN8
Enamine Real

#15

Last edited by Andras 1m

Send a message... Send

Calculated Properties

		Current	Pinned
▲	Mass	316.42	300.38
▼	cLogP	2.68	2.75
▼	TPSA (Å²)	62.78	66.71
	pKa (str. acidic)	10.24	
▼	pKa (str. basic)	8.99	9.73
▲	FSP3	0.29	0.13
▲	Solubility (mM)	0.77	0.58
	H-bond acceptors	3	3
▲	H-bond donors	2	1

CNS MPO

		Current	Pinned
▲	MPO Score	5.01	4.96

HERG binding

Predicted activity: Inactive (0.52)

P_ChemBL value: 6

Reference for most similar (0.32): Bioorg. Med. Chem. Lett. 2012 v22 1 p120

SCHEME 7. Comparison of one selected analog to the reference compound in Design mode. ▶

Conclusions

Our results reveal that ligand design can be fostered by providing instant access to extra large compound collections, knowledge bases and predicted properties. Therefore, it facilitates widening the scope of the chemical space and helps escape an availability bias that would constrain the series of ideas and increase the likelihood of project failure due to insufficient freedom to operate.

The rich visualization in design mode and the idea management in overview mode help prioritizing the compound series and decision making in Marvin Live.

[1] Wager, Travis T., et al. ACS *Chem. Neurosci.*, **2010**, 1 (6), pp 435–449

[2] Papadatos, George, et al. *Nucleic Acids Res.* **2016** Jan 4; 44(Database issue): D1220–D1228. Database downloaded 01 May 2017

[3] https://nodejs.org/ Accessed 04 June 2017

[4] https://www.chemaxon.com/products/madfast/ Accessed 04 June 2017

[5] Shivanyuk, A. N., et al. *Chemistry Today* **2007**, 25, 58-59. Database downloaded 01 May 2017

[6] Tolmachev, Andrey, et al. *ACS Combinatorial Science* **2016**, 18.10, 616-624.

[7] https://docs.chemaxon.com/display/docs/Chemical+Hashed+Fingerprint Accessed 04 June 2017

[8] https://www.chemaxon.com/products/calculator-plugins/molecular-modelling/ Accessed 04 June 2017

[9] Galambos J., et al. ACS *J. Med. Chem.*, **2017**, 60 (6), pp 2470–2484

ChemAxon Ltd, 1031 Budapest, Zahony u. 7, Hungary

ChemAxon

www.chemaxon.com · ml-support@chemaxon.com