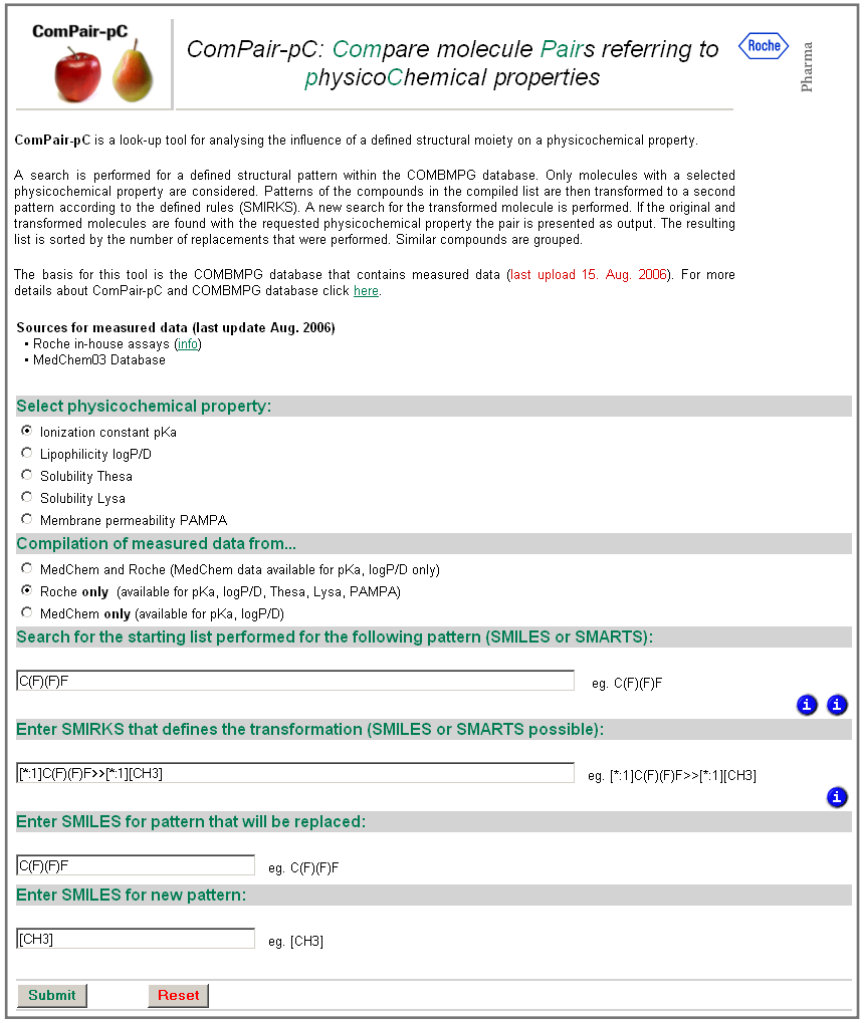
LUCID: Supervised Multidimensional Optimization of Compounds Using Matched Molecular Pairs

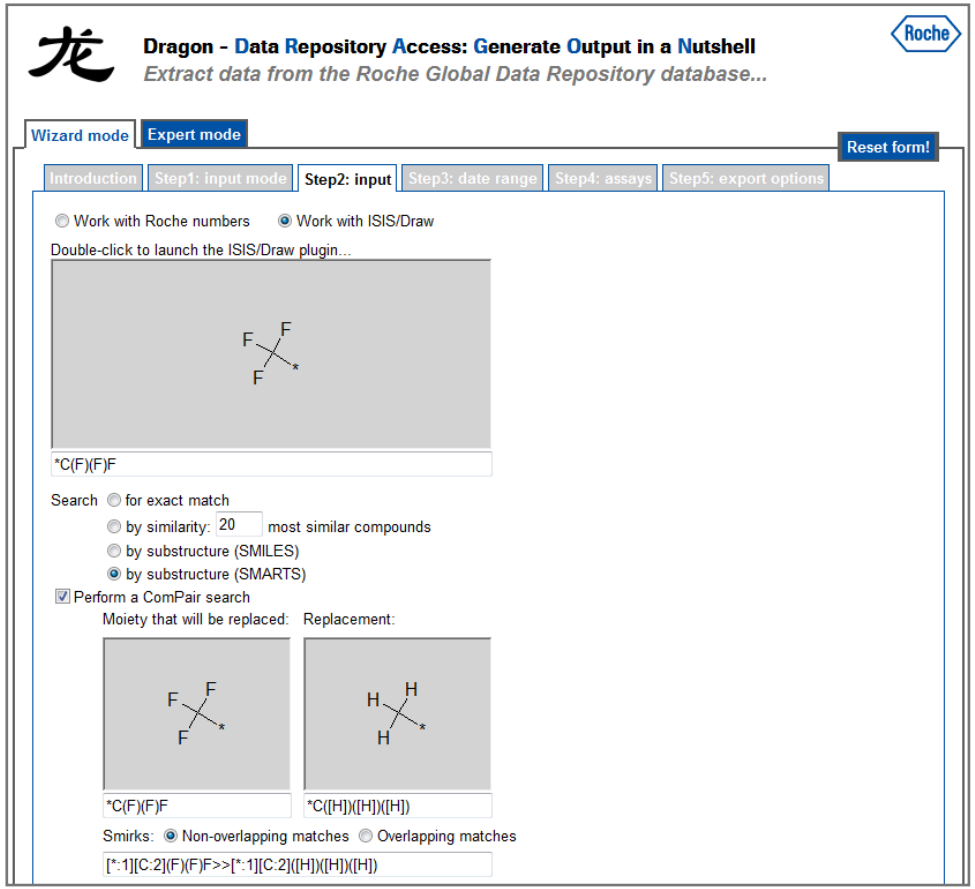
Grégori Gerebtzoff

Bruce Nauman, 'One Hundred Live and Die'

History of Matched Molecular Pairs at Roche

2006: ComPair-pC; 2009: ComPair in Dragon





Computationally efficient algorithm to identify MMPs[1]

Usage and limitations of Hussain's method

Problem:

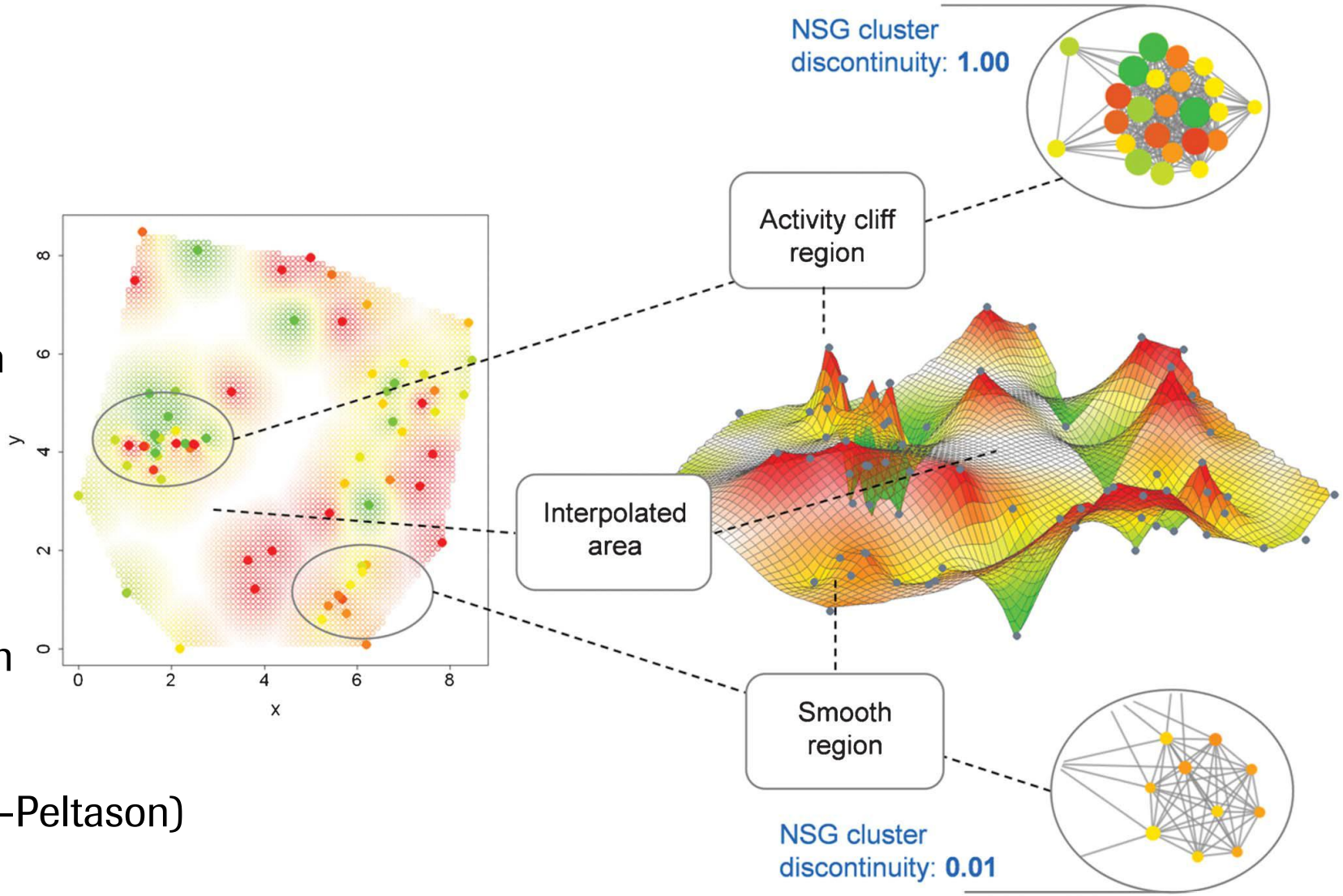
→the Hussain's algorithm identifies many MMPs and generates "too many" transformations

→in the original publication, 5.3 million MMPs identified (2.5 million transformations) for a dataset of 330'000 compounds generating 21.7 million fragmentations

Usage:

→mainly identification and prediction of activity cliffs

→Readily available at Roche:
SAR Map and SAR Tree (Lisa Sach-Peltason)
http://www.youtube.com/watch?v=gFIfyrBtvPo



[1] J. Hussain & C. Rea, J. Chem. Inf. Model. 2010, 50 (3), 339

Illustration from: D. Stumpfe & J. Bajorath, RSC Adv. 2012, 2, 369

Some thoughts on what we want to achieve

- The ComPair approach allows us to answer the question:
  “What is the effect of the replacement of side chain X to side chain Y on various properties (hepatic clearance, hERG, LogD)”

- What people expect to get in the future:
  Easy guidelines = a catalog of transformations to solve MedChem issues

- New paradigm: LUCID
  Instead of searching for the effect of specific replacements, LUCID helps the chemists optimizing various properties by suggesting potentially beneficial replacements

  LUCID is based on the Hussain’s method, but uses a lower number of relevant MMPs

Can we use Hussain's method to extend the ComPair approach?

How to reduce the number of MMPs and select relevant MMPs

Reduction of the number of MMPs:

→ Side chains must be smaller than x (17) atoms

→ Core part of the molecules must be greater than x (5) atoms

Selection of relevant MMPs:

→ Define the size of the replacement

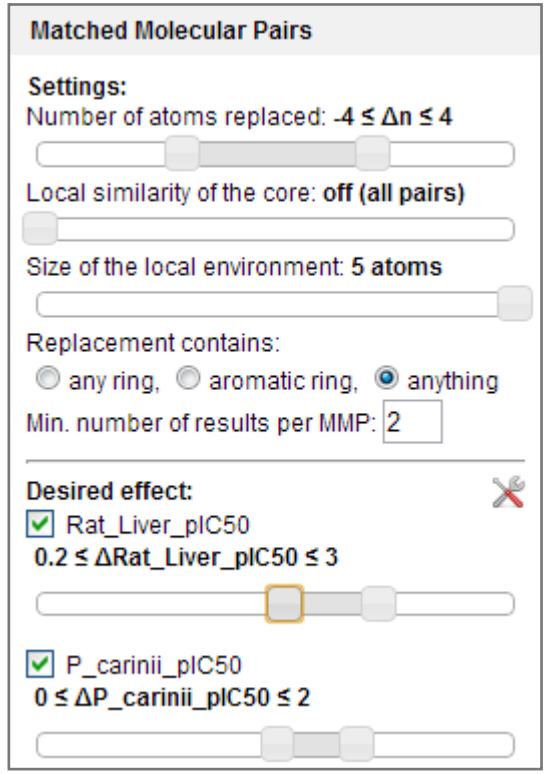
→ Define the local similarity of the core

→ Select required features in the replacement

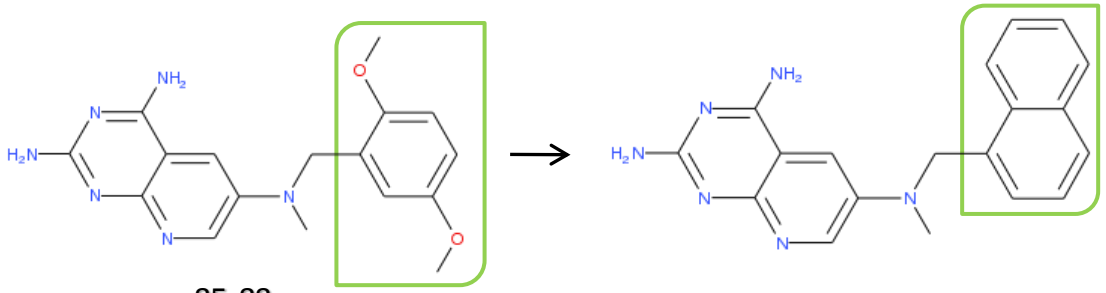
→ Filter out transformations with less than n (2) examples
(n=1 → ~90% of the transformations)

→ List of properties (physchem, safety, …) to optimize

→ User-defined site(s) of transformation



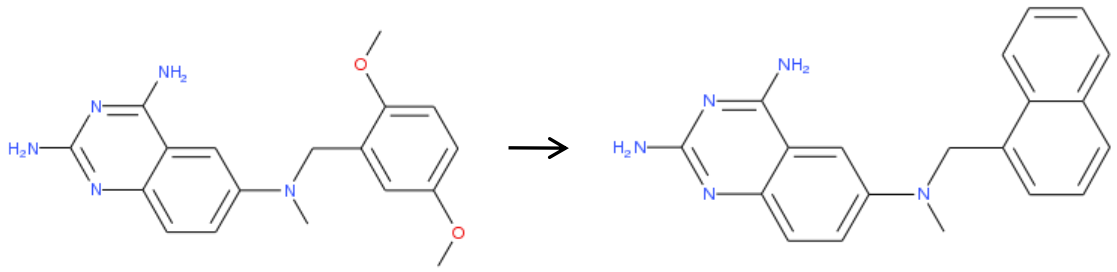
Chemical similarity at the environment of the attachment point



25-22

P carinii pIC50 = 7.76
T gondii pIC50 = 8.20

P carinii pIC50 = ??
T gondii pIC50 = ??



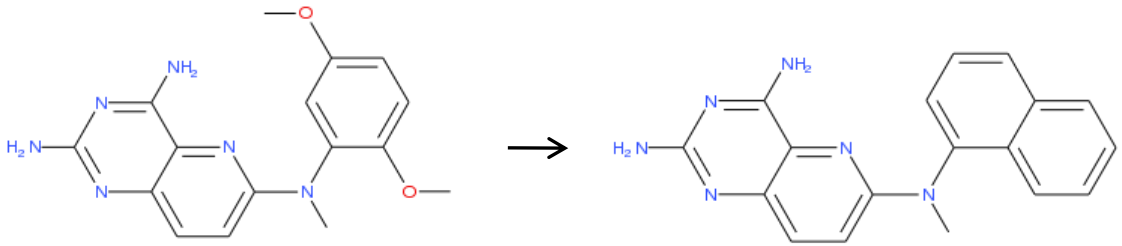
15-11

P carinii pIC50 = 7.06
T gondii pIC50 = 7.52

Δ = 0.71
Δ = 0.15

15-15

P carinii pIC50 = 7.77
T gondii pIC50 = 7.68



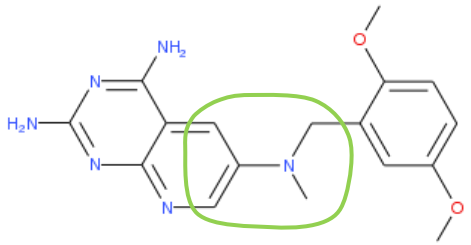
22-20

P carinii pIC50 = 7.47
T gondii pIC50 = 7.39

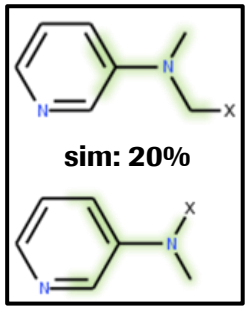
Δ = -2.18
Δ = -1.71

22-23

P carinii pIC50 = 5.29
T gondii pIC50 = 5.68



AllChem.GetMorganFingerprint(Chem.MoLFromSmiles('[*]CN(C)c1cnccc1'), 4, fromAtoms=[0], useFeatures=True)
AllChem.GetMorganFingerprint(Chem.MoLFromSmiles('[*]N(C)c1cnccc1'), 4, fromAtoms=[0], useFeatures=True)
DataStructs.DiceSimilarity(fp1, fp2)



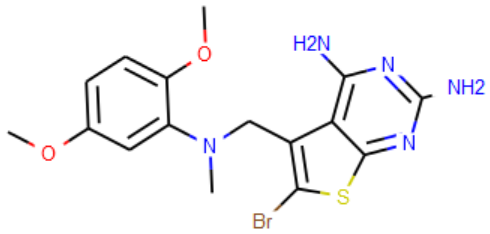
→ The chemical environment around the attachment point is stored (FCFP_X)

→ The similarity is calculated between the environment of the parent molecule and the requested MMPs

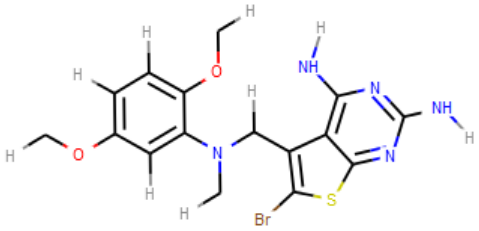
Similarity	n	Δ (P carinii)	Δ (T gondii)
Low (all pairs)	17	0.48 (SD=0.89)	0.19 (SD=0.73)
Medium	5	0.71 (SD=0.67)	0.15 (SD=0.5)
High	1	0.71	0.15

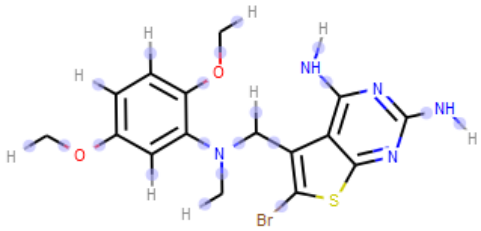
User-defined sites of transformation

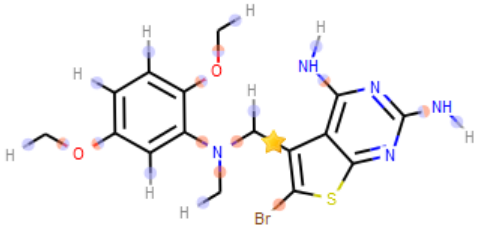
LUCID: towards an interactive visualization of Chemical Space and MMPs

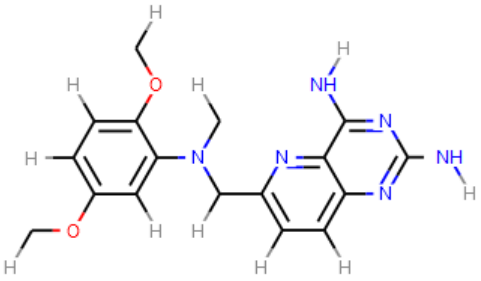


36-6D

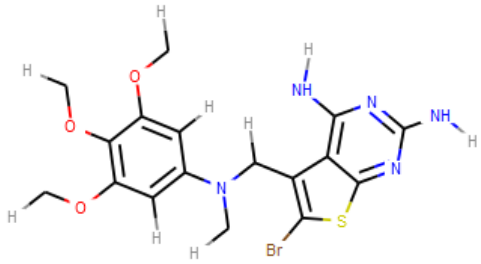








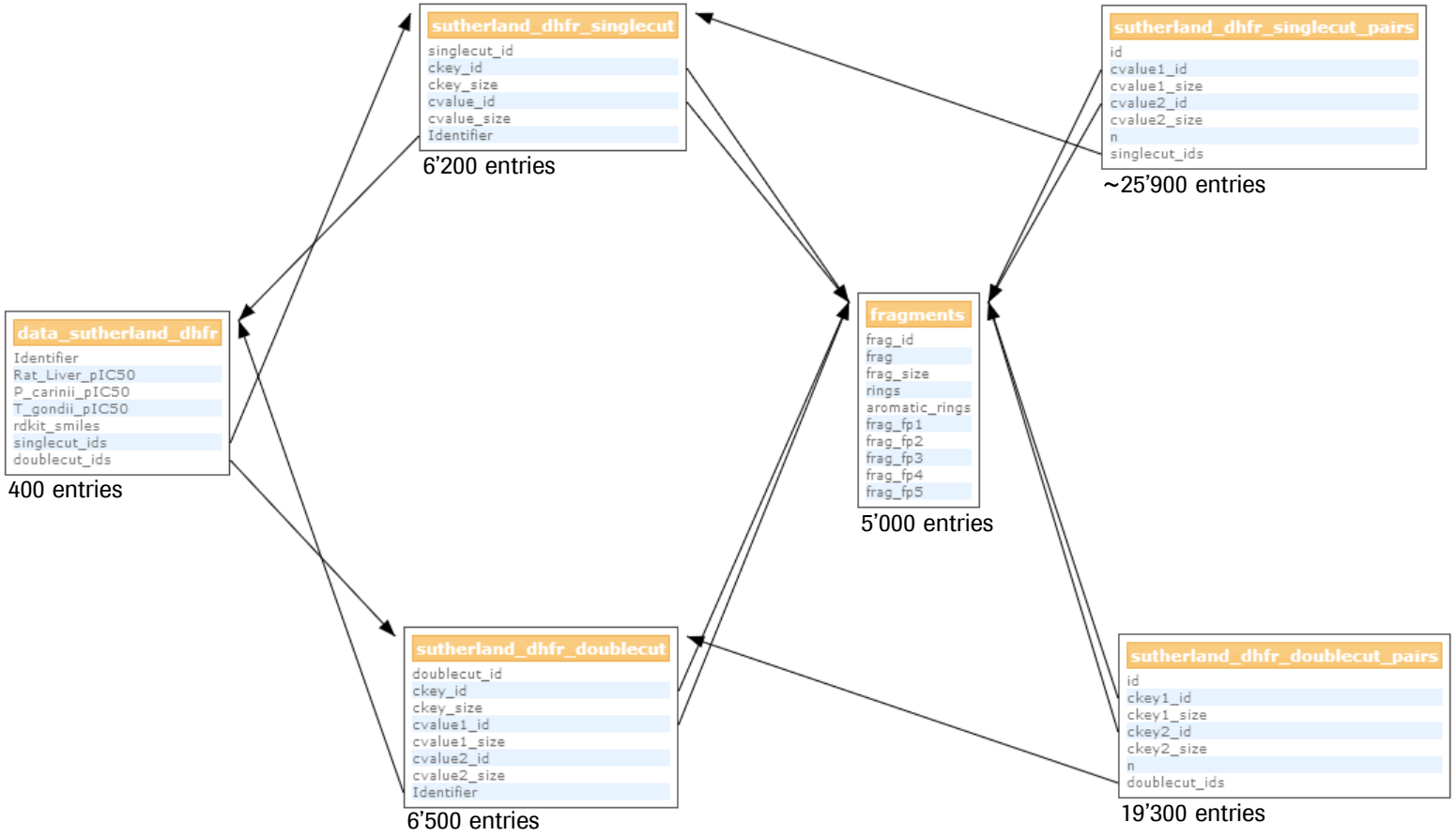
21-18



36-6C

Sutherland et al. (doi:10.1023/B:JCAM.0000047814.85293.da)

Structure of the LUCID database



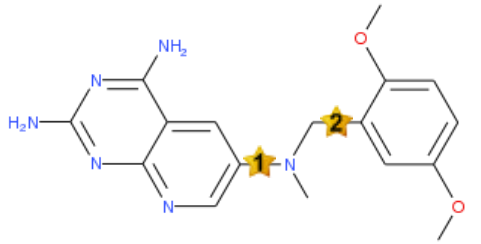
Data table

Identifier	Rat_Liver_pIC50	P_carinii_pIC50	T_gondii_pIC50	rdkit_smiles
10-2B	5.40894	5.2636	5.72125	Nc1nc(N)c2cc(COC(=O)c3ccccc3)cnc2n1
10-2C	6.63827	6.60206	7.25181	Cc1c(COC(=O)c2ccccc2)cnc2nc(N)nc(N)c
10-2D	5.1549	5.00877	5.23657	Nc1nc(N)c2nc(COC(=O)c3ccccc3)cnc2n1
10-3A	7.58503	6.25181	7.30103	COc1cc(C(=O)OCc2ccc3nc(N)nc(N)c3c2)c
10-4A	6.37675	5.61979	6.65758	COc1ccc(OC)c(C(=O)OCc2ccc3nc(N)nc(N
10-5A	6.30103	5.85387	6.85387	COc1ccc(C(=O)OCc2ccc3nc(N)nc(N)c3c2
10-6A	6.45593	6.33724	6.82391	COc1cc(OC)cc(C(=O)OCc2ccc3nc(N)nc(N
11-AH2503	5.65758	6.20761	6	CCc1nc2nc(N)nc(N)c2nc1CC
11-AH2504	6.38722	5.69897	7.74473	CCc1nc2nc(N)nc(N)c2nc1-c1ccccc1OC
11-GR92754	6.49485	7.08619	7.55284	CC(C)Cc1nc2nc(N)nc(N)c2nc1CC(C)C

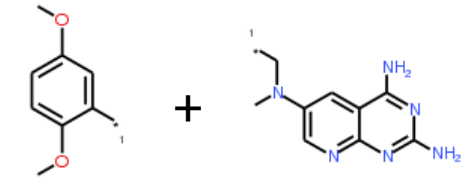
Singlecut, Doublecut, and fragments tables

mmp_singlecut_id	ckey_id	ckey_size	cvalue_id	cvalue_size	Identifier
2959487	2777833	1	2779296	24	25-22
2959488	2777833	1	2779299	24	25-22
2959486	2777833	1	2779301	24	25-22
2959496	2777834	1	2779295	24	25-22
2959497	2777834	1	2779300	24	25-22
2959483	2777869	2	2779297	23	25-22
2959484	2777869	2	2779298	23	25-22
2959480	2777949	10	2778561	15	25-22
2959482	2778312	13	2778557	12	25-22
2959485	2778316	11	2778560	14	25-22
2959495	2778557	12	2778312	13	25-22
2959489	2778560	14	2778316	11	25-22
2959492	2778561	15	2777949	10	25-22
2959490	2779295	24	2777834	1	25-22
2959493	2779296	24	2777833	1	25-22
2959500	2779297	23	2777869	2	25-22
2959481	2779298	23	2777869	2	25-22
2959499	2779299	24	2777833	1	25-22
2959491	2779300	24	2777834	1	25-22
2959494	2779301	24	2777833	1	25-22

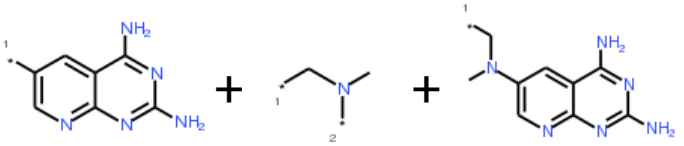
mmp_doublecut_id	ckey_id	ckey_size	cvalue1_id	cvalue1_size	cvalue2_id	cvalue2_size	Identifier
3620	2780337	1	2777949	10	2778560	14	25-22
3628	2780356	3	2777949	10	2778557	12	25-22
3615	2780395	2	2778316	11	2778557	12	25-22
3613	2780409	1	2777833	1	2779297	23	25-22
3612	2780409	1	2779298	23	2777833	1	25-22
3616	2780522	9	2777833	1	2778561	15	25-22
3629	2780527	8	2777869	2	2778561	15	25-22
3626	2780890	10	2777833	1	2778560	14	25-22
3610	2780892	12	2777833	1	2778557	12	25-22
3630	2780896	11	2777869	2	2778557	12	25-22
3614	2780899	9	2777869	2	2778560	14	25-22
3621	2781142	13	2778316	11	2777833	1	25-22
3618	2781145	14	2777949	10	2777833	1	25-22
3625	2781148	13	2778316	11	2777834	1	25-22
3624	2781149	11	2778312	13	2777834	1	25-22
3619	2781152	14	2777949	10	2777834	1	25-22
3609	2781986	23	2777833	1	2777834	1	25-22
3611	2781987	21	2777869	2	2777869	2	25-22
3617	2781988	23	2777834	1	2777834	1	25-22
3622	2781989	22	2777869	2	2777834	1	25-22
3623	2781990	22	2777833	1	2777869	2	25-22
3627	2781991	23	2777833	1	2777833	1	25-22



25-22



2777949 + 2778561



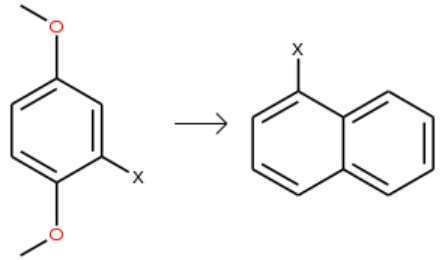
2778557 + 2780356 + 2778561

frag_id	frag	frag_size	frag_fp1	frag_fp2	frag_fp3	frag_fp4	frag_fp5
2777949	[*:1]c1cc(OC)ccc1OC	10	ÿÿÿÿÁü¿	ÿÿÿÿ«¦ÊÁü¿	ÿÿÿÿ$Pe«¦ÊÁü¿	ÿÿÿÿ$Pez'g`«¦ÊÁü...	ÿÿÿÿ$PeoØè/z'g`«¦...
2778557	[*:1]c1cnc2nc(N)nc(N)c2c1	12	ÿÿÿÿÁü¿	ÿÿÿÿ«¦ÊÁü¿	ÿÿÿÿ«¦ÊÁü¿Mãsç	ÿÿÿÿ«¦Ê Áü¿R¸4æMã...	ÿÿÿÿ¯¡û«¦ÊÁü¿R¸...
2778561	[*:1]CN(C)c1cnc2nc(N)nc(N)c2c1	15	ÿÿÿÿ=ü¿	ÿÿÿÿ=ü¿J«=ì	ÿÿÿÿ=ü¿J«=ìô‹%ð	ÿÿÿÿ„%bF=ü¿J«=ìô‹...	ÿÿÿÿ„%bF £C÷~=ü¿J«...
2780356	[*:1]CN([*:2])C	3	ÿÿÿÿ=ü¿Ãü¿	ÿÿÿÿšæÒµ=ü¿Ãü¿J«...	ÿÿÿÿaó ¥ŸšæÒµ=ü¿Ãü...	ÿÿÿÿaó ¥ŸšæÒµ=ü¿Ãü...	ÿÿÿÿaó ¥ŸšæÒµ=ü¿Ãü...

fp.ToBinary( )
DataStructs.UIntSparseIntVect( )

Singlecut_pairs table

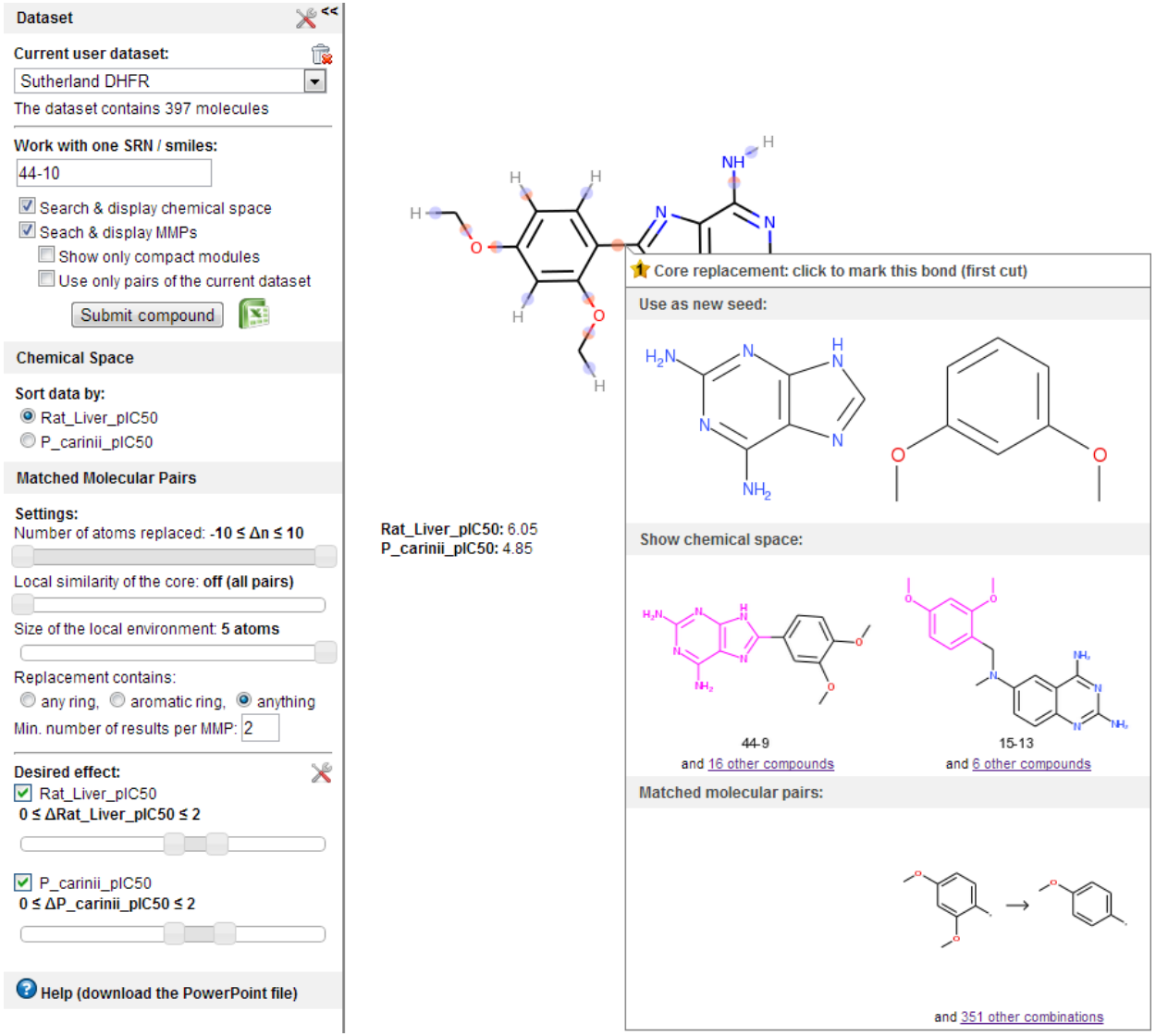
id	cvalue1_id	cvalue1_size	cvalue2_id	cvalue2_size	n	mmp_singlecut_ids
351	2777949	10	2778139	12	25	2956302|2956699|16,2959492|2957696|15,2958654|2958...
309	2777949	10	2778076	10	17	2959358|2957637|16,2960705|2958435|15,2957478|2961...
315	2777949	10	2778859	8	13	2959358|2960629|16,2960705|2959855|15,2957478|2959...
425	2777949	10	2778066	6	11	2960365|2958279|15,2957478|2958449|15,2959492|2957...
349	2777949	10	2778107	10	10	2956302|2959400|16,2960705|2960725|15,2957105|2957...
307	2777949	10	2777972	10	9	2959358|2957914|16,2960705|2960637|15,2957478|2959...
246	2777949	10	2777908	8	8	2959358|2956202|16,2957478|2957672|15,2959150|2957...
280	2777949	10	2777921	7	8	2959358|2956236|16,2960705|2956255|15,2957478|2959...
311	2777949	10	2778331	10	8	2959358|2959334|16,2960705|2960745|15,2957478|2957...
294	2777949	10	2777941	8	5	2959358|2956285|16,2960705|2956960|15,2957478|2956...
317	2777949	10	2779256	12	5	2959358|2959395|16,2960705|2960660|15,2959717|2959...
353	2777949	10	2778605	12	5	2956302|2958513|16,2957105|2961974|15,2959492|2957...
423	2777949	10	2778052	18	5	2960365|2960053|15,2960967|2962194|14,2961356|2960...
711	2777949	10	2777964	8	5	2960705|2956946|15,2957478|2956723|15,2959150|2956...
995	2777949	10	2778928	8	5	2957478|2959301|15,2959150|2959197|14,2958312|2960...
1179	2777949	10	2778166	13	5	2960581|2958901|15,2962010|2956774|13,2960104|2956...
264	2777949	10	2777911	7	4	2959358|2956212|16,2960705|2956928|15,2957478|2957...
658	2777949	10	2777932	7	4	2960705|2956239|15,2957478|2959069|15,2959150|2957...
10381	2777949	10	2778558	13	4	2962010|2958223|13,2957067|2961737|13,2961277|2957...



2959358|2957637|16,2960705|2958435|15,
2957478|2961128|15,2957105|2957172|15,
2958876|2958887|15,2959717|2960608|15,
2959150|2961101|14,2960362|2956552|14,
2961779|2961754|14,2959903|2957124|14,
2959509|2959472|14,2960967|2960918|14,
2959739|2961748|14,2960082|2960067|14,
2961199|2962103|14,2958404|2961043|13,
2961689|2961675|13

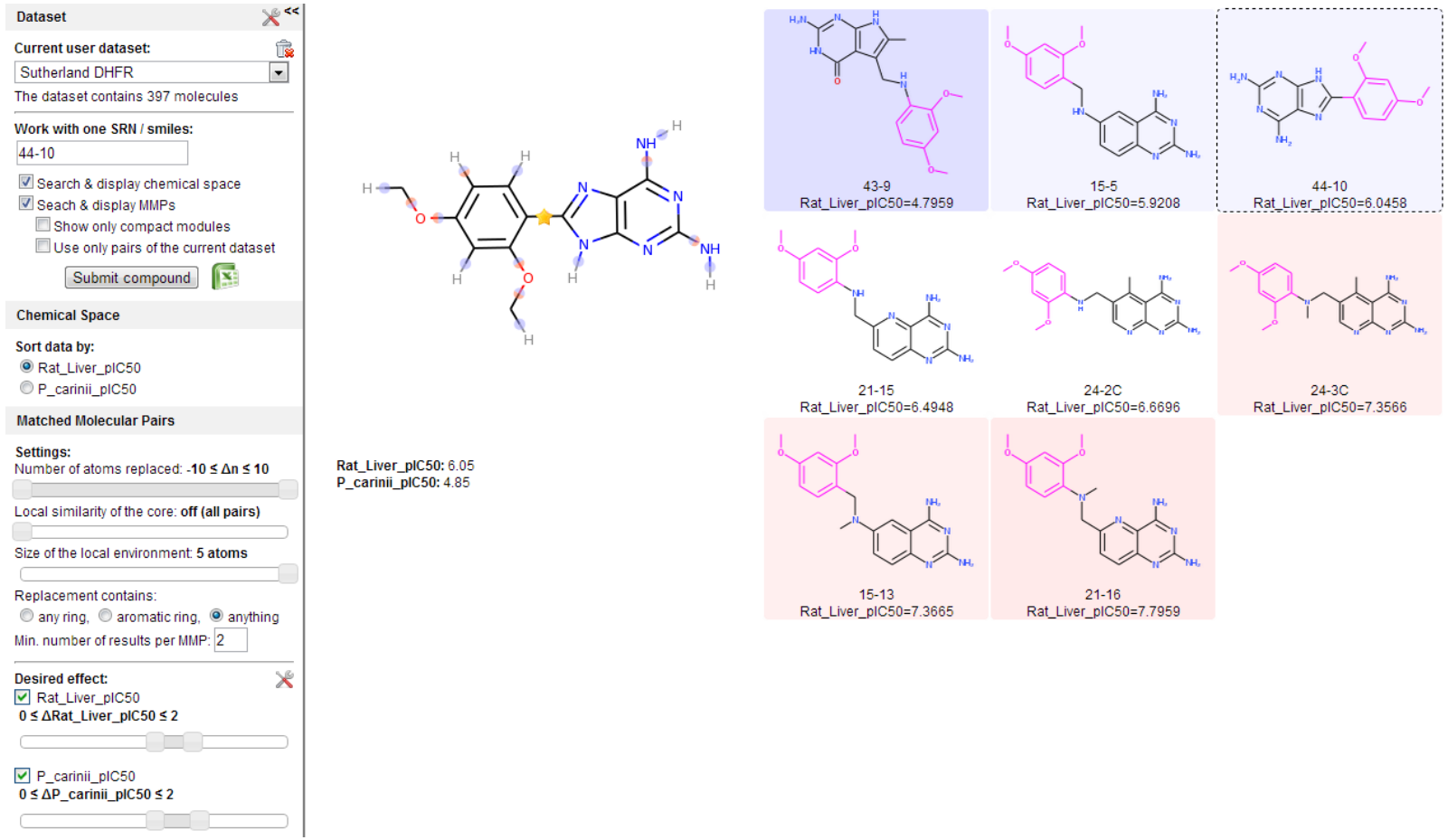
LUCID web interface

Interactive selection of the site of transformation



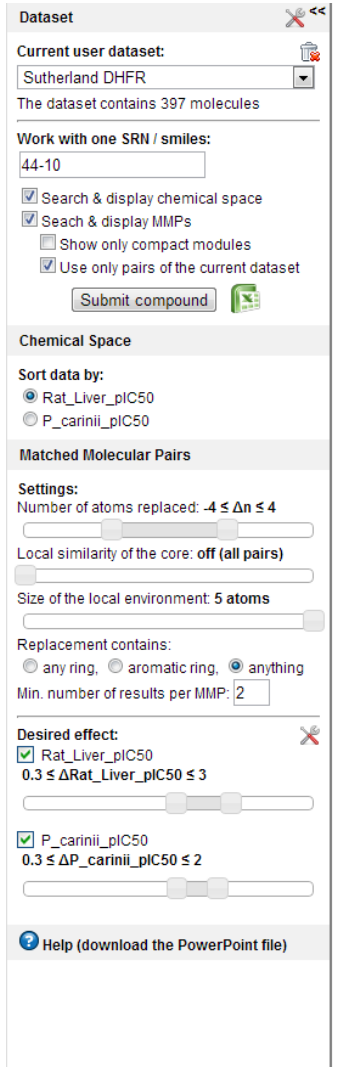
LUCID web interface

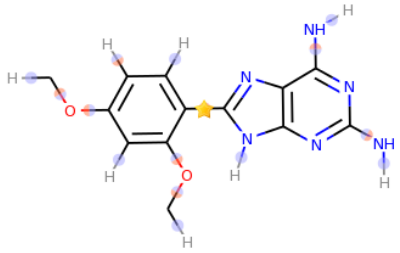
Chemical space around one constant fragment, colored by the selected property



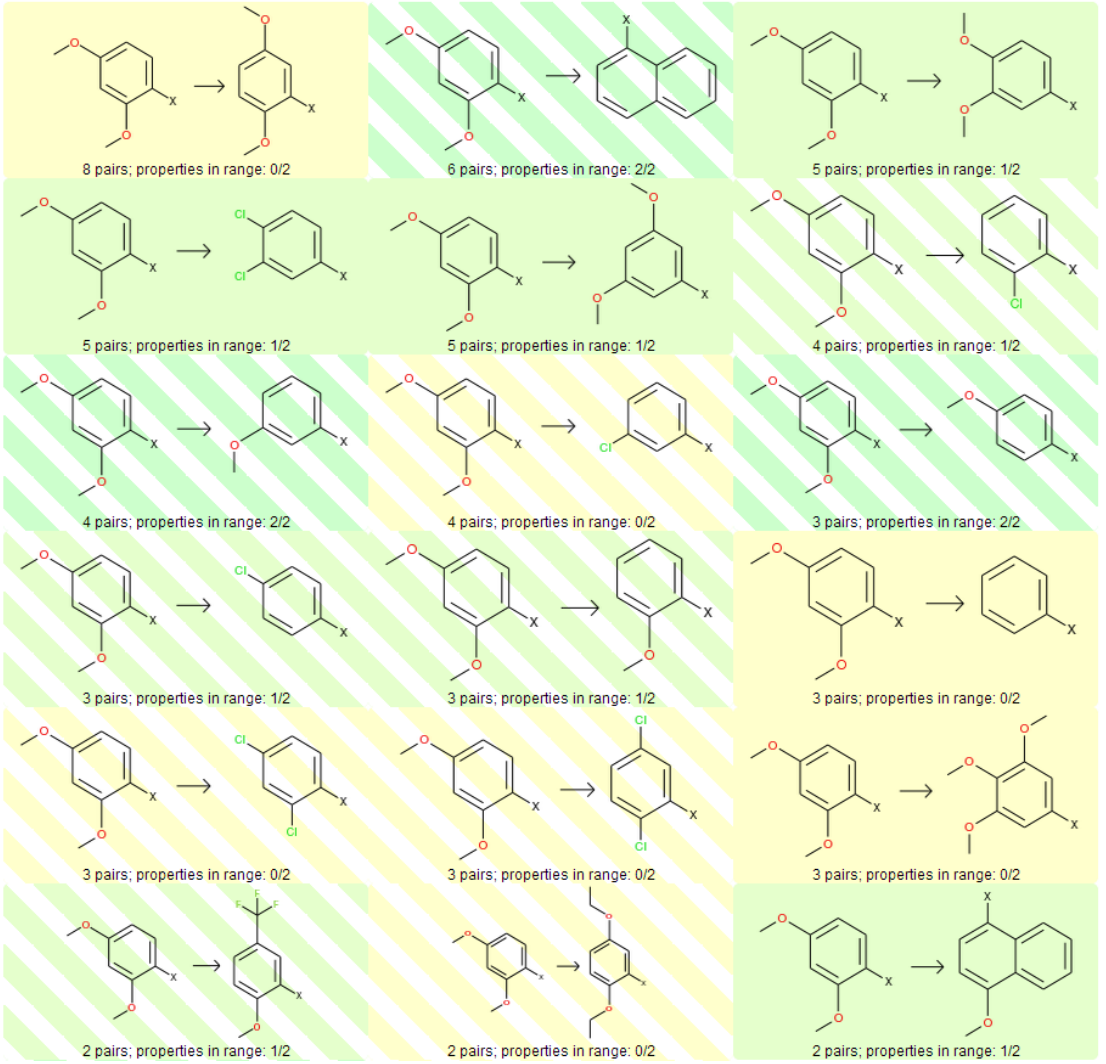
LUCID web interface

Transformations at the selected site, colored by the number of properties within the selected range



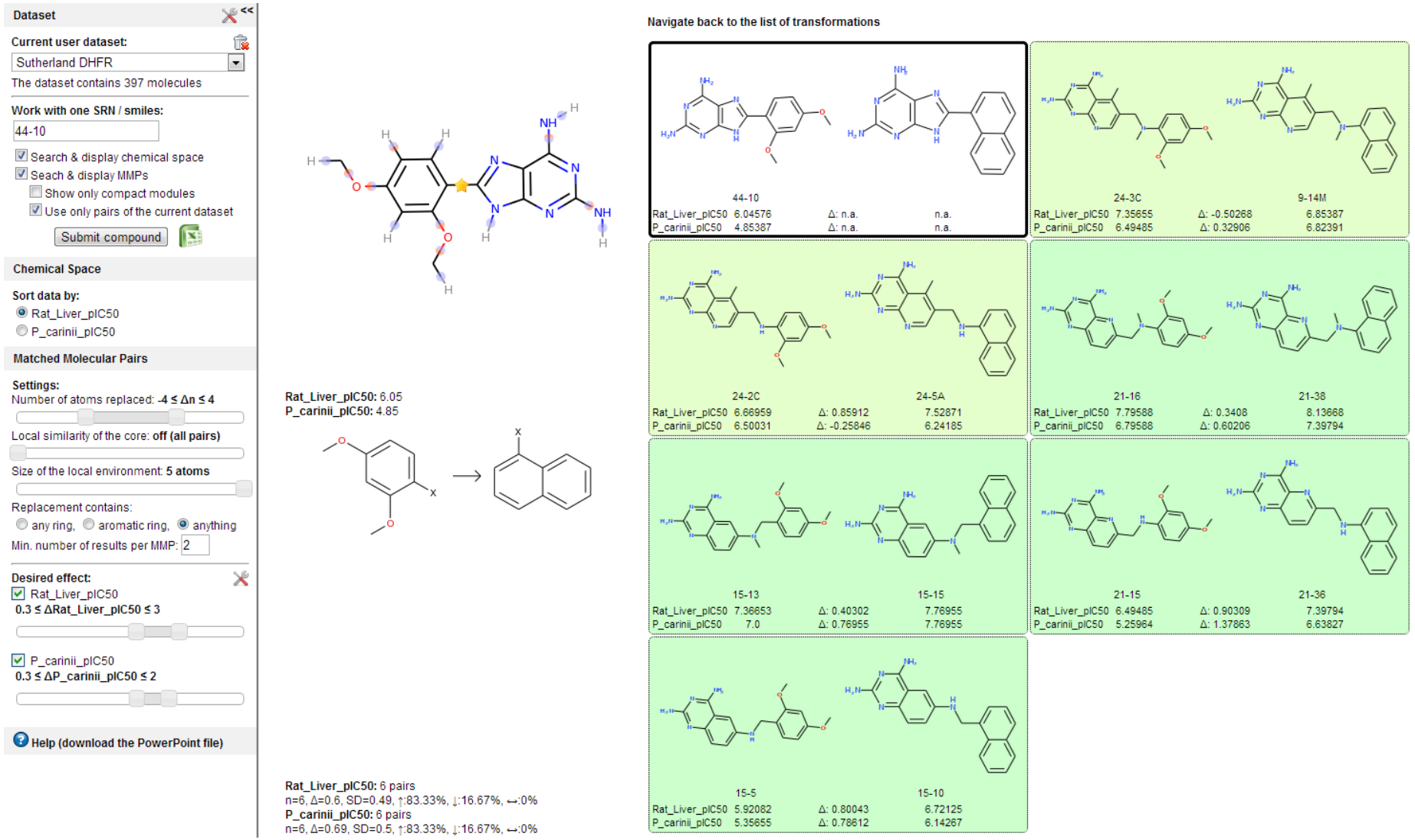


Rat_Liver_pIC50: 6.05
P_carinii_pIC50: 4.85



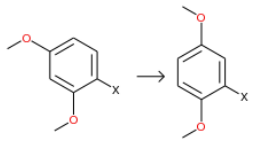
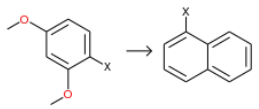
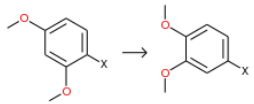
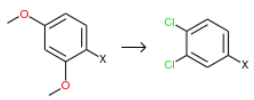
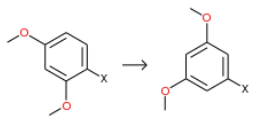
LUCID web interface

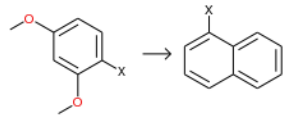
Pairs of compounds (MMPs) corresponding to the selected transformation



Excel export

Transformations (left) and MMPs of one particular transformation (right)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1			Local Similarity		user_Rat_Liver_pIC50							user_P_carinii_pIC50						
2	Transformation	Transform	Size	Strength	N	Median	StdErr	StdDev	Increase	Decrease	Neutral	N	Median	StdErr	StdDev	Increase	Decrease	Neutral
3		44-7	5	0	8	-0.3	0.14	0.4	2	5	1	8	0.021	0.15	0.43	3	3	2
4		Nc1nc2[nH	5	0	6	0.6	0.2	0.49	5	1	0	6	0.69	0.2	0.5	5	1	0
5		44-9	5	0	5	0.77	0.51	1.1	4	1	0	5	0.25	0.16	0.36	3	0	2
6		44-11	5	0	5	0.02	0.52	1.2	2	2	1	5	0.51	0.22	0.49	3	1	1
7		44-8	5	0	5	0.15	0.37	0.83	3	1	1	5	0.3	0.19	0.42	4	0	1

	A	B	C	D	E	F	G	H	I
1				user_Rat_Liver_pIC50			user_P_carinii_pIC50		
2	Compound1	Compound2	Local Simi	Value1	Value2	Delta	Value1	Value2	Delta
3	44-10	Nc1nc2[nH]c(-c3cccc	1	6.04576	n.a.	n.a.	4.85387	n.a.	n.a.
4	24-3C	9-14M	0.17	7.35655	6.85387	-0.50268	6.49485	6.82391	0.32906
5	24-2C	24-5A	0.17	6.66959	7.52871	0.85912	6.50031	6.24185	-0.25846
6	21-16	21-38	0.17	7.79588	8.13668	0.3408	6.79588	7.39794	0.60206
7	15-13	15-15	0.17	7.36653	7.76955	0.40302	7	7.76955	0.76955
8	21-15	21-36	0.17	6.49485	7.39794	0.90309	5.25964	6.63827	1.37863
9	15-5	15-10	0.17	5.92082	6.72125	0.80043	5.35655	6.14267	0.78612

Doing now what patients need next