

# PERFORMANCE OF SUB-PHARMACOPHORE MODELS AS SEEDS IN DRUG DISCOVERY

Workshop in  
Cheminformatics  
Sep 2014  
EPFL Lausanne

Modest Korff, Joel Freyss, Thomas Sander



# TOPICS

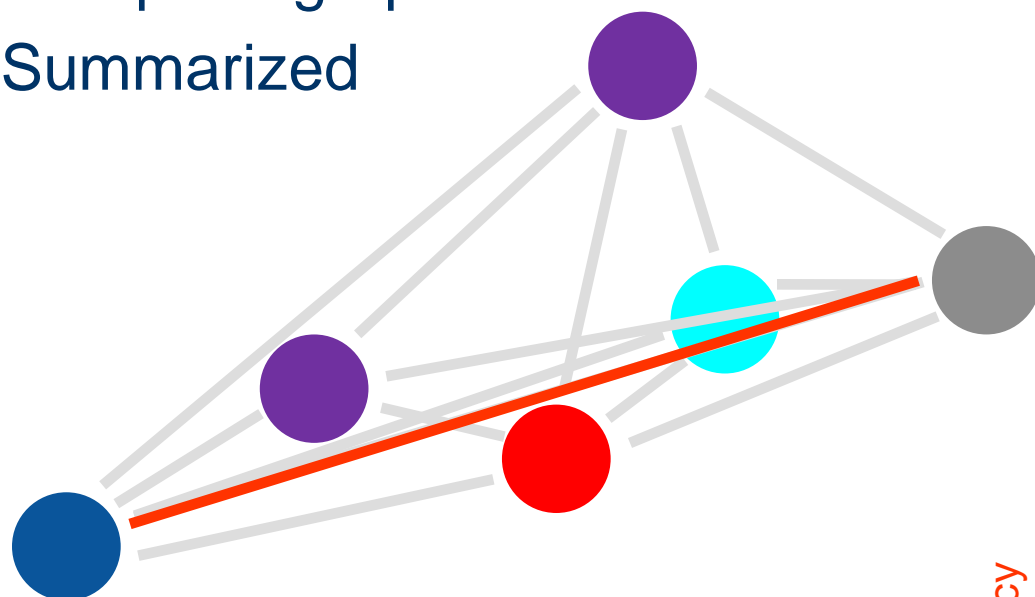
---

- ▶ **Task**
- ▶ **Flexophore descriptor**
- ▶ **Sub-Flexophores**
- ▶ **Improving pharmacophore models**
- ▶ **Results and conclusions**

## **Improve pharmacophore models for successful virtual screening**

# THE FLEXOPHORE DESCRIPTOR

Complete graph  
Summarized



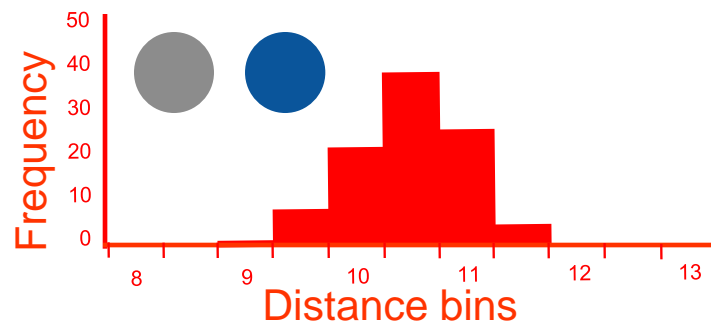
Nodes:

Pharmacophore points

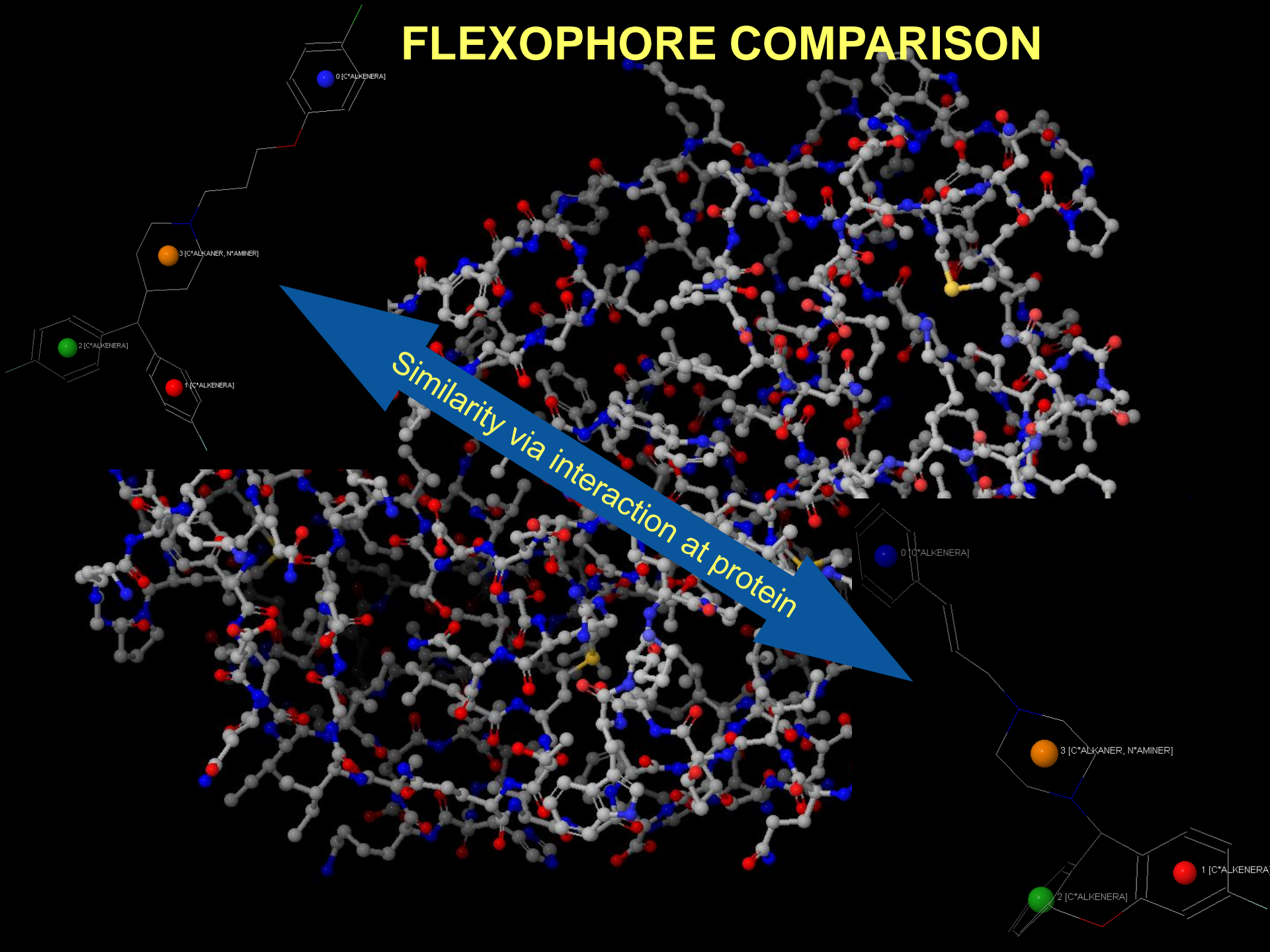


Edges:

Distance histograms

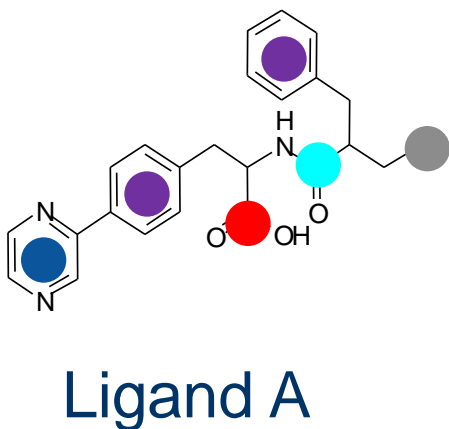


# FLEXOPHORE COMPARISON

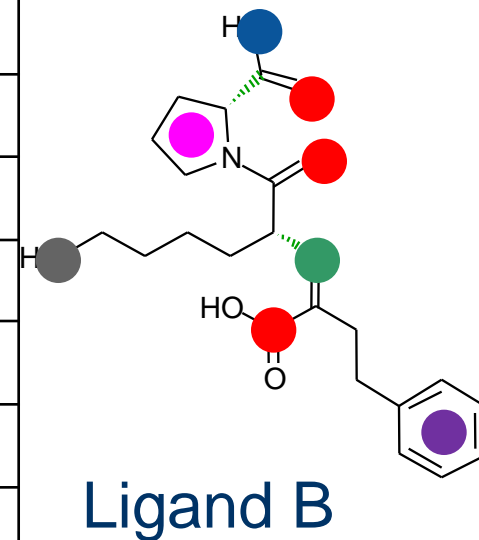


# SIMILARITY METRIC FOR BIOISOSTERS

## a Protein-ligand interaction statistics



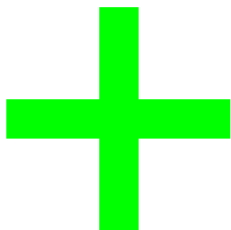
PP Types	1.	2.	3.	4.	5.	6.
1. C Alkane (prim.)	1.00	0.72	0.83	0.36	0.37	0.30
2. C Alkane (sec.)	0.72	1.00	0.51	0.38	0.40	0.38
3. C Carbonyl	0.83	0.51	1.00	0.41	0.45	0.45
4. O Carbonyl	0.36	0.38	0.41	1.00	0.98	0.88
5. O Ether	0.37	0.40	0.45	0.98	1.00	0.85
6. O Alcohol	0.30	0.38	0.45	0.88	0.85	1.00



$$\text{Sim}_{A,B} = f(\text{Pharmacophore point pairs bioisosterism})$$

# HOW TO IMPROVE A PHARMACOPHORE MODEL?

---

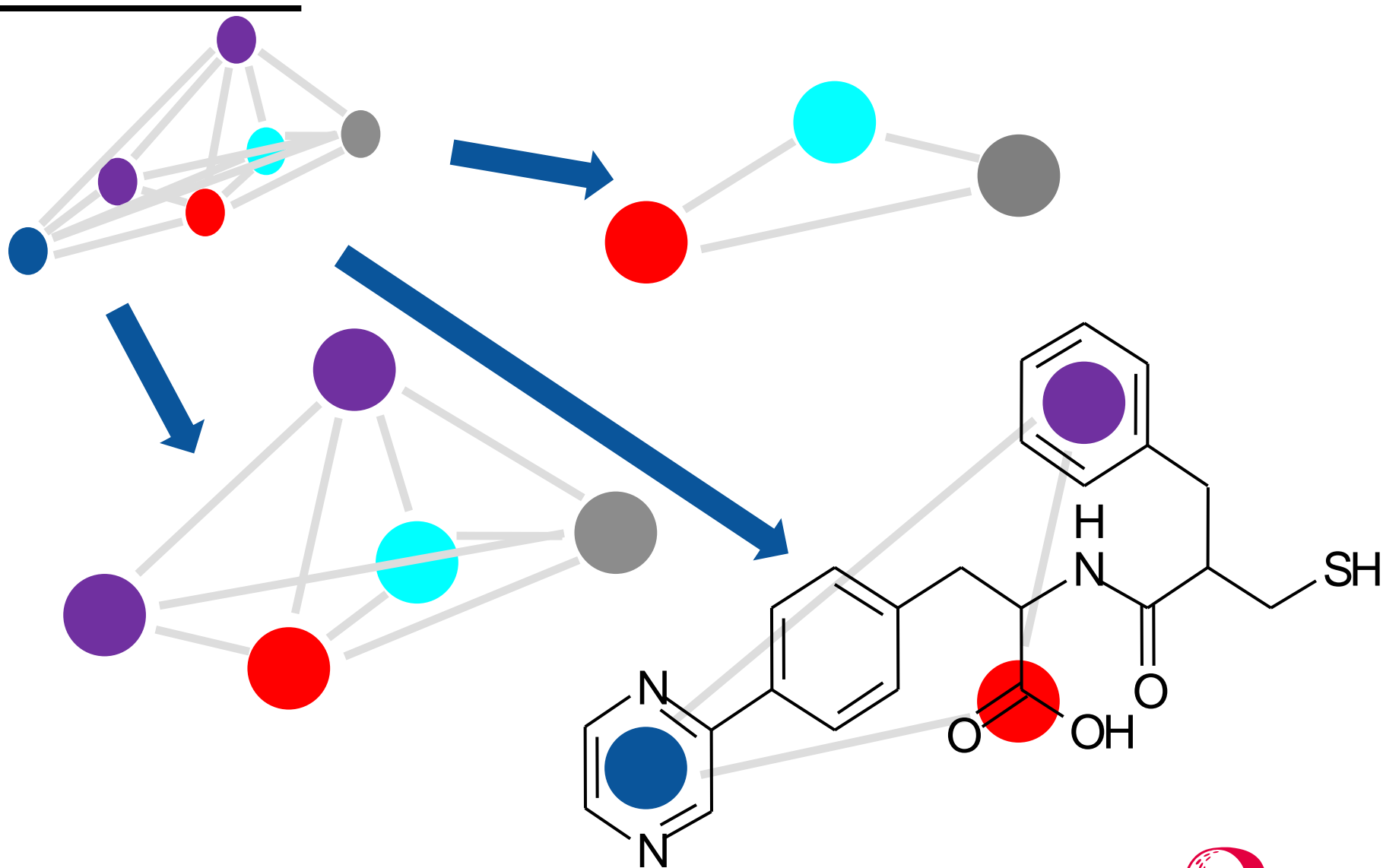


Two changeable components

- ▶ Pharmacophore points
  - Substructure based
  - Easy to manipulate on/off
- ▶ Distance histograms
  - Edges of complete graph  $((n * n) - n) / 2$
  - I.e. 7 PPNodes → 21 distance histograms
  - Histograms, not ranges
  - Meaningful manipulation difficult



# SUB PHARMACOPHORES





# COMBINATORIAL EXPLOSION

How many subpharmacophores?

Possible combinations  $\binom{n}{k}$

# PPoints in query molecule

# PPoints in Sub Pharmacophore	PPoints	4	6	8	10	12	14
	3	4	20	56	120	220	364
5		6	56	252	792	2002	
7			8	120	792	3432	
9				10	220	2002	
11					12	364	

# Possible combinations

# HOW TO MEASURE IMPROVEMENT?

---

## Need for validation

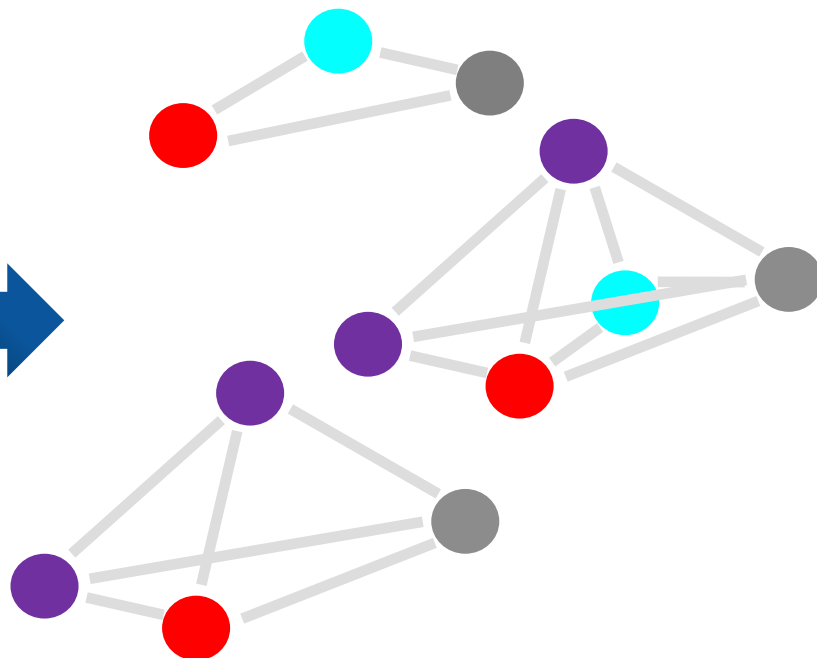
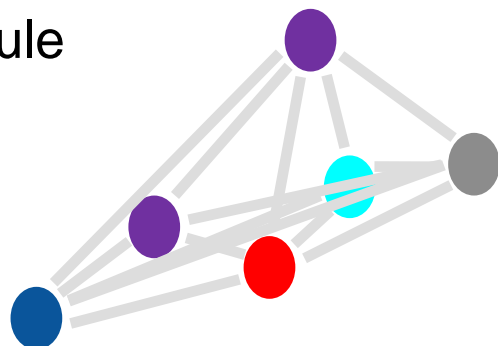
### DUD dataset

- ▶ 39 targets
- ▶ 241 data sets
- ▶ 10-50 seeds (actives)
- ▶ Filled up to 1000 with decoys (assumed inactives)
  - Physico chemical properties like seeds

Huang, N.; Shoichet, B. K.; Irwin, J. J., Benchmarking Sets for Molecular Docking. *J. Med. Chem.* **2006**, *49*, (23), 6789-6801.S

# TEST SETUP

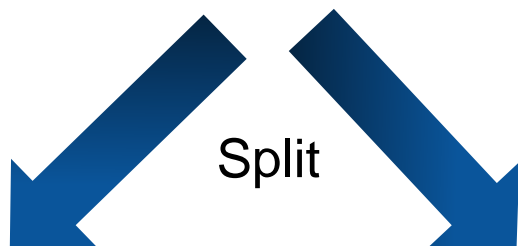
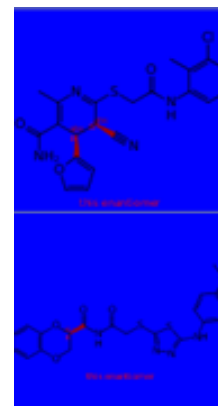
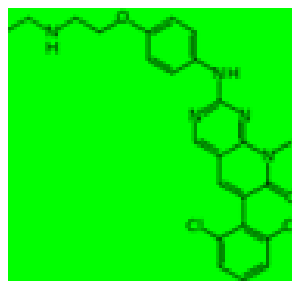
Probe molecule



▶ Seeds (actives)

▶ Decoys (inactives)

} 1000



Split

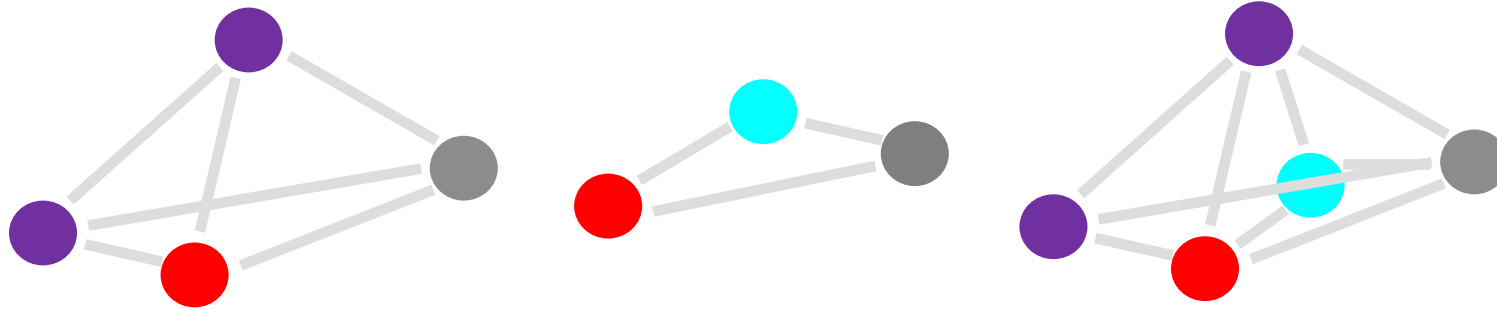
Train (500)

Test (500)

# RUNNING A TEST

---

Probe SubFlexophores



Compare SubFlexophores with train Flexophores

Compare best performing SubFlexophores with test Flexophores

Calculate enrichment

# RESULTS ALL TESTS

---

- ▶ Data sets 241
- ▶ Probe SubFlexophores: 20-200 per probe molecule
- ▶ Compare to train
- ▶ Top SubFlexophores 1-10 per probe molecule
- ▶ Compare to test

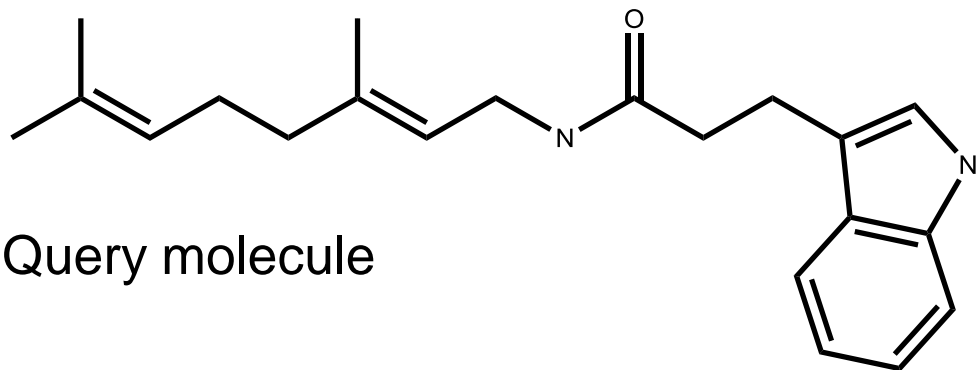
## Summary

**Enrichment SubFlexophore test                      127**

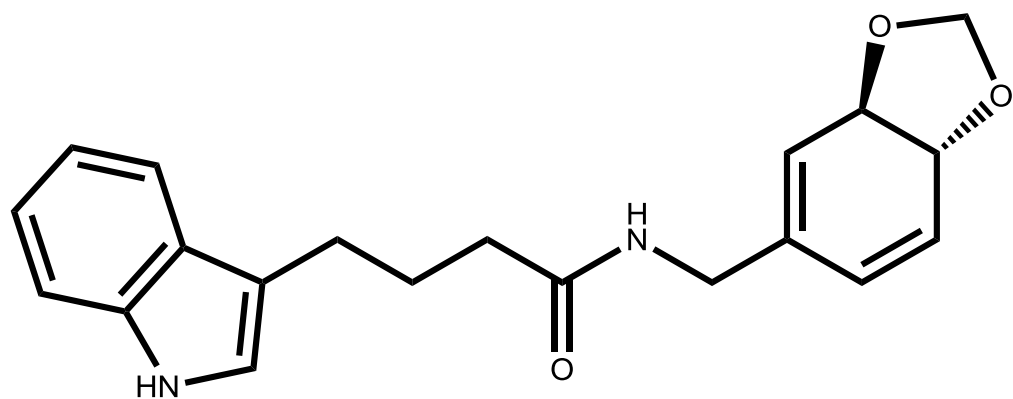
**Enrichment full Flexophore test                      86**

**Best SubFlexophore better than full Flexophore                      91**

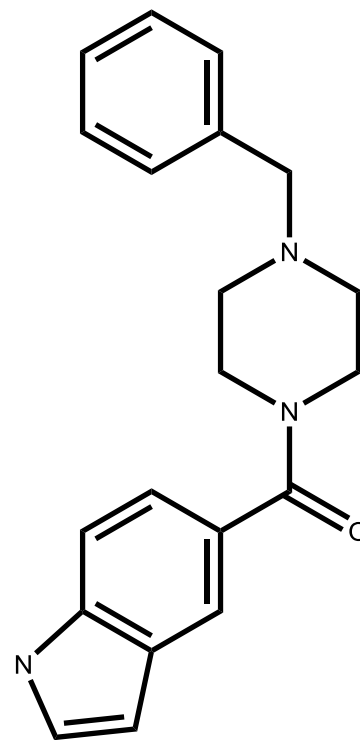
# SUCCESSFUL SCAFFOLD HOPPING



Query molecule



Found seeds



# FLEXOPHORE EDITOR: PLAYGROUND FOR SCIENTISTS

FlexophoreEditor

File Sub-Flexophore Examples Training Help

#	Name	Enrichment
1	Sub1.0	70
2	Sub054	70
3	Sub027	70
4	Sub1.60	60
5	Sub1.44	60
6	Sub1.27	60
7	Sub1.19	60
8	Sub1.12	60
9	Sub1.02	60
10	Sub097	60
11	Sub010	60
12	Sub1.62	50
13	Sub1.58	50
14	Sub1.59	50
15	Sub1.56	50
16	Sub1.55	50
17	Sub1.54	50
18	Sub1.47	50
19	Sub1.46	50
20	Sub1.38	50
21	Sub1.31	50
22	Sub1.29	50
23	Sub1.28	50
24	Sub1.26	50
25	Sub1.19	50
26	Sub1.15	50
27	Sub084	50
28	Sub075	50
29	Sub076	50
30	Sub072	50
31	Sub016	50
32	Train0001.0	50
33	Sub1.52	40
34	Sub1.45	40
35	Sub1.35	40
36	Sub1.34	40
37	Sub1.33	40
38	Sub1.23	40
39	Sub1.03	40
40	Sub085	40
41	Sub093	40
42	Sub081	40
43	Sub070	40
44	Sub068	40
45	Sub066	40
46	Sub052	40
47	Sub045	40
48	Sub043	40
49	Sub028	40
50	Sub017	40
51	Sub1.42	30
52	Sub1.13	30
53	Sub089	30
54	Sub033	30
55	Sub024	20
56	Sub003	20
57	Sub012	20
58	Sub008	20

Filter Min-max distance [Å]  
10 [ ] 50

Flexophore Editor - written by Modest Korff - Actelion ...

ar\_ligandsAllFlexophoreViz  
Molecules 1001  
Decoys 970  
Seeds 30  
Test 1

Molecule	Similarity	Id
	0.807	15
	0.777	370
	0.766	9
	0.731	18
	0.724	7
	0.705	208
	0.700	2
	0.690	17
	0.680	1
	0.655	654
	0.642	0

Enrichment 70.0(1%)

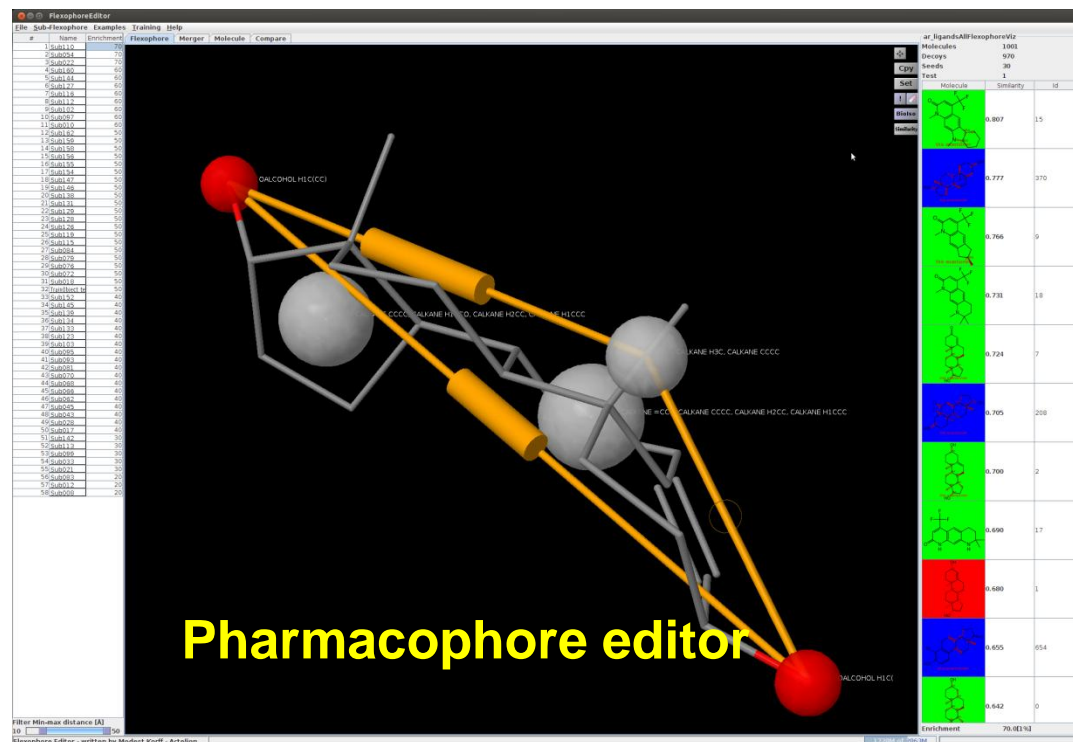
1/250M of 2963M

# IMPLEMENTATION DETAILS

- ▶ In-house development
- ▶ Java programming language
- ▶ Jmol for pharmacophore visualization
- ▶ Multi thread similarity calculations

Training data

SubFlexophores





# CONCLUSIONS

---

- ▶ Impossible to outsmart combinatorial explosion of possible pharmacophore models
- ▶ Medicinal chemist may construct test and validation data
- ▶ Combination of brute force computer power plus medicinal chemist know how resulted in improved pharmacophore models

# THANK YOU!

