



Hit expansion approaches using combinations of bioisosteric transformations, similarity searches and generalized substructure searches

Andreas Bergner, PhD
Principal Scientist, Computational Chemistry

MipTec Basel
22 September 2010



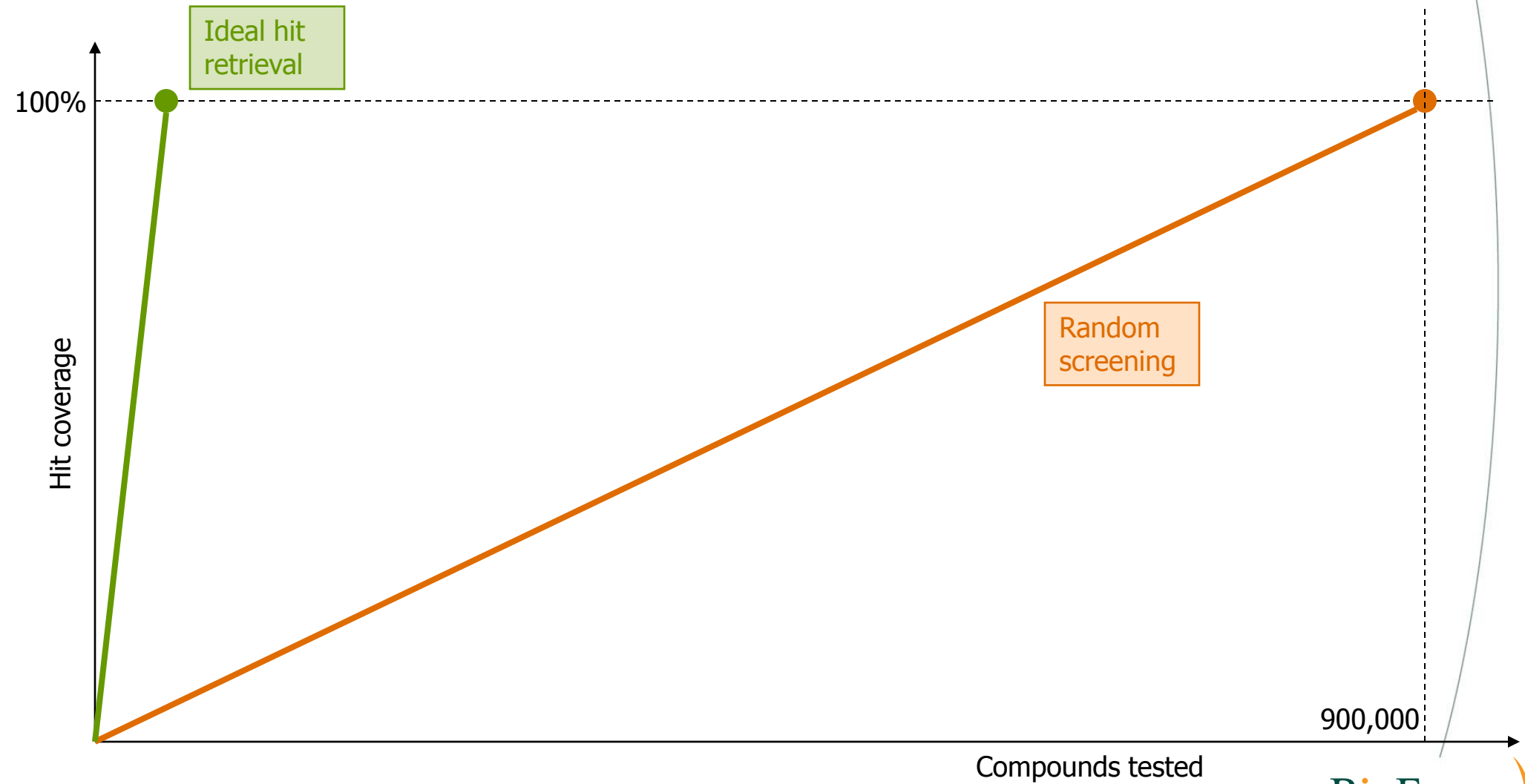


Outline

- Screening paradigms and hit expansion
- Demonstration data set and visualization
- Methods
 - fingerprint-based similarity searching and Turbosim
 - bioisosteric transformations and enumerations
 - ring generalization
 - ring assembly modifications
- Results

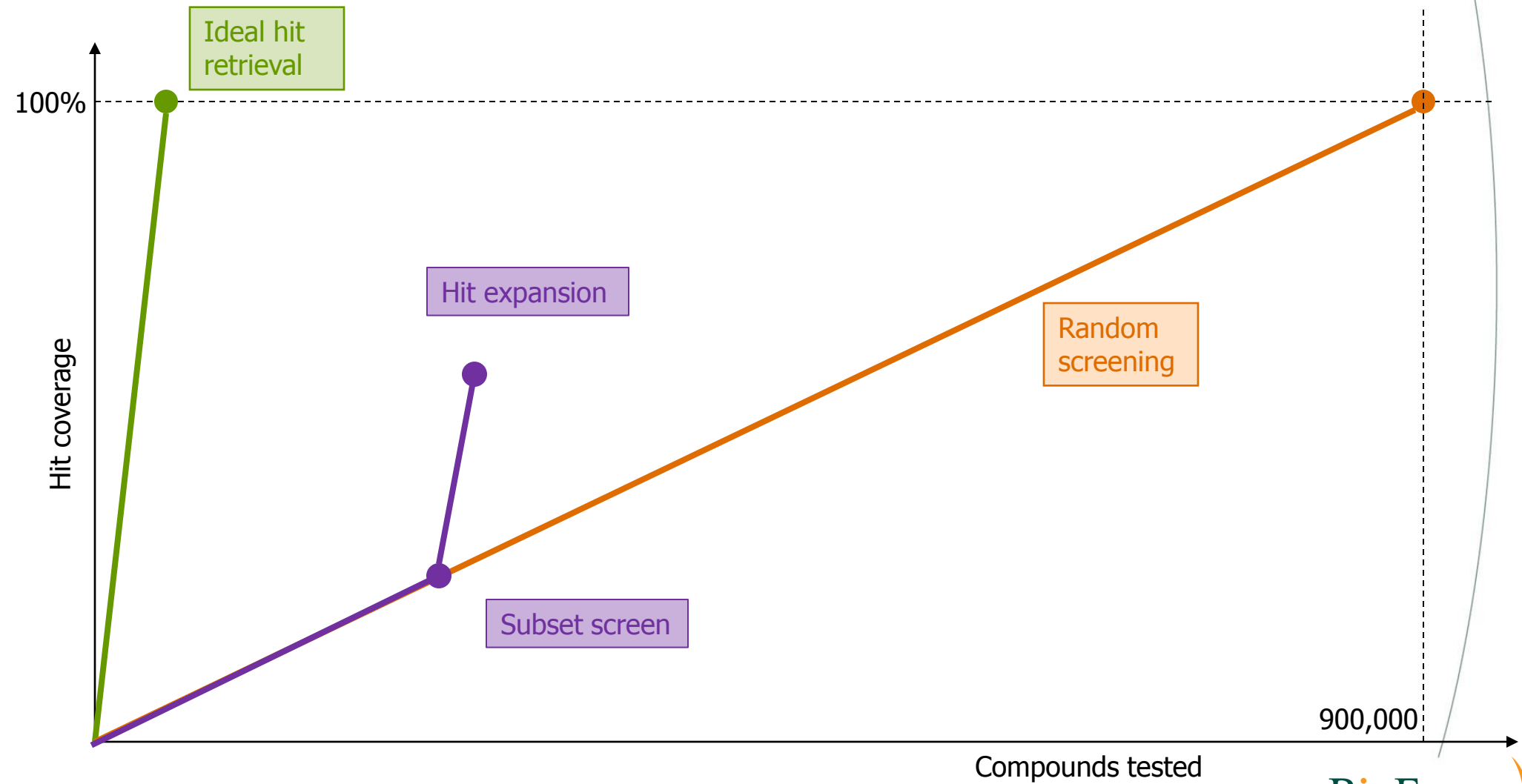


Screening paradigms



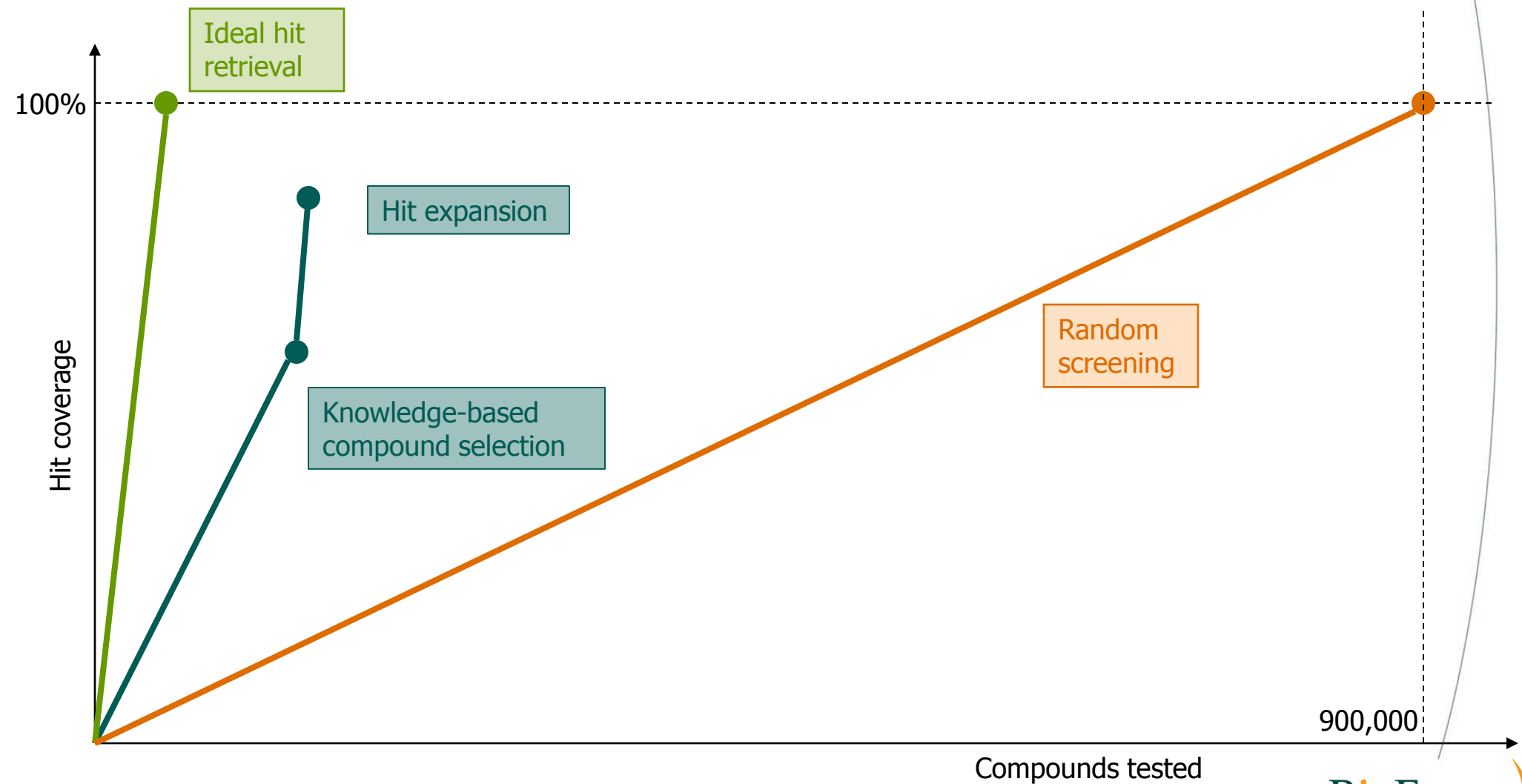


Screening paradigms





Screening paradigms





Hit expansion

- Consolidate and expand knowledge about active regions in chemical space
 - find derivatives of active scaffolds and establish or broaden SAR information
 - 'early' scaffold hopping
 - strengthen knowledge base for direct use in H2L campaign
 - distill and exploit knowledge about chemical features that contribute to activity
 - rescue active scaffolds from low-activity data



Outline

- Screening paradigms and hit expansion
- **Demonstration data set and visualization**
- Methods
 - fingerprint-based similarity searching and Turbosim
 - bioisosteric transformations and enumerations
 - ring generalization
 - ring assembly modifications
- Results



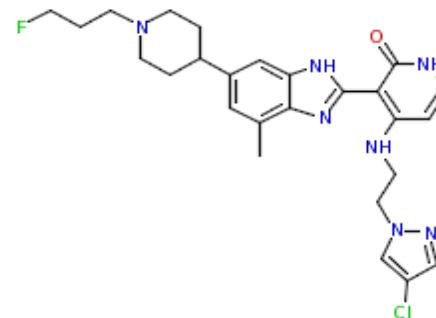
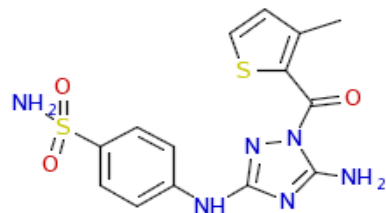
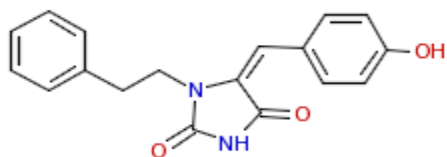
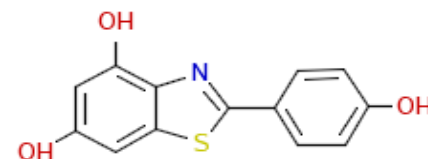
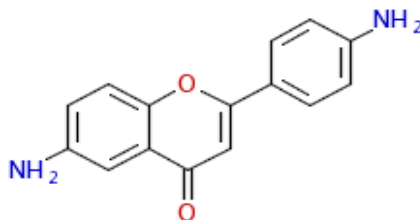
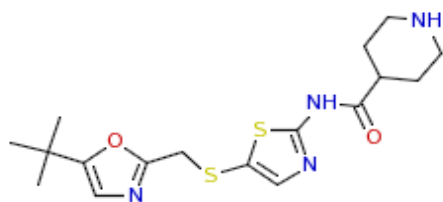
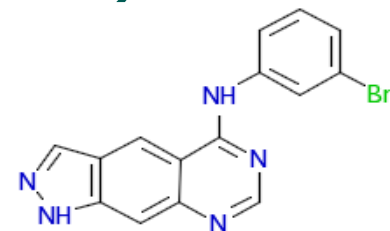
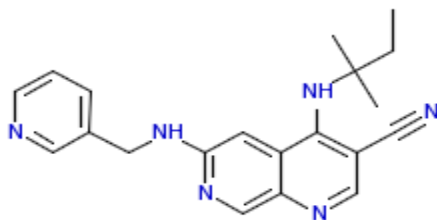
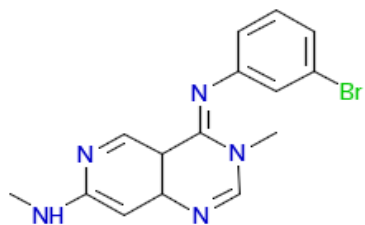
Demonstration data set

- ChEMBL database
 - www.ebi.ac.uk/chembl
- EGFR (Tyr kinase)
 - 2863 compounds with published IC50 value
 - 962 Murcko scaffolds
 - 1142 clusters (MACCS/T=0.8)
 - 738 clusters (MACCS/T=0.7)
 - large activity range (sub-nM to mM)



Demonstration data set

Selected set of active compounds (templates)

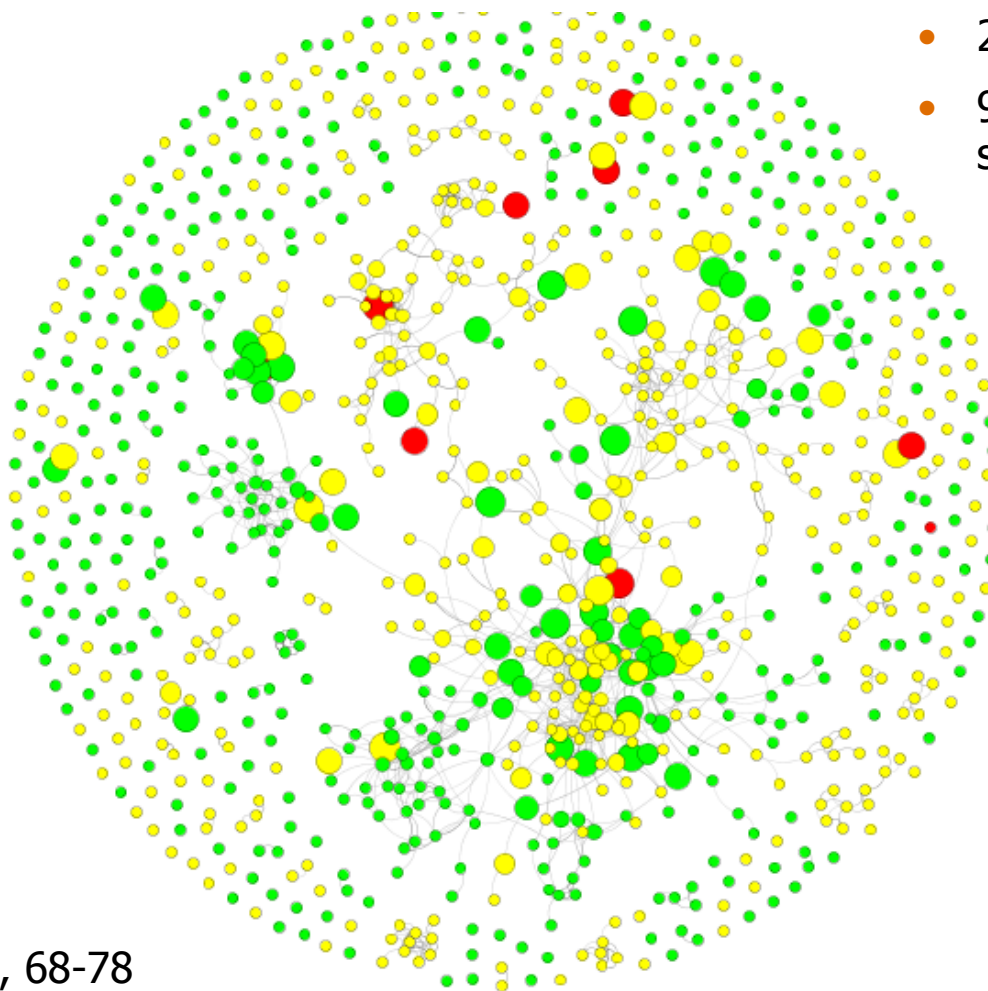


Neighbors in chemical space

Network-like similarity graphs

Template
Active
Inactive

- 9 templates
- 2'854 other actives
- 900'000 BioFocus screening collection



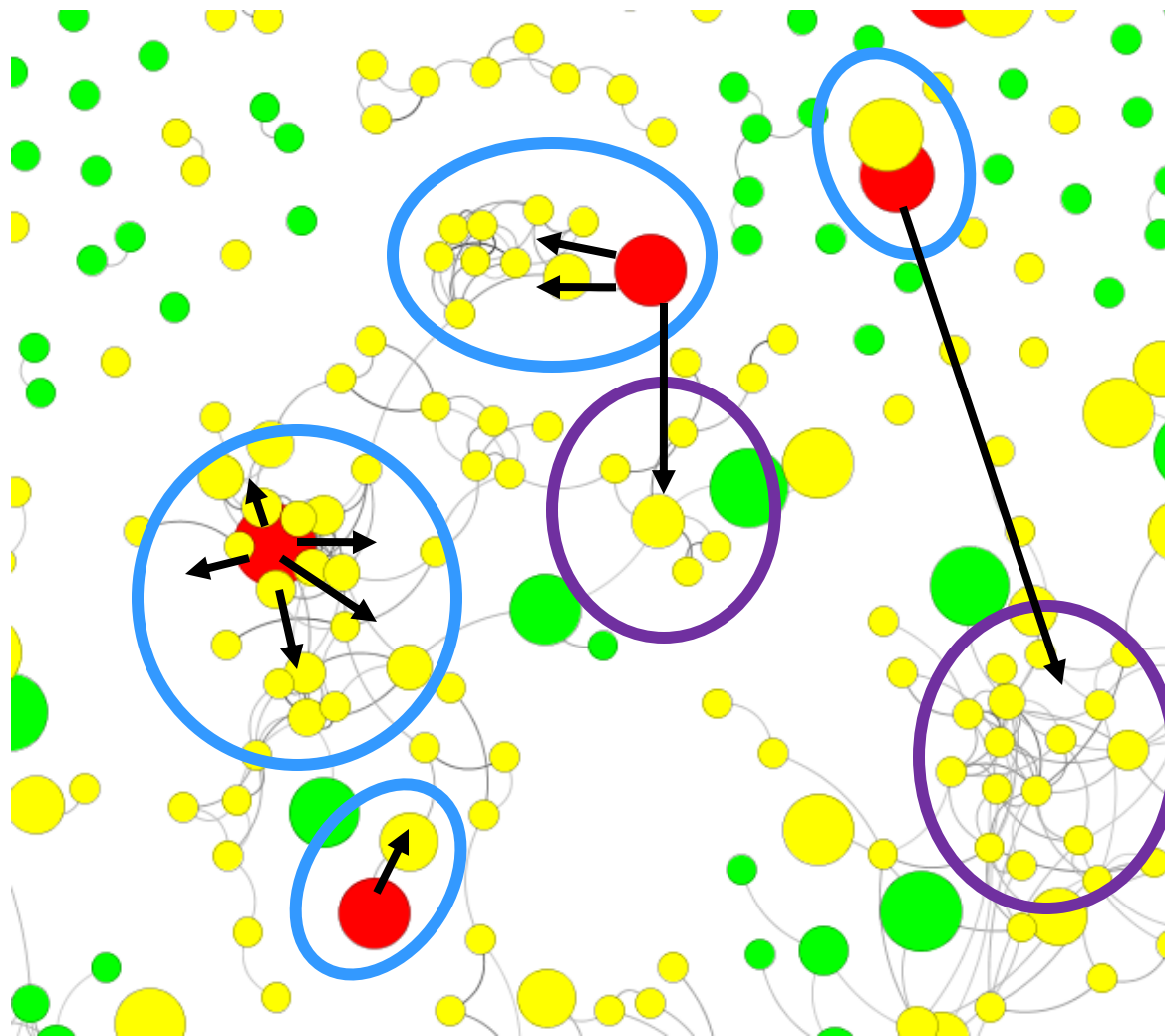
- Saranea:
Loukine et al,
J.Chem.Inf.Mod., 2010, 68-78



Neighbors in chemical space

Probing chemical space from active points

Template
Active
Inactive



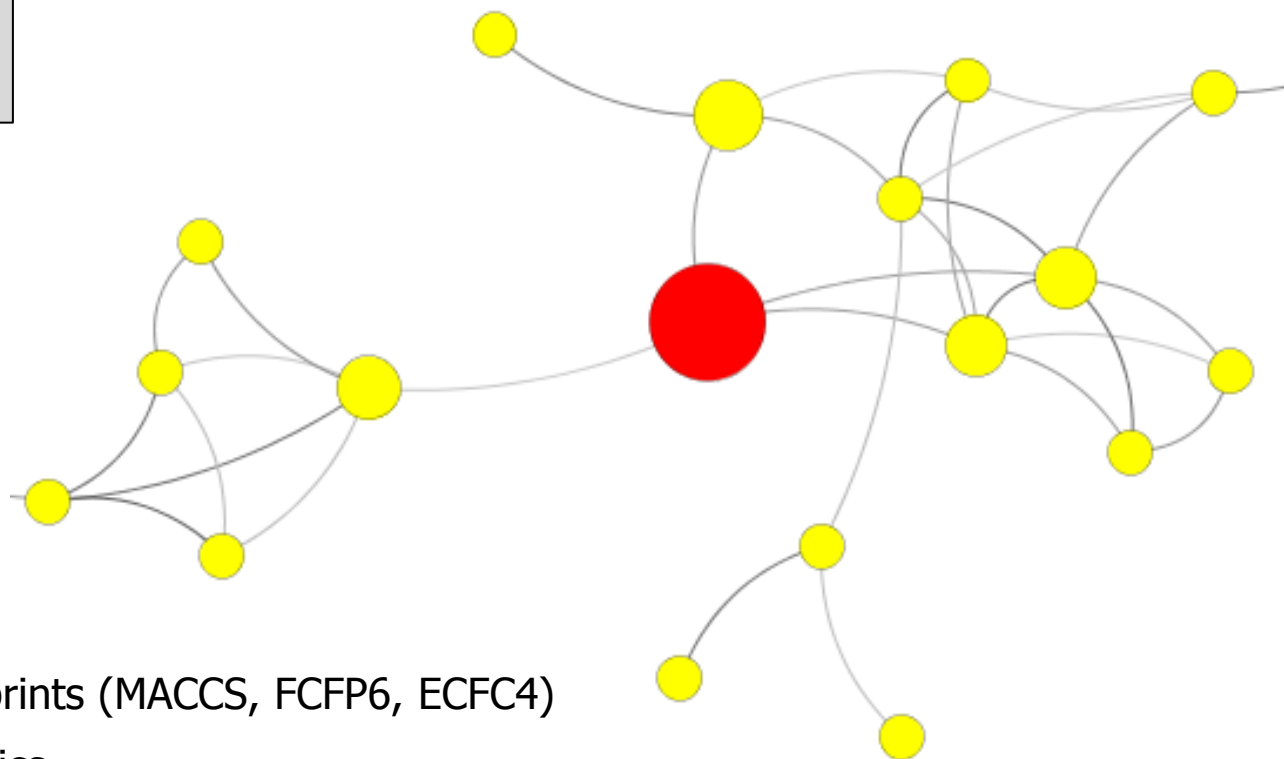


Outline

- Screening paradigms and hit expansion
- Demonstration data set and visualization
- **Methods**
 - fingerprint-based similarity searching and Turbosim
 - bioisosteric transformations and enumerations
 - ring generalization
 - ring assembly modifications
- Results

Fingerprint-based similarity searches

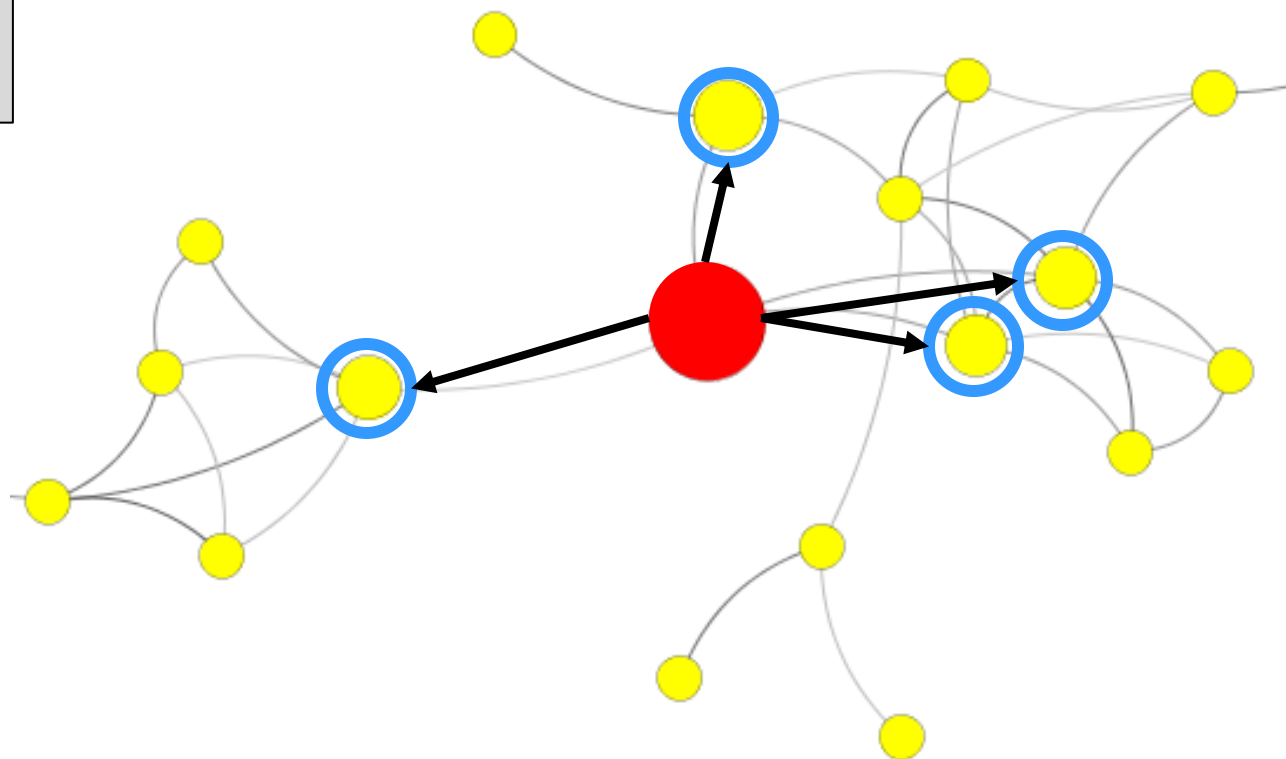
Template
Active



- Various fingerprints (MACCS, FCFP6, ECFC4)
- Tanimoto metrics

Turbosim

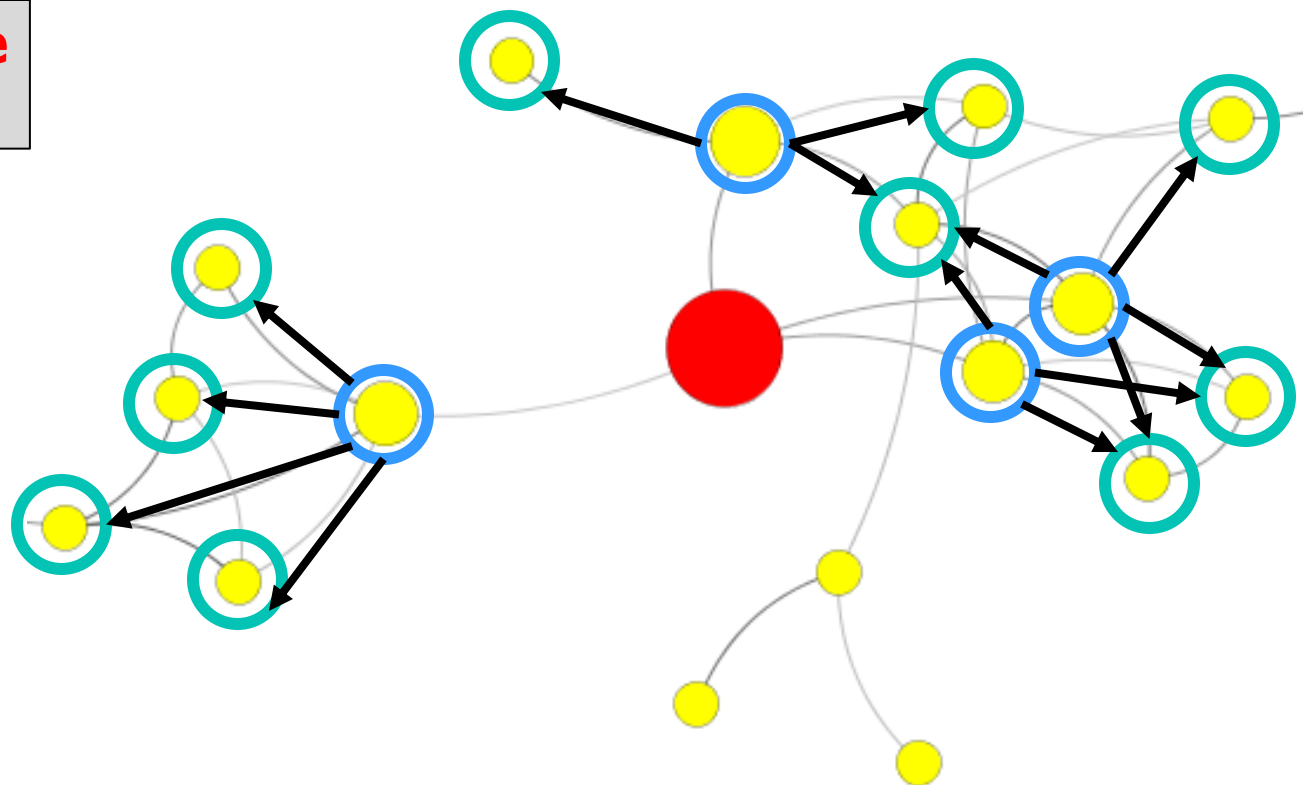
Template
Active



- Hert et al., J.Med.Chem., 2005, 7049-7054

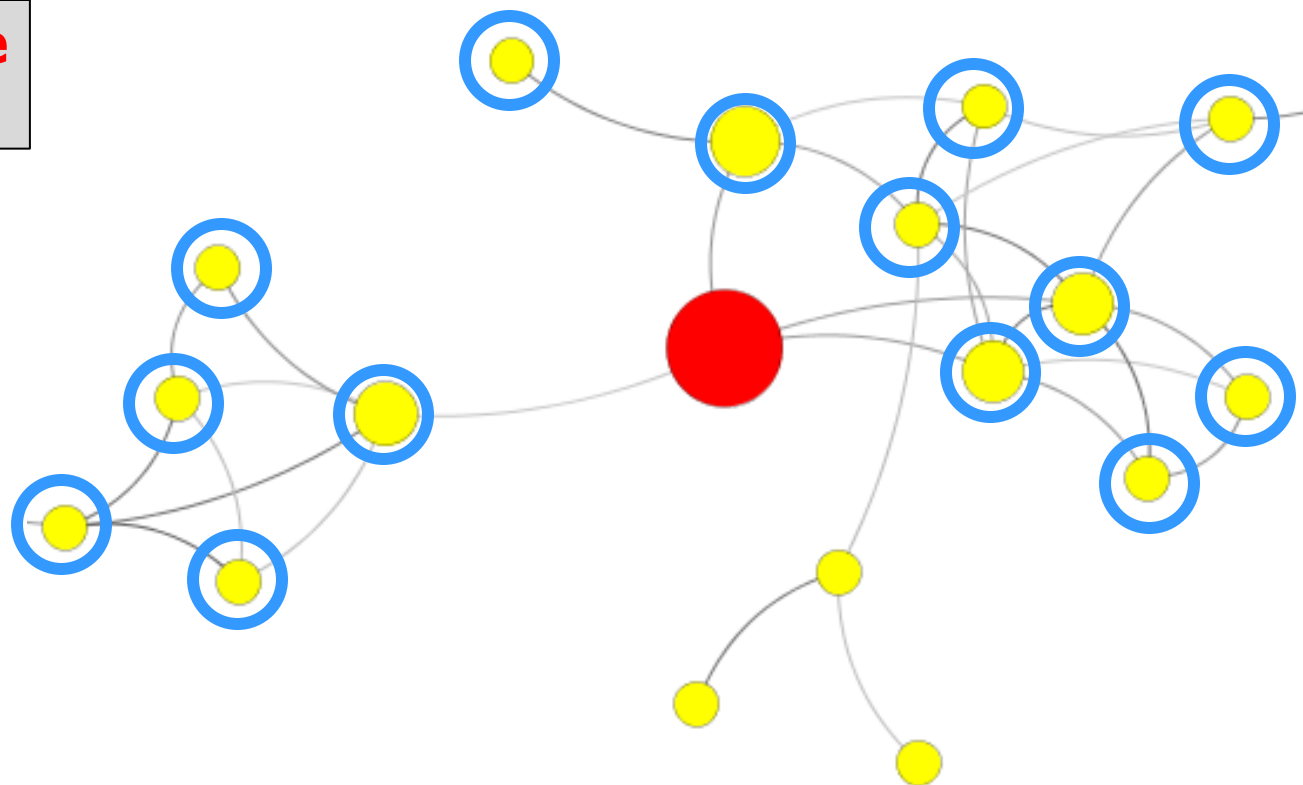
Turbosim

Template
Active



Turbosim

Template
Active





Outline

- Screening paradigms and hit expansion
- Demonstration data set and visualization
- **Methods**
 - fingerprint-based similarity searching and Turbosim
 - bioisosteric transformations and enumerations
 - ring generalization
 - ring assembly modifications
- Results



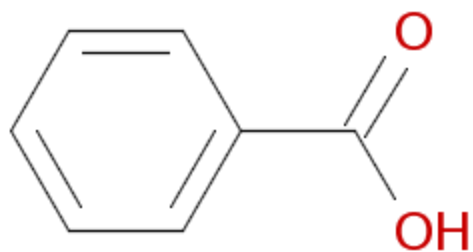
Extend similarity searches

- Like Turbosim: carry out two- or multi-step searches
- Requirements
 - Fast
 - Highly automated
 - Generally applicable
 - simple 2D methods
 - no 3D methods (3D re-core, pharmacophores)
- Bioisosteres: Alternative concept of similarity

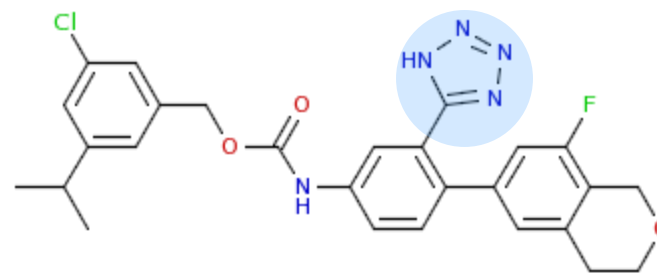
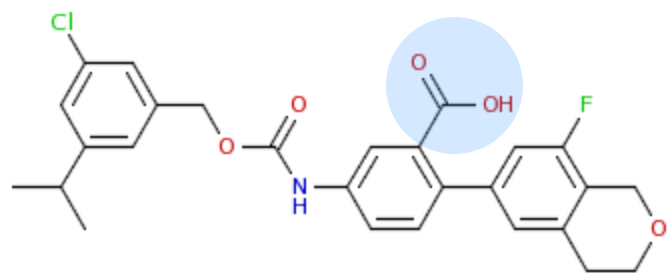
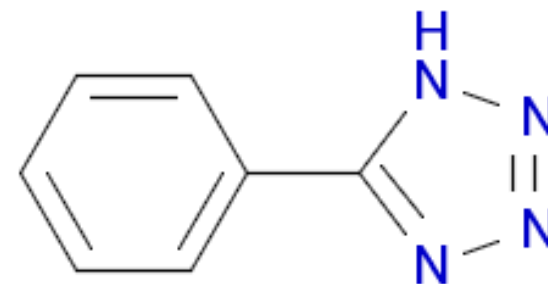
- Sheridan, J.Chem.Inf.Comput.Sci, 2002, 103-108



Bioisosteres and similarity

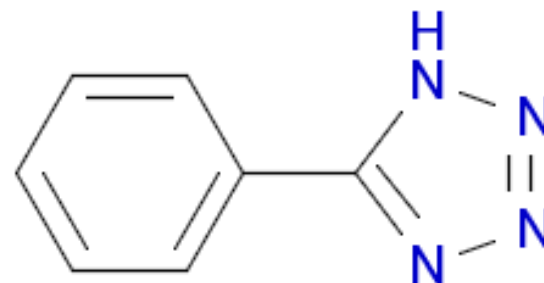
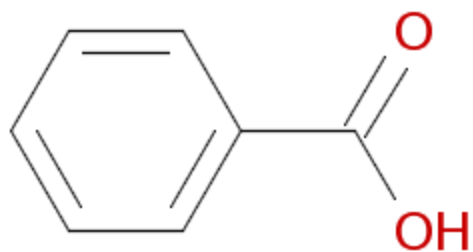


- T=0.11 (MACCS)
- T=0.25 (FCFP_6)



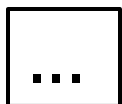
- T=0.61 (MACCS)
- T=0.92 (FCFP_6)

Bioisosteres and similarity

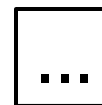
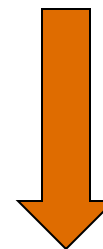


- Virtualize template using bioisosteric transformation

- Similarity searches
- Turbosim

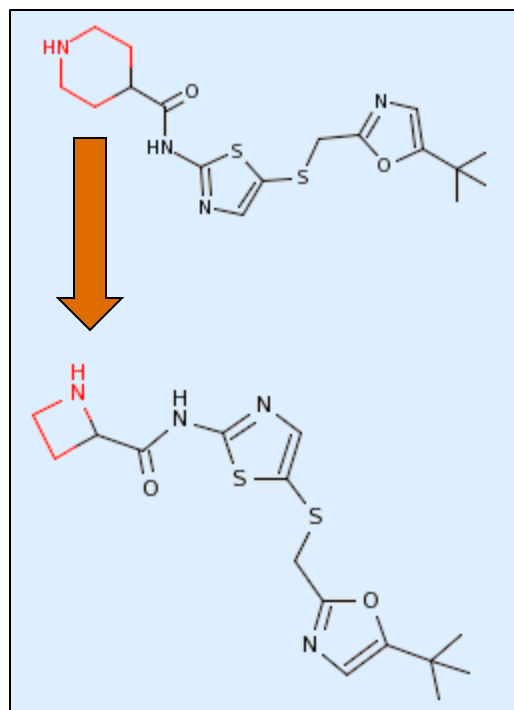
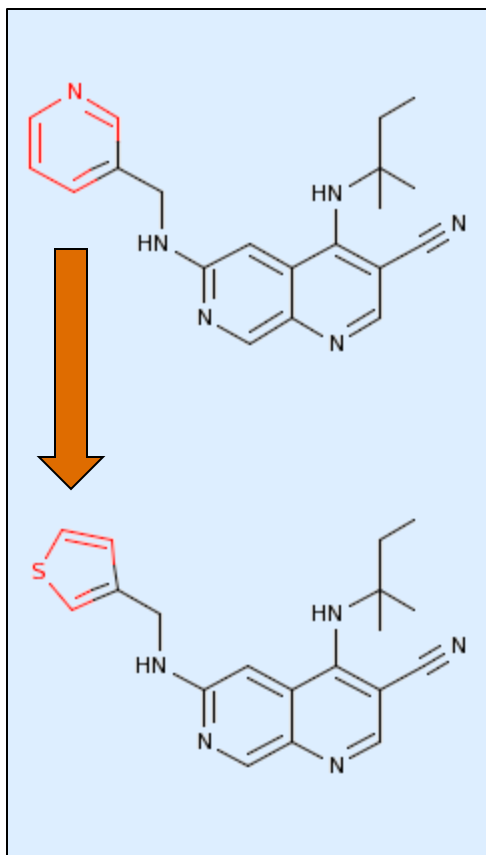


- Similarity searches
- Turbosim

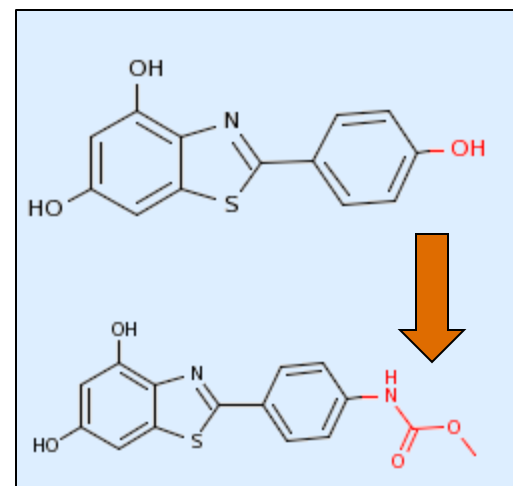
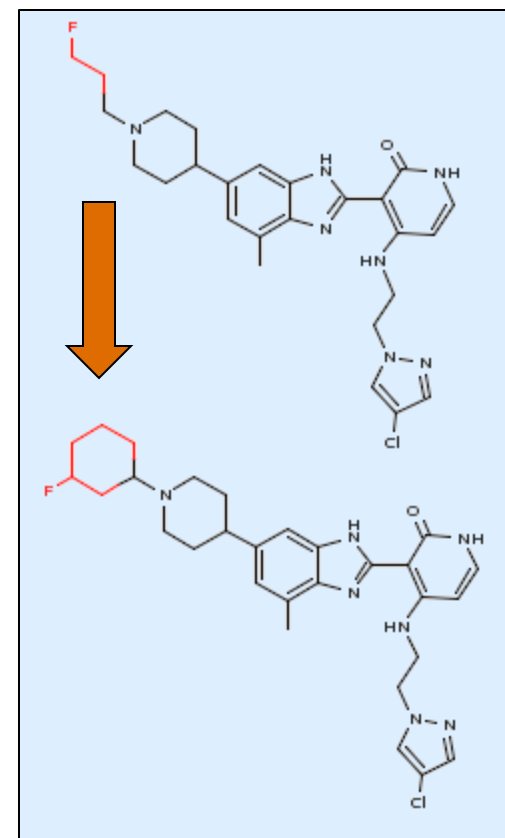


Bioisosteric transformations

- Transformations:
- 1272 1R
- 900 2R
- 218 3R



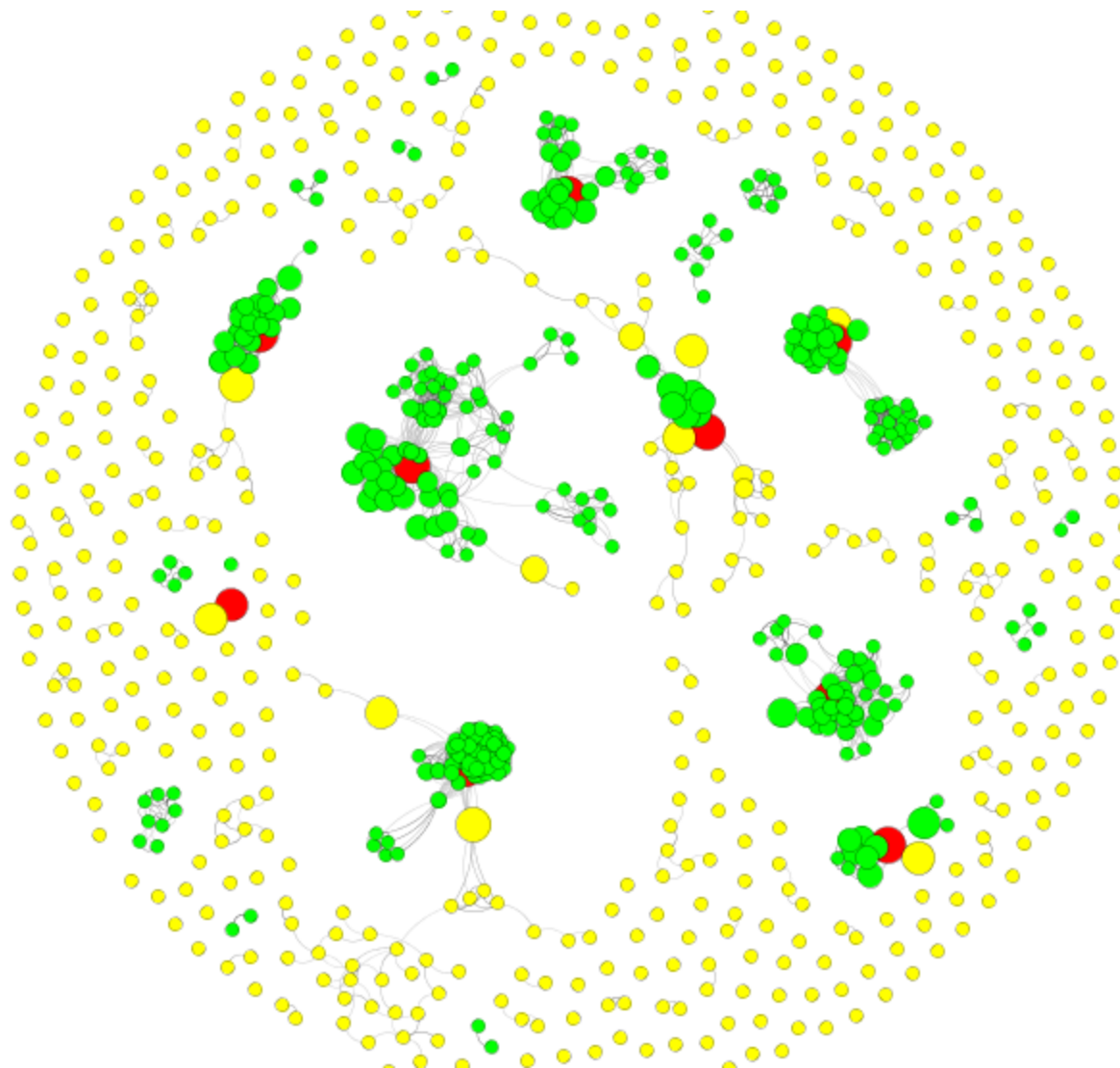
- Wagener & Lommerse, *J.Chem.Inf.Mod.*, 2006, 677-685





Enumerate bioisosteres

Template
Active
Bioisosteres

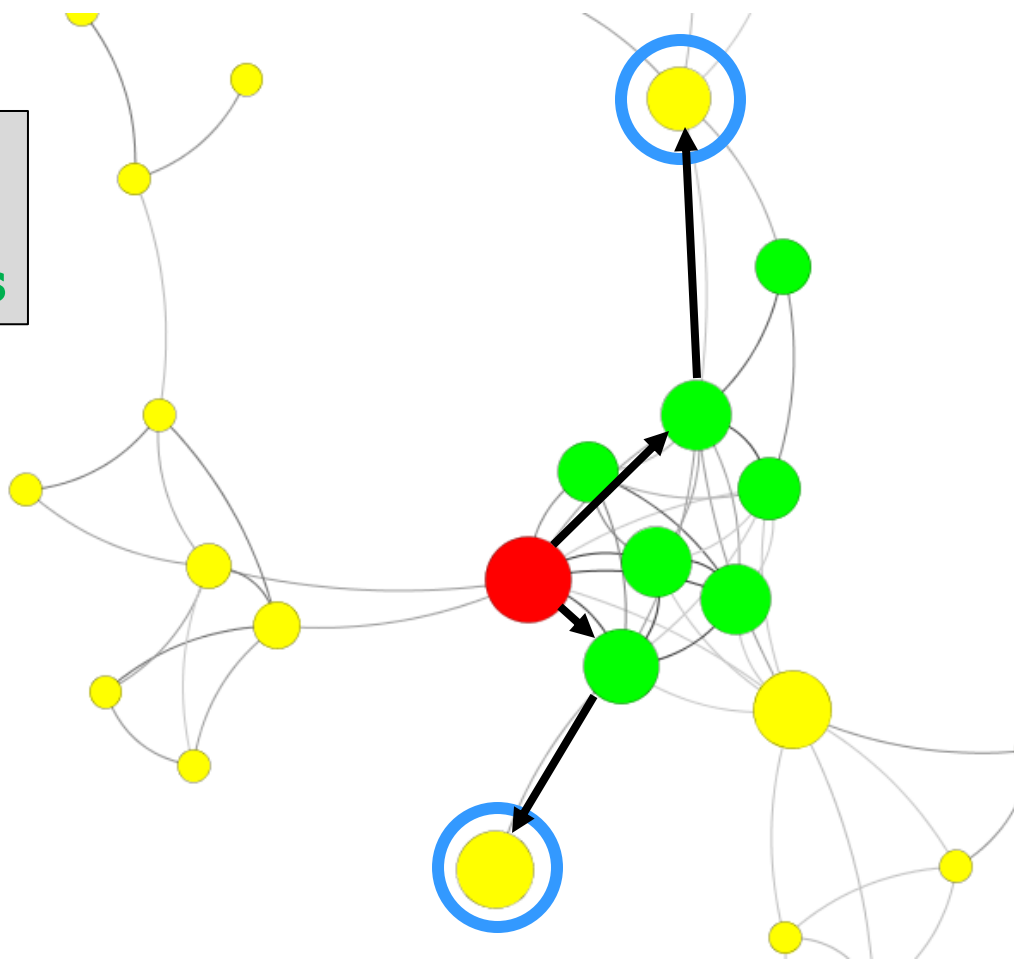




Bioisosteric replacement

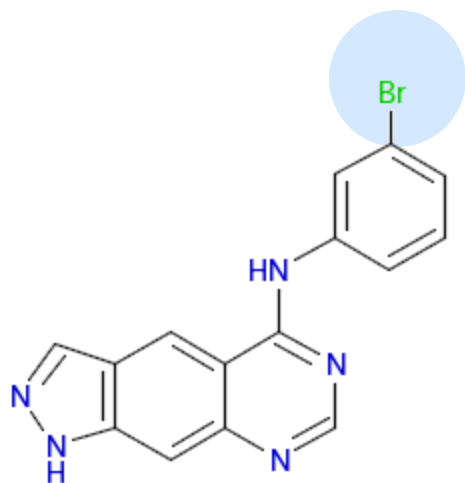
Find similar compounds via virtual molecules

Template
Active
Bioisosteres

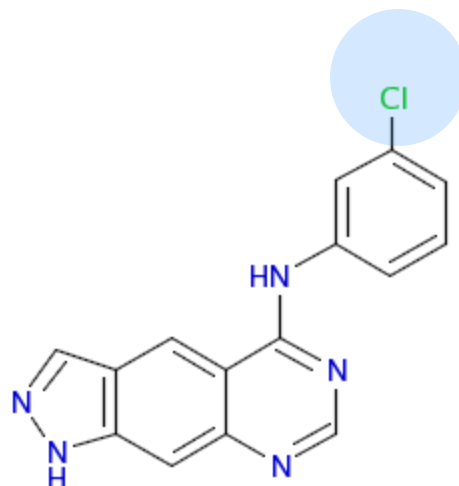


Bioisosteres and similarity

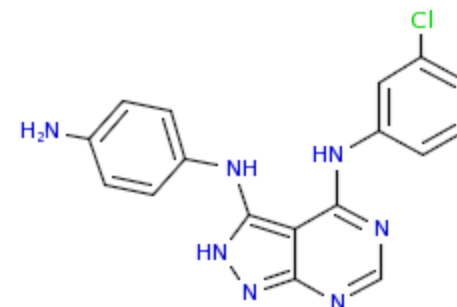
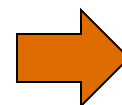
- Overall
- T=0.83 (MACCS)



Template



Virtual

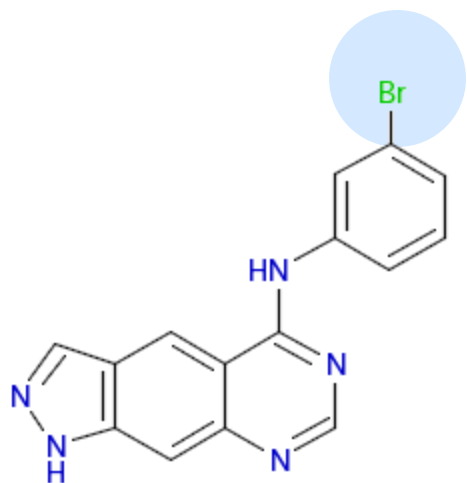


Active

- T=0.88 (MACCS)

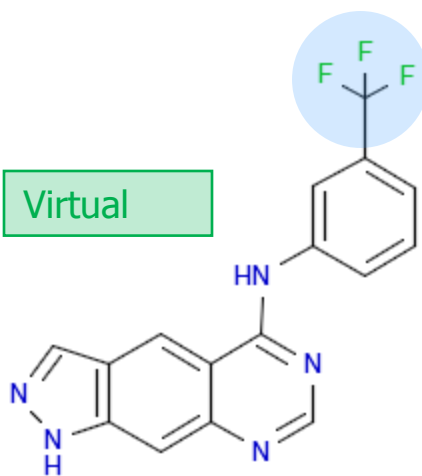
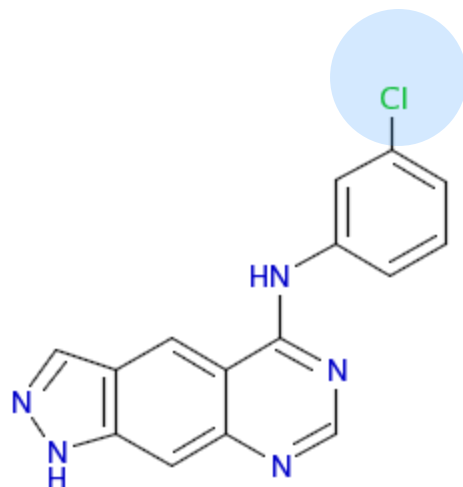


Bioisosteres and similarity



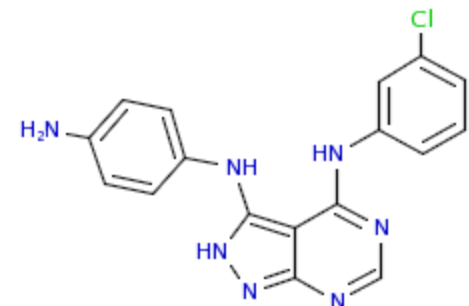
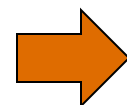
Template

- Overall
- T=0.65 (ECFC_4)



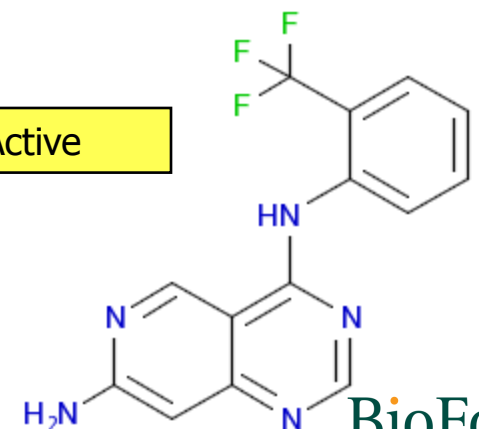
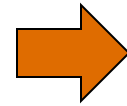
Virtual

25



Active

- T=0.75 (ECFC_4)



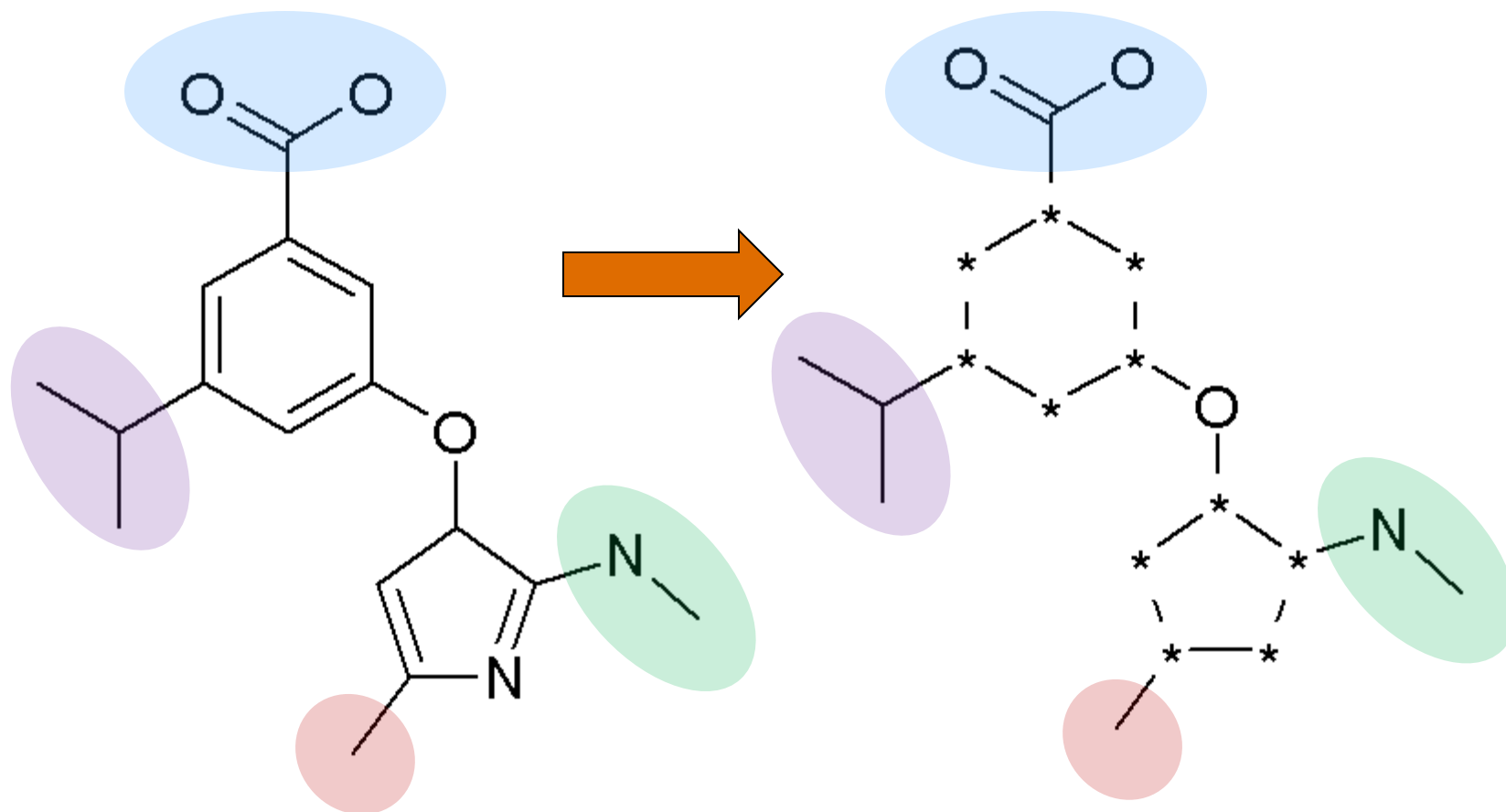


Outline

- Screening paradigms and hit expansion
- Demonstration data set and visualization
- **Methods**
 - fingerprint-based similarity searching and Turbosim
 - bioisosteric transformations and enumerations
 - **ring generalization**
 - ring assembly modifications
- Results

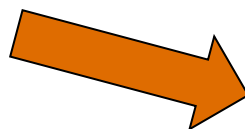
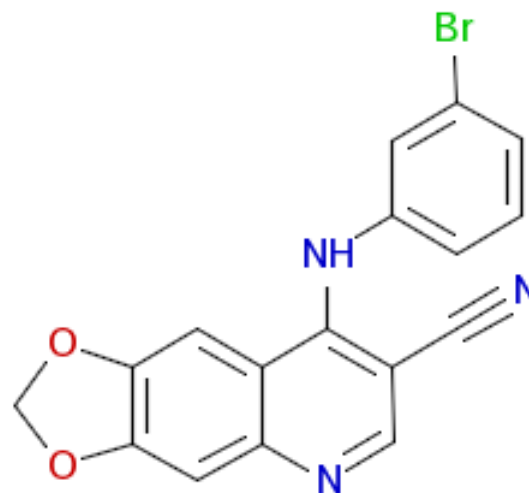
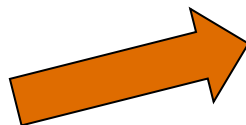
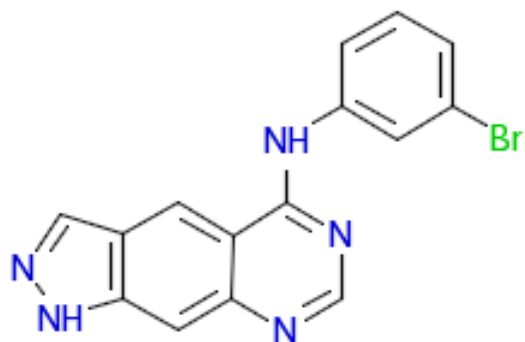
Ring generalization

- SMARTS-based substructure search

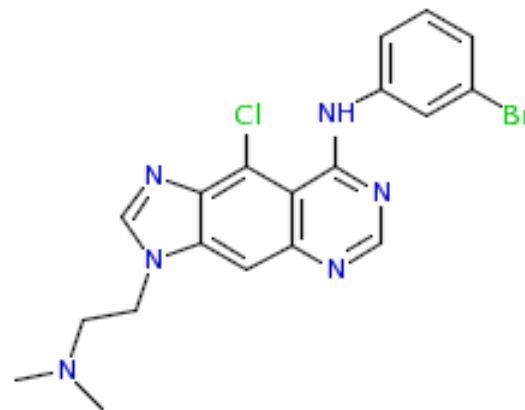


Ring generalization

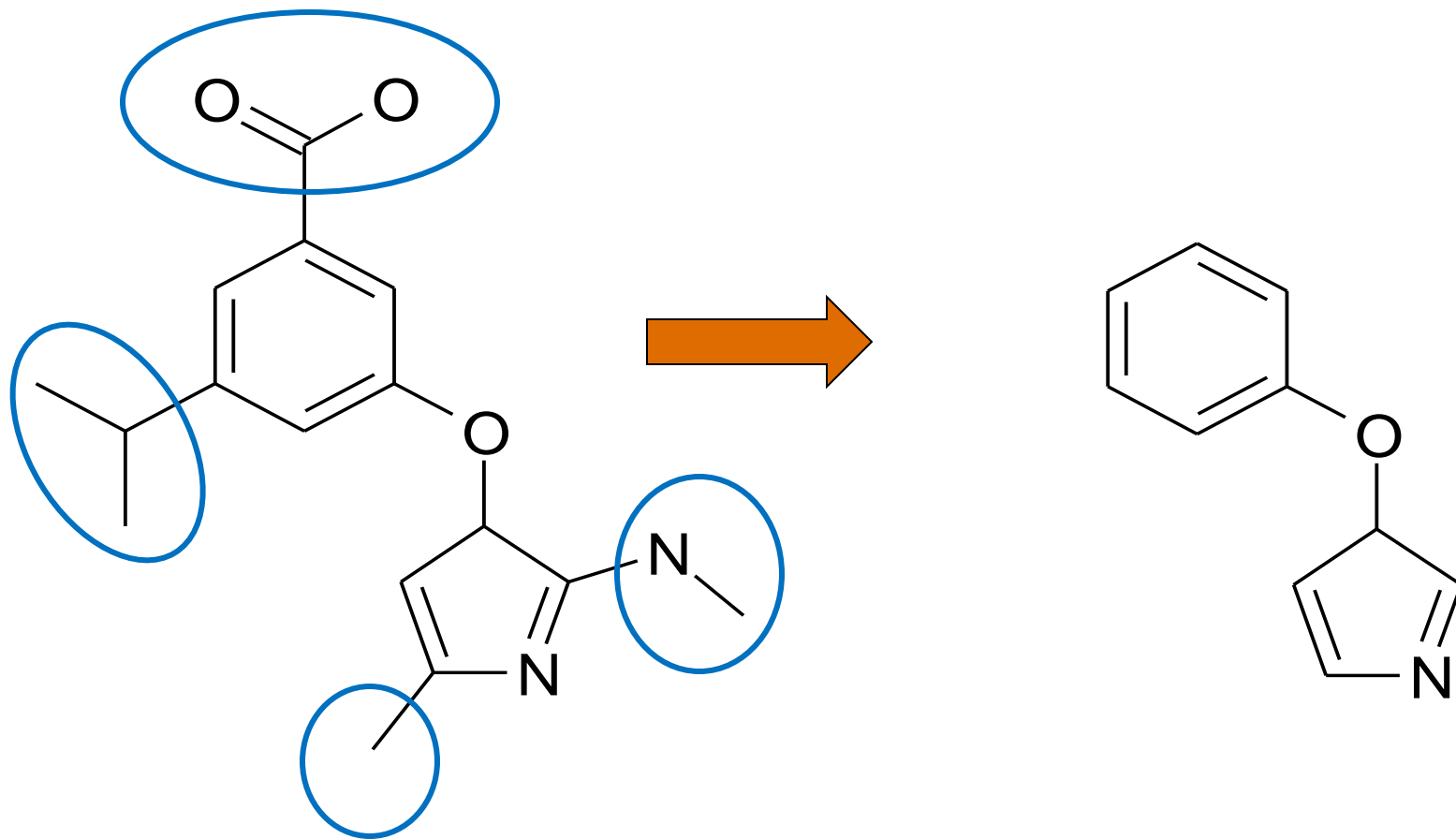
- T=0.56 (MACCS)
- T=0.25 (FCFP_6)



- T=0.55 (MACCS)
- T=0.40 (FCFP_6)

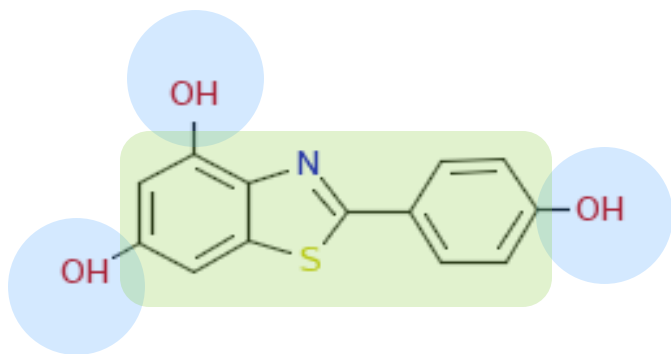


Murcko scaffolds

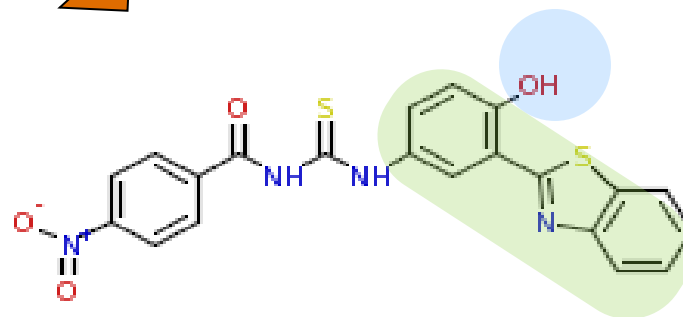
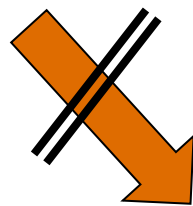


- Bemis & Murcko, J.Med.Chem., 1996, 2887-2893

Murcko scaffolds

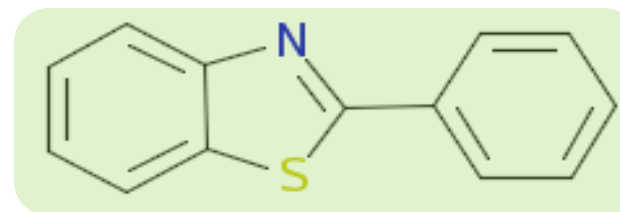
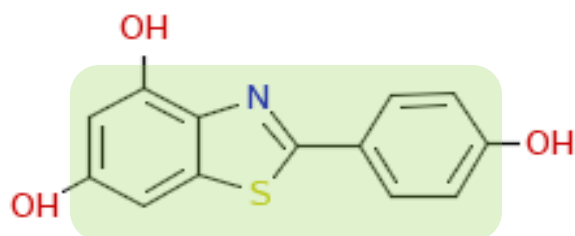


- Substructure search
- Similarity search
- T=0.48 (MACCS)



Murcko scaffolds

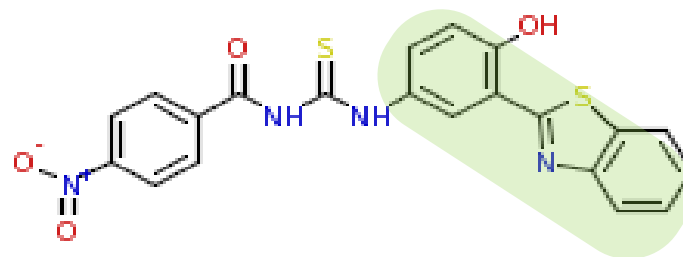
- T=0.62 (MACCS)



- T=0.48 (MACCS)



- Substructure search
- T=0.34 (MACCS)

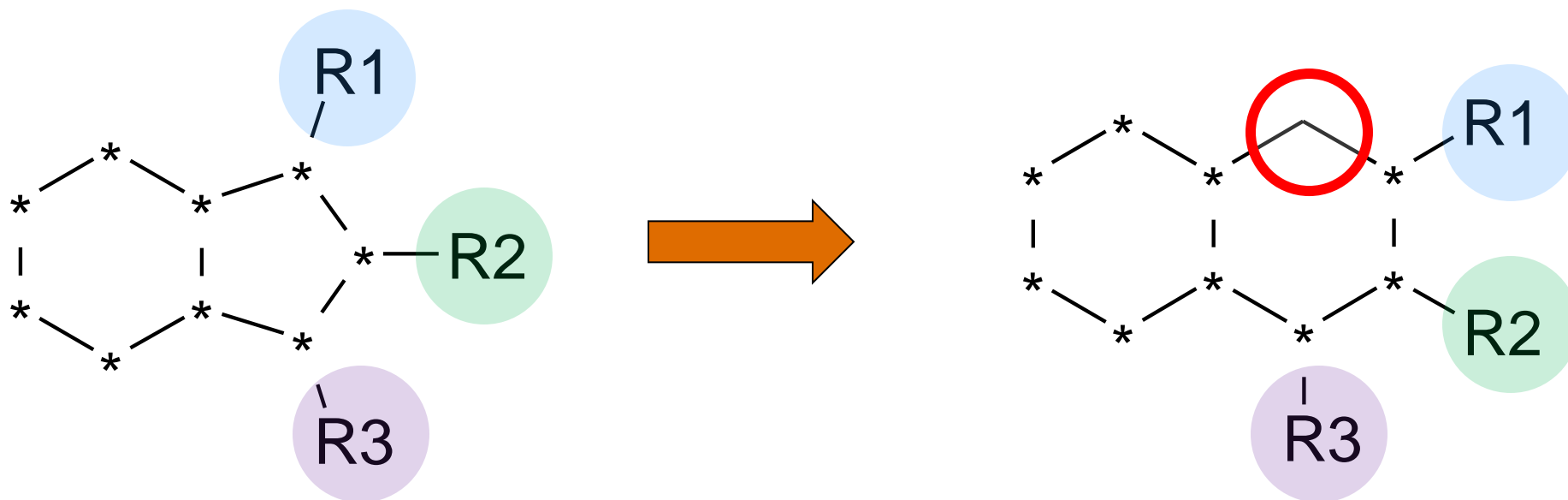




Outline

- Screening paradigms and hit expansion
- Demonstration data set and visualization
- **Methods**
 - fingerprint-based similarity searching and Turbosim
 - bioisosteric transformations and enumerations
 - ring generalization
 - ring assembly modifications
- Results

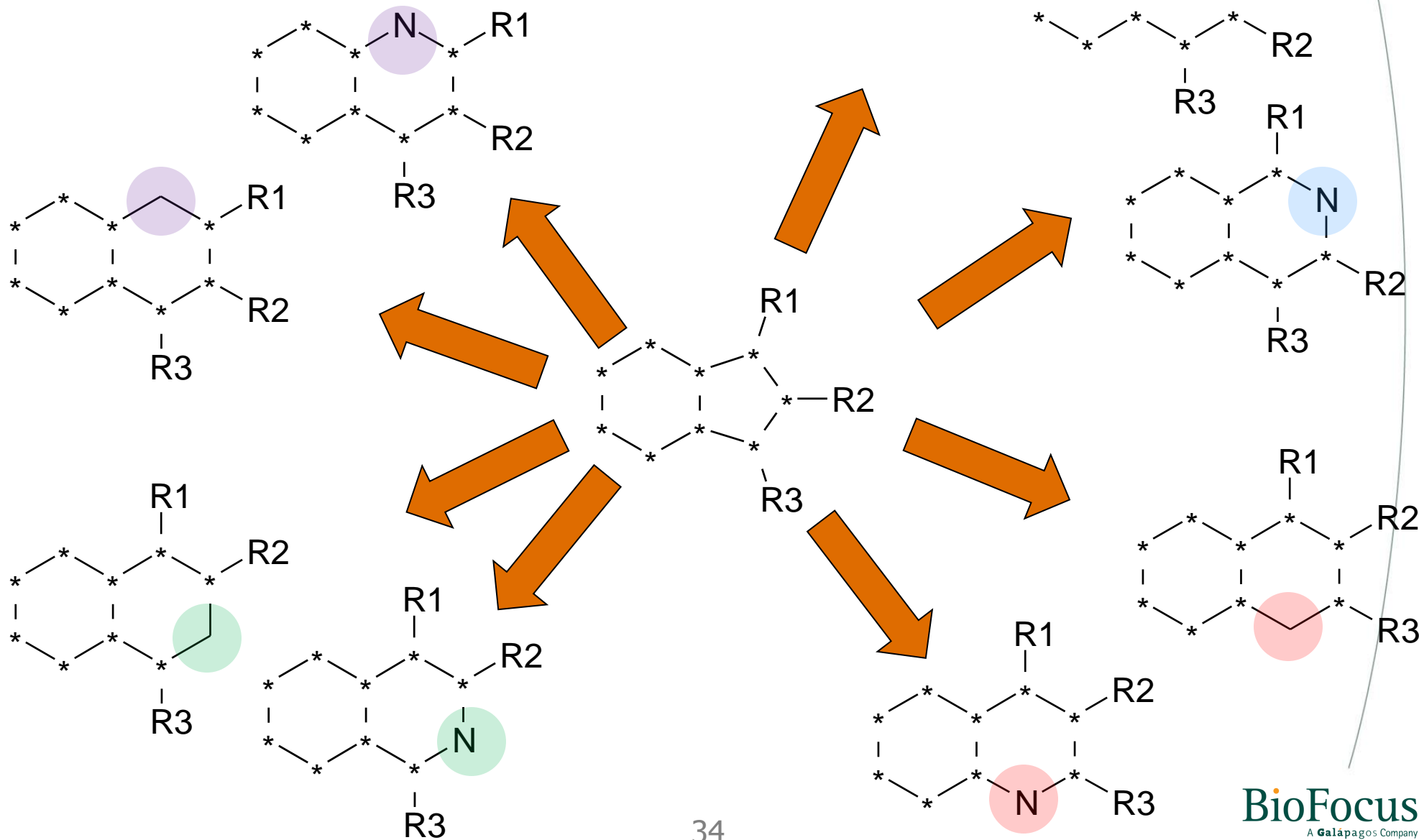
✈ Modifications of annealed rings



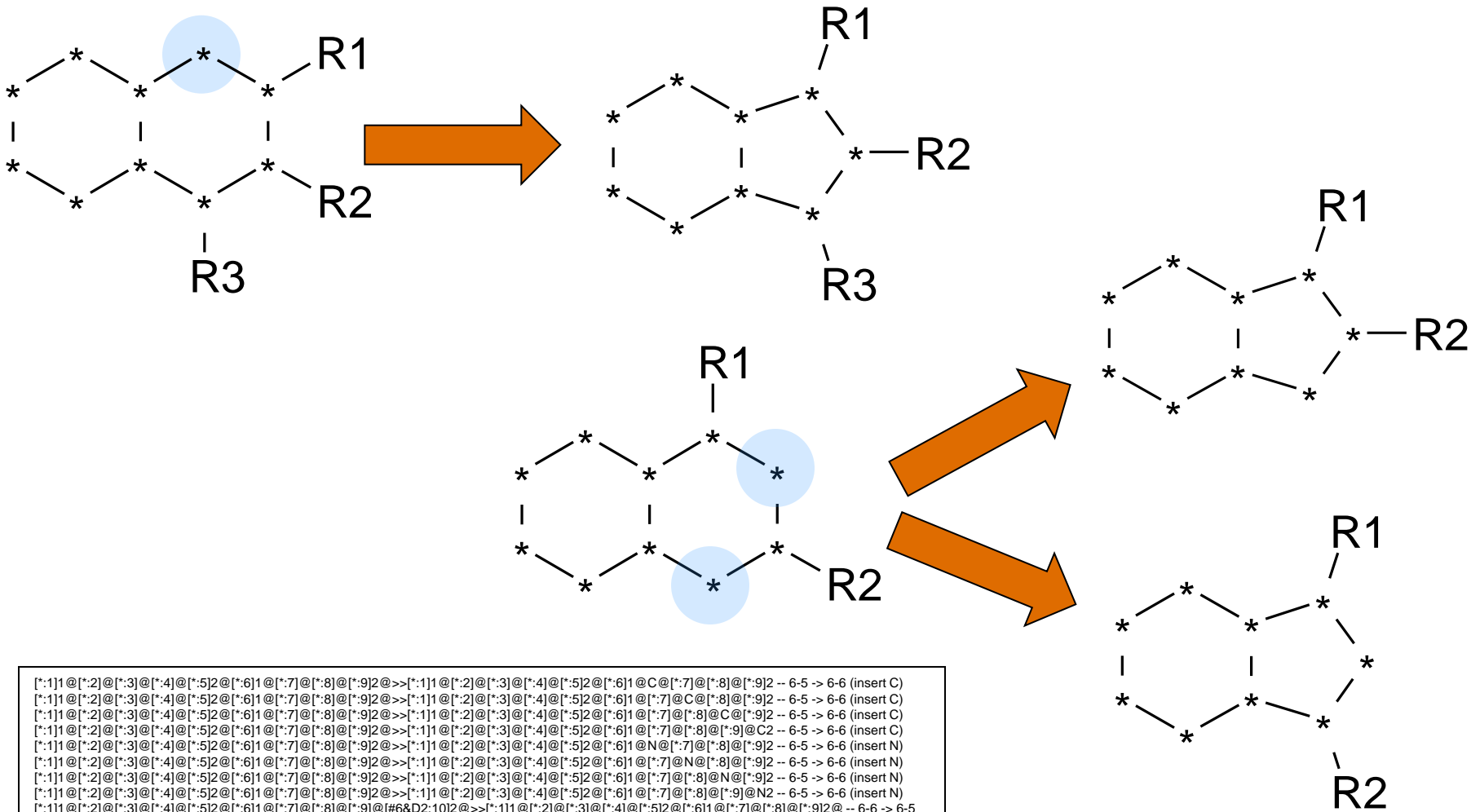
- SMIRKS-based transformations
- Similarity searches



6-5 to 6-6 ring expansion



6-6 to 6-5 ring contraction

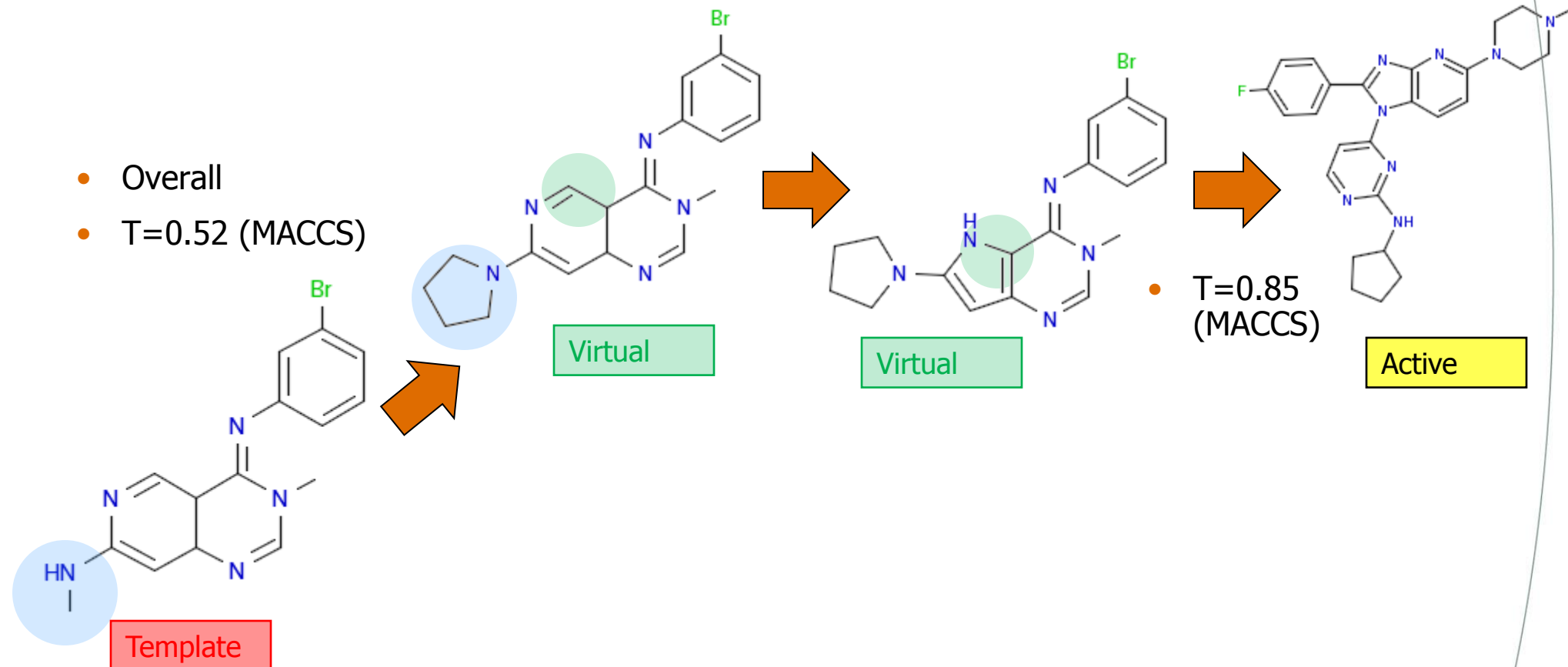


```
[*:1]1@[*:2]@[*:3]@[*:4]@[*:5]2@[*:6]1@[*:7]@[*:8]@[*:9]2@->[*:1]1@[*:2]@[*:3]@[*:4]@[*:5]2@[*:6]1@C@[*:7]@[*:8]@[*:9]2 -- 6-5 -> 6-6 (insert C)
[*:1]1@[*:2]@[*:3]@[*:4]@[*:5]2@[*:6]1@[*:7]@[*:8]@[*:9]2@->[*:1]1@[*:2]@[*:3]@[*:4]@[*:5]2@[*:6]1@[*:7]@C@[*:8]@[*:9]2 -- 6-5 -> 6-6 (insert C)
[*:1]1@[*:2]@[*:3]@[*:4]@[*:5]2@[*:6]1@[*:7]@[*:8]@[*:9]2@->[*:1]1@[*:2]@[*:3]@[*:4]@[*:5]2@[*:6]1@[*:7]@[*:8]@C@[*:9]2 -- 6-5 -> 6-6 (insert C)
[*:1]1@[*:2]@[*:3]@[*:4]@[*:5]2@[*:6]1@[*:7]@[*:8]@[*:9]2@->[*:1]1@[*:2]@[*:3]@[*:4]@[*:5]2@[*:6]1@[*:7]@[*:8]@[*:9]@C2 -- 6-5 -> 6-6 (insert C)
[*:1]1@[*:2]@[*:3]@[*:4]@[*:5]2@[*:6]1@[*:7]@[*:8]@[*:9]2@->[*:1]1@[*:2]@[*:3]@[*:4]@[*:5]2@[*:6]1@N@[*:7]@[*:8]@[*:9]2 -- 6-5 -> 6-6 (insert N)
[*:1]1@[*:2]@[*:3]@[*:4]@[*:5]2@[*:6]1@[*:7]@[*:8]@[*:9]2@->[*:1]1@[*:2]@[*:3]@[*:4]@[*:5]2@[*:6]1@[*:7]@N@[*:8]@[*:9]2 -- 6-5 -> 6-6 (insert N)
[*:1]1@[*:2]@[*:3]@[*:4]@[*:5]2@[*:6]1@[*:7]@[*:8]@[*:9]2@->[*:1]1@[*:2]@[*:3]@[*:4]@[*:5]2@[*:6]1@[*:7]@[*:8]@N@[*:9]2 -- 6-5 -> 6-6 (insert N)
[*:1]1@[*:2]@[*:3]@[*:4]@[*:5]2@[*:6]1@[*:7]@[*:8]@[*:9]2@->[*:1]1@[*:2]@[*:3]@[*:4]@[*:5]2@[*:6]1@[*:7]@[*:8]@[*:9]@N2 -- 6-5 -> 6-6 (insert N)
[*:1]1@[*:2]@[*:3]@[*:4]@[*:5]2@[*:6]1@[*:7]@[*:8]@[*:9]@[#6&D2:10]2@->[*:1]1@[*:2]@[*:3]@[*:4]@[*:5]2@[*:6]1@[*:7]@[*:8]@[*:9]2@ -- 6-6 -> 6-5
[*:1]1@[*:2]@[*:3]@[*:4]@[*:5]2@[*:6]1@[*:7]@[*:8]@[#6&D2:9]@[*:10]2@->[*:1]1@[*:2]@[*:3]@[*:4]@[*:5]2@[*:6]1@[*:7]@[*:8]@[*:10]2@ -- 6-6 -> 6-5
[*:1]1@[*:2]@[*:3]@[*:4]@[*:5]2@[*:6]1@[*:7]@[#6&D2:8]@[*:9]@[*:10]2@->[*:1]1@[*:2]@[*:3]@[*:4]@[*:5]2@[*:6]1@[*:7]@[*:9]@[*:10]2@ -- 6-6 -> 6-5
[*:1]1@[*:2]@[*:3]@[*:4]@[*:5]2@[*:6]1@[#6D2:7]@[*:8]@[*:9]@[*:10]2@->[*:1]1@[*:2]@[*:3]@[*:4]@[*:5]2@[*:6]1@[*:8]@[*:9]@[*:10]2@ -- 6-6 -> 6-5
```

Combined approaches

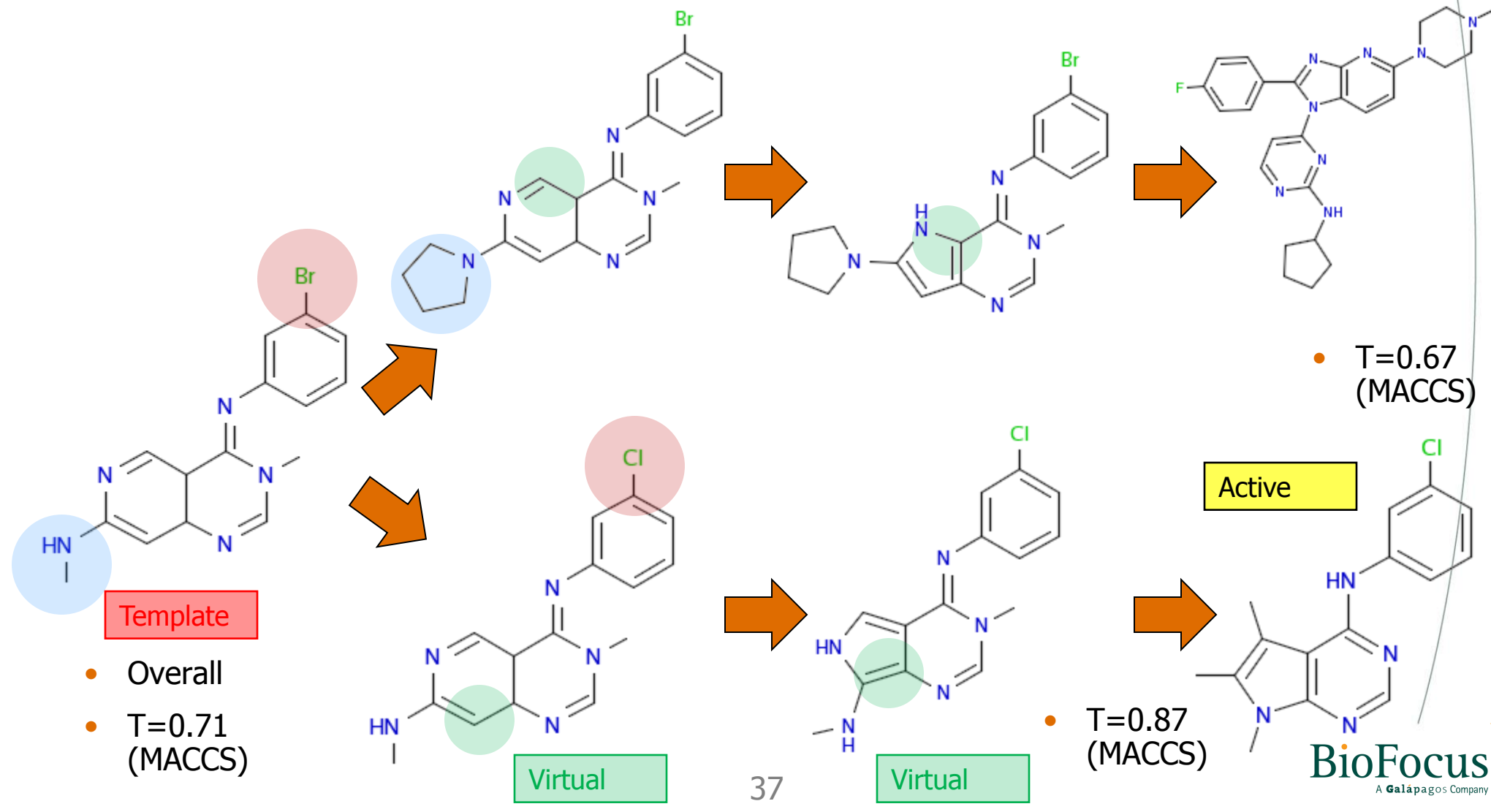
Bioisosteric replacement and ring modifications

- Overall
- T=0.52 (MACCS)



Combined approaches

Bioisosteric replacement and ring modifications





Outline

- Screening paradigms and hit expansion
- Demonstration data set and visualization
- Methods
 - fingerprint-based similarity searching and Turbosim
 - bioisosteric transformations and enumerations
 - ring generalization
 - ring assembly modifications
- Results

Retrieval with fingerprint-based methods

- 9 templates (query structures)
- 256 true actives found (out of 2'863)
- 566 from 900'000 compound library (excluding SFK)
- Sim (MACCS) = 0.85
- Sim (FCFP6) = 0.5
- Sim (ECFC4) = 0.75
- Turbo (MACCS) = 0.9
- Turbo (FCFP6)= 0.6

Approach	EGFR test set		BioFocus collection	
	Count	Exclusive	Count	Exclusive
Sim_MACCS	32	1	111	71
Sim_FCFP6	110	16	27	17
Sim_ECFC4	229	136	425	414
Turbo_MACCS	21		55	16
Turbo_FCFP6	35	4	0	
Combined Total	256	157	566	518

Retrieval with combined methods

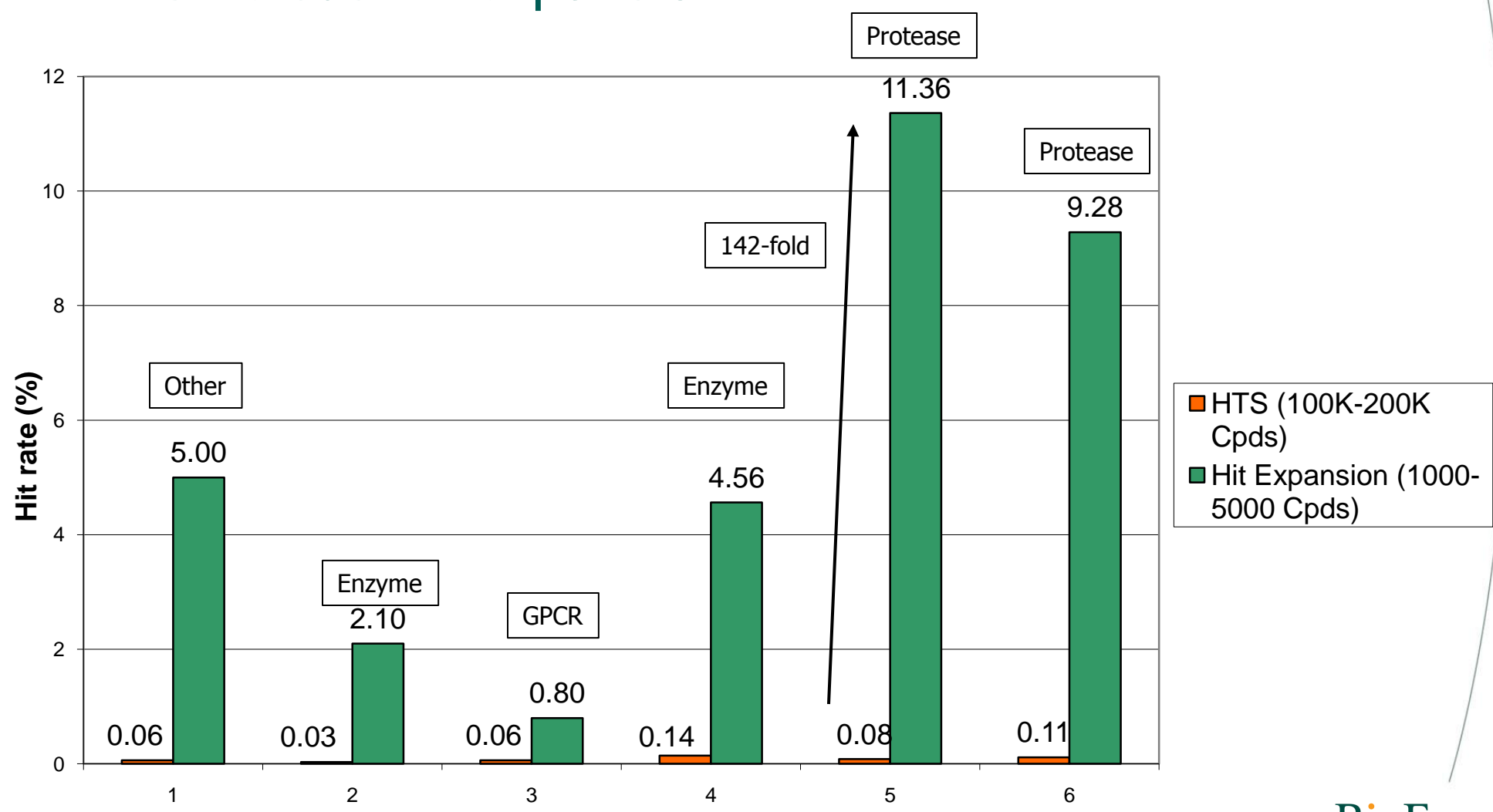
Approach	EGFR test set		BioFocus collection	
	Count	Exclusive	Count	Exclusive
Sim_MACCS	32		111	32
Sim_FCFP6	110		27	1
Sim_ECFC4	229	18	425	32
Turbo_MACCS	21		55	2
Turbo_FCFP6	35		0	
BioisoSim_MACCS	24		328	190
BioisoSim_FCFP6	122		38	2
BioisoSim_ECFC4	169		971	519
BioisoTurbo_MACCS	18		157	43
BioisoTurbo_FCFP6	57		17	
Genring	45	4	14	14
Murcko	78		1688	
Bioiso_Murcko	78		1736	40
Sim_ECFC4_Bioiso_RingMod	114		317	141
Sim_ECFC4_RingMod	191	5	26	
Sim_FCFP6_Bioiso_RingMod	148		9	
Sim_FCFP6_RingMod	137		7	
Sim_MACCS_Bioiso_RingMod	18	7	578	312
Sim_MACCS_RingMod	9		101	23
Turbo_FCFP6_Bioiso_RingMod	161	7	24	
Turbo_FCFP6_RingMod	140		10	
Turbo_MACCS_Bioiso_RingMod	13		310	50
Turbo_MACCS_RingMod	3		75	
Combined Total	433	41	3916	1401

- 177 more true actives were found
- Overrepresentation of certain methods



Hit rates

HTS versus hit expansion





Summary

- Various 2D similarity methods/metrics can be combined with fingerprint-based similarity searches for probing active chemical space
 - bioisosteric transformations
 - ring manipulations
- Increase the number of actives detected in hit expansion set while keeping the “noise” manageable
 - adjust/omit individual pipelines (e.g. Murcko), apply diverse subset selection
 - assess and prioritize consensus of search methods
 - balanced representation of templates
- Powerful tool for hit expansion



Acknowledgements

- Serge P. Parel
- Doris Hafenbradl



Thank you !