DrugbyDesign

Shape Diversity in Fragment Library Design: Theory and Practice

Ken Brameld
(formerly
Roche Palo Alto)

Overview

- Shape characteristics of a typical Rule-of-3 fragment library.
- Improving shape diversity with simple metrics.
  - Properties of standard dimension.
  - A general strategy of “shape-hole” identification.
- Sources of shape diversity
  - Protein-ligand Complexes
  - Cambridge Structural Database (CSD)
- Complexity and shape vs. hit rates.

Fragment Screening – The Basics

- "Fragments" are low MW molecules screened at high concentration.
  - **Rule-of-three*: MW < 300, HBD + HBA <= 3, ClogP <= 3**
- Typical screening methods include crystallography, NMR, or surface plasmon resonance spectroscopies.
- Common justifications for embarking on a fragment screen.
  - Screening less complex molecules covers more chemical space.
  - Starting from a fragment lead generates a more efficient binder with better physicochemical properties.

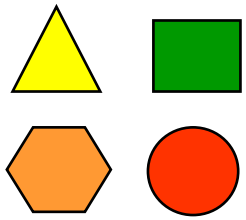
  - *Company X is doing it, we better do it too.*
  - *Desperation… (We've tried everything else, and still don't have a lead for this target.)*

DrugbyDesign

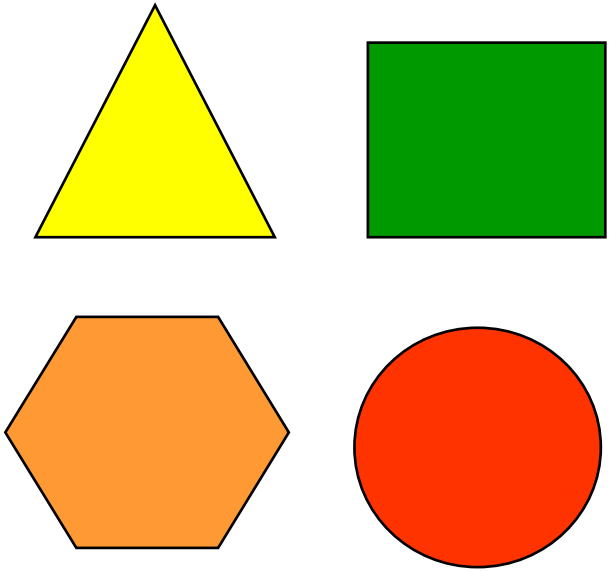
*Congreve, et. al. Drug Discov. Today, 2003, 8, 876–877.

Shape-space: An Abstraction with Useful Implications

- Caveats:
  - Shape *and* electrostatics are important, but electrostatics will be ignored today.
  - Shape-space is an abstract concept that is defined by the objects in the space.
  - These findings are for fragments and may not scale to larger molecules.
- When are shapes similar?



Size matters.



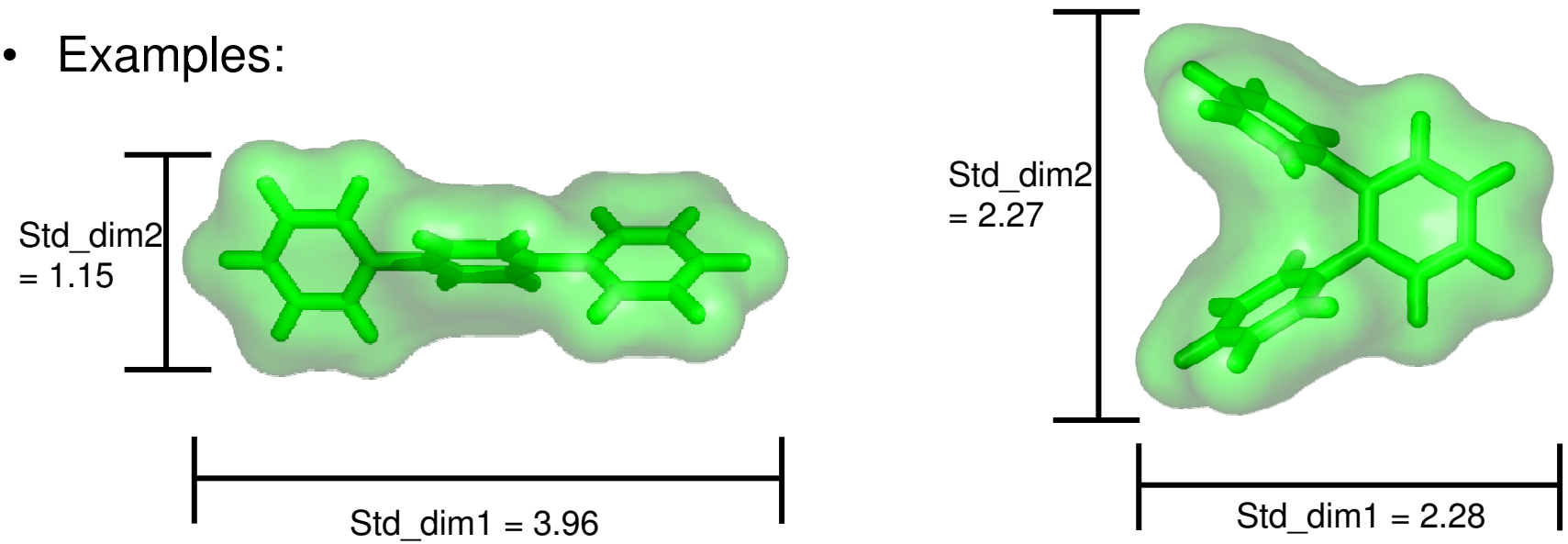
- Similarity is a relative measure – it depends on your perspective.
- Shape space for large molecules has been explored*, how does it scale to fragments?

DrugbyDesign

*Haigh et. al. JCIM 2005, 45, 673-684.

Shape Metric of Standard Dimension

- Std_Dim = “Square root of the largest three eigenvalues of the covariance matrix of atomic coordinates”.
- Reflects the normalized “length,” “height,” and “width” of a molecule.
- Calculated with MOE.
- Examples:

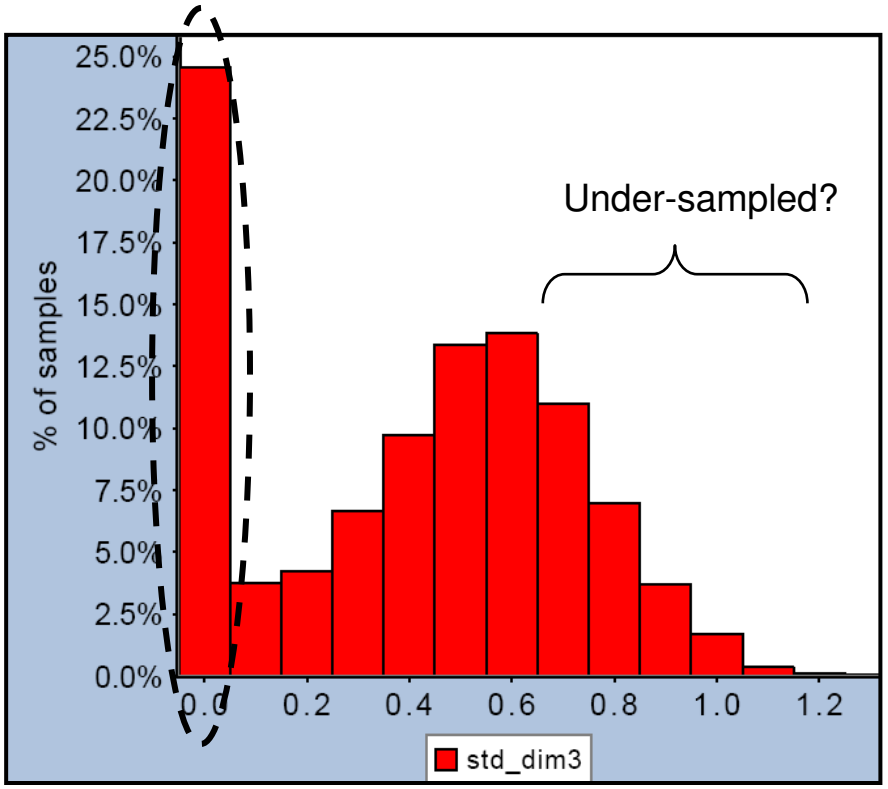


- Standard dimension 3 (std_dim3) is zero for planar molecules.

Characteristics of a typical fragment library

"Aren't all fragments just flat heterocycles?" – Anonymous chemist

- Standard molecular dimensions can be used to crudely characterize molecular shape.
- ZINC "clean" fragments, clustered to create a library of ~11k molecules.
  - ~25% of the molecules are planar.
  - ~75% have std_dim3 < 0.7
- Is this a reasonable distribution?

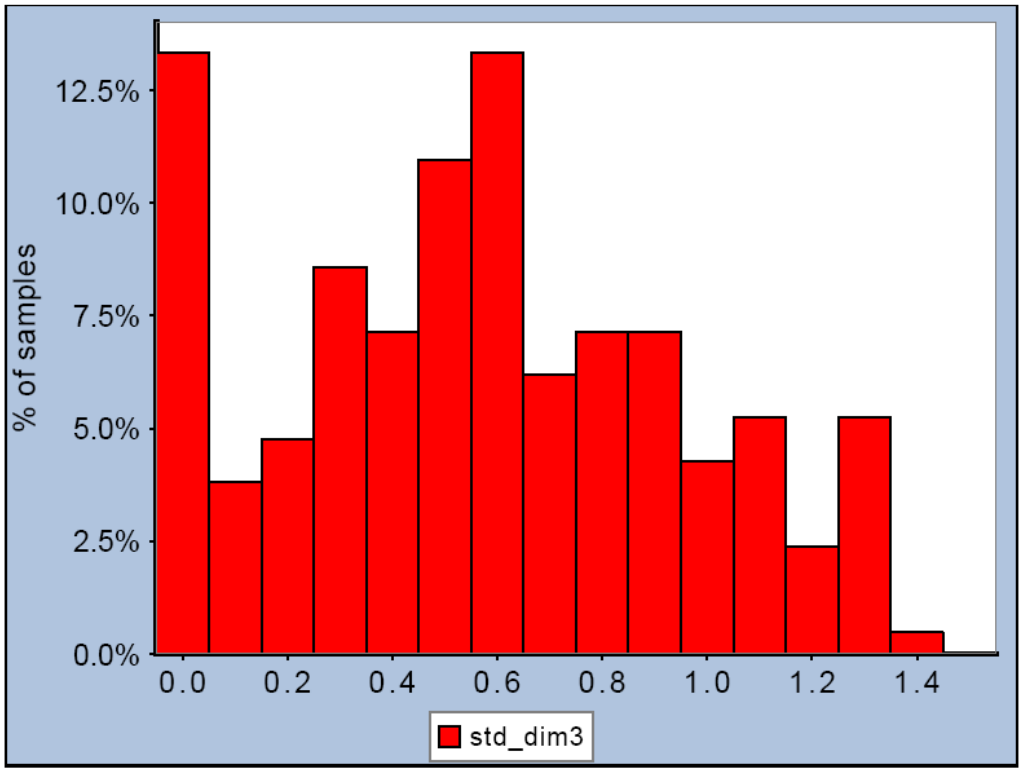


DrugbyDesign

Fragments That Bind Proteins are NOT Flat Heterocycles

- Searched the PDB and Roche internal crystal structures for all unique “rule-of-three” compliant fragments.
- A wide range of std_dim3 values are observed.
  - ~13% are planar molecules
  - ~40% have std_dim3 > 0.7.
- Maybe shape should be considered in fragment library design… but how to measure shape diversity?

Distribution of std_dim3 for ~1400 PDB Fragments



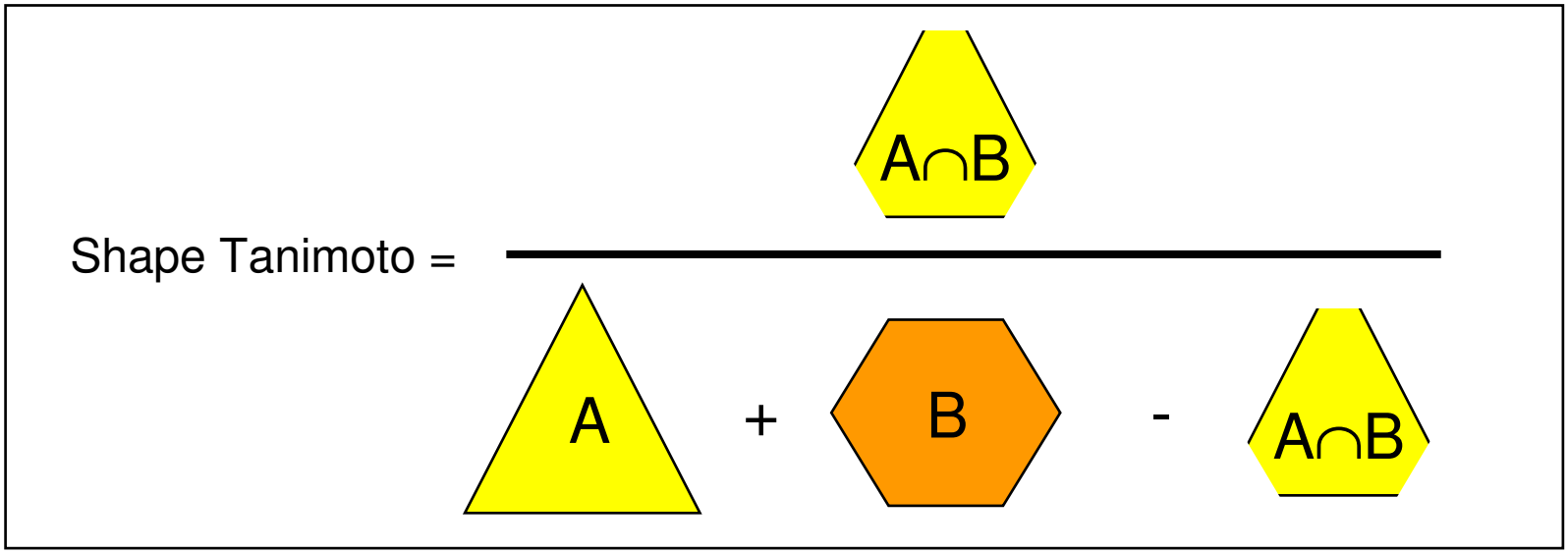
DrugbyDesign

ROCS: A Tool to Evaluate Shape Similarity

(ROCS – Rapid Overlay of Chemical Structures)

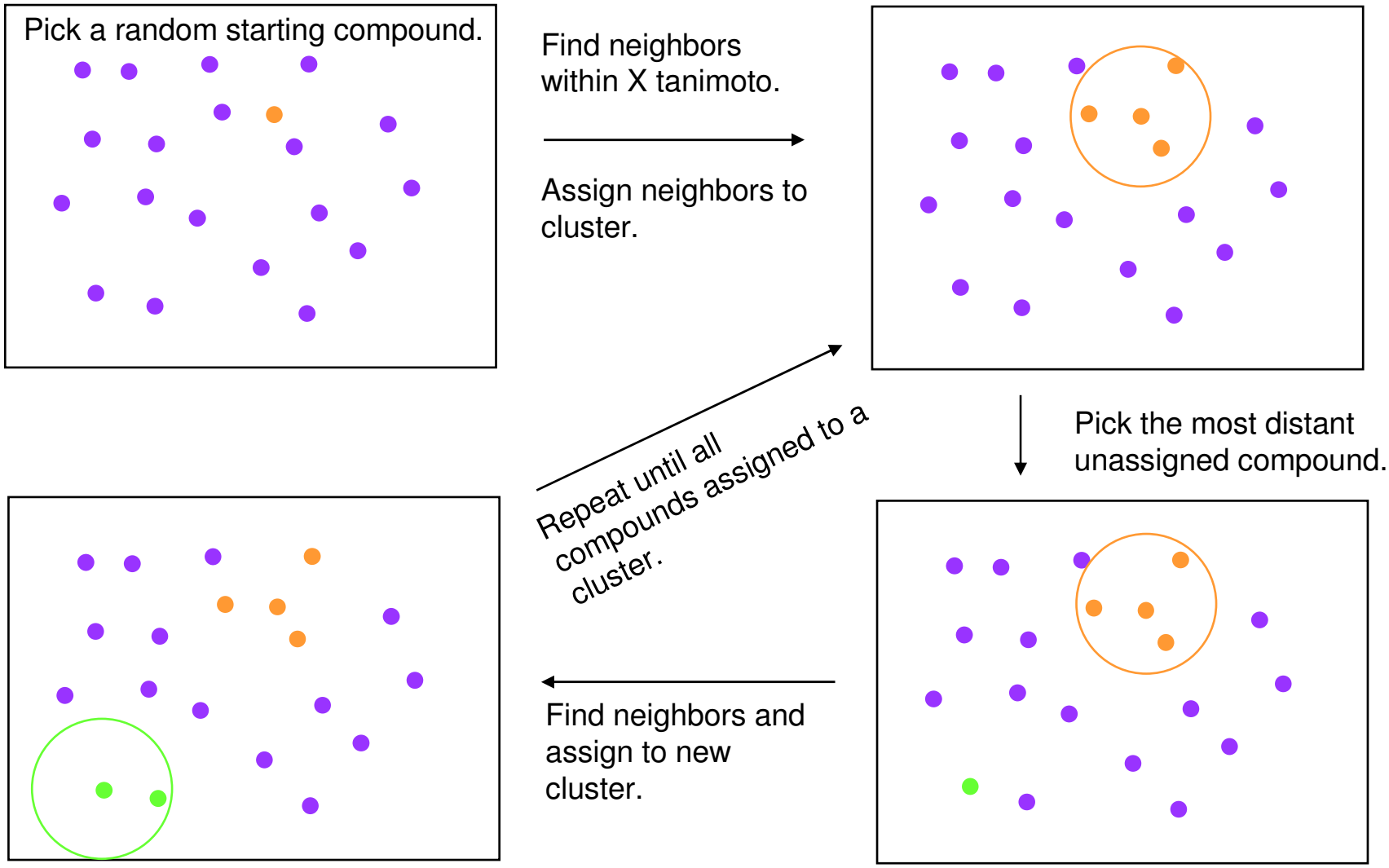
- How similar are these two shapes?

- Similarity is measured by shape Tanimoto (values 0.0 – 1.0; 1= identical shape).
- Shape Tanimoto coefficient illustrated:



DrugbyDesign

Hierarchical Clustering To Measure Shape Diversity

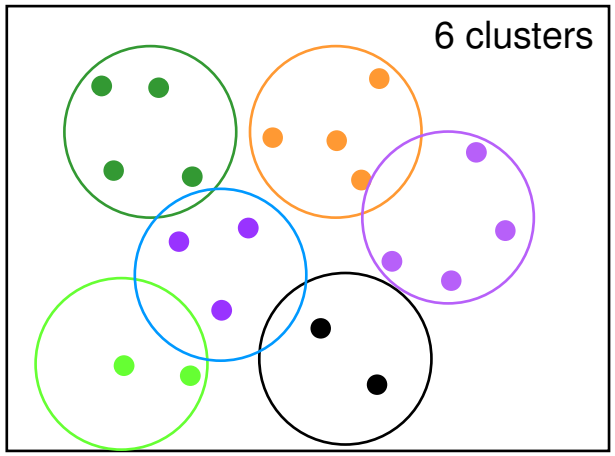


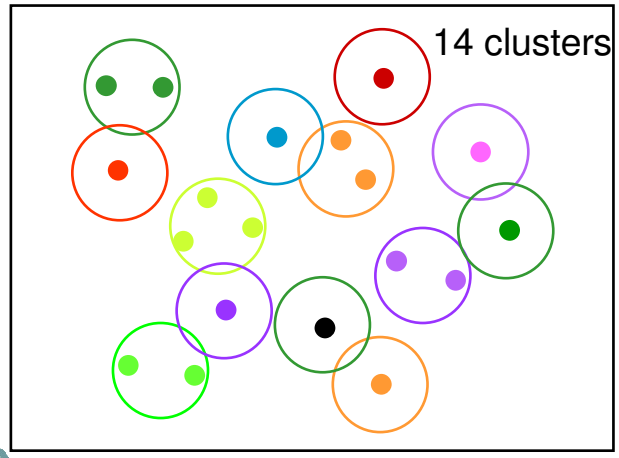
DrugbyDesign

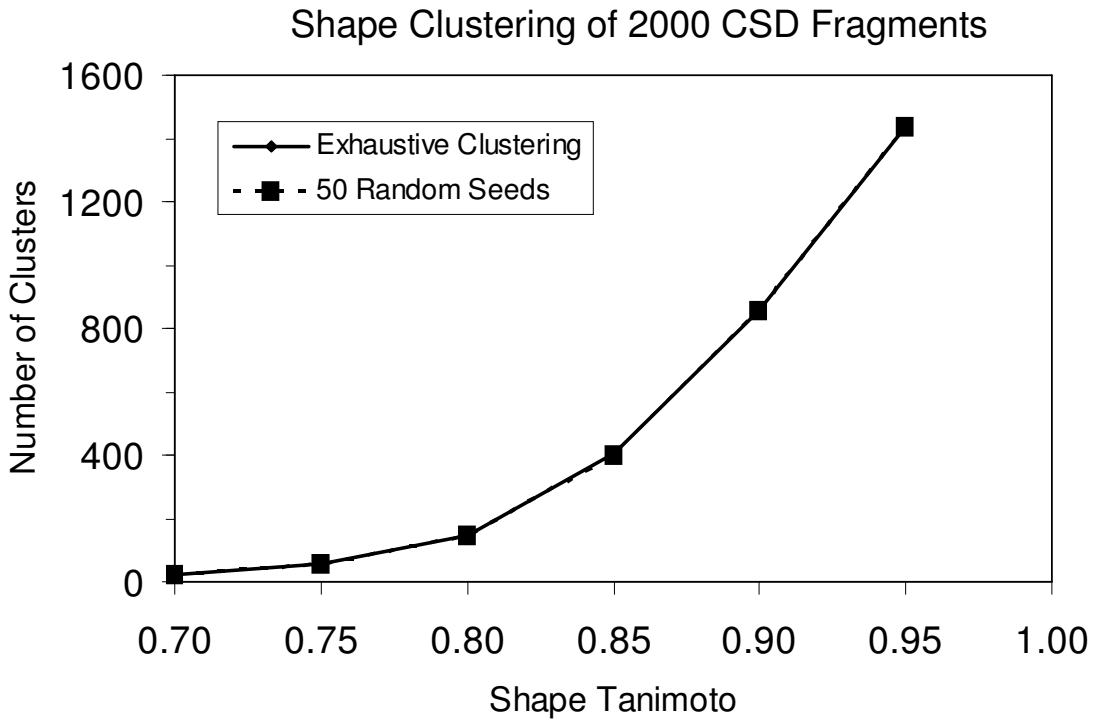
Haigh et. al. JCIM 2005, 45, 673-684

Diversity Depends on the Definition of Similarity

- Number of clusters used to measure diversity: **More Clusters = More Diversity**
- Number of clusters depends on the similarity radius: **Smaller radius = More Clusters**





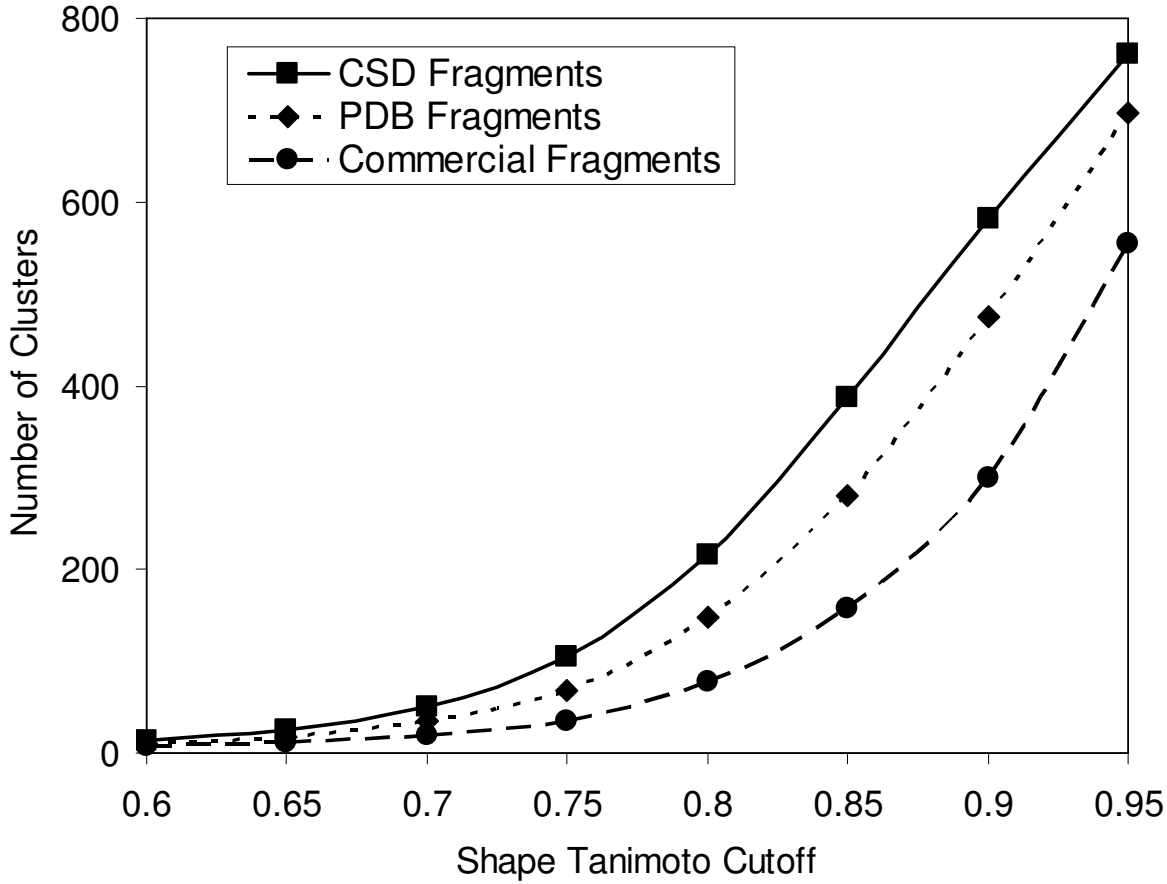


ROCS Shape Tanimoto used as the similarity metric.

DrugbyDesign

Commercial Fragment Library has Poor Shape Diversity

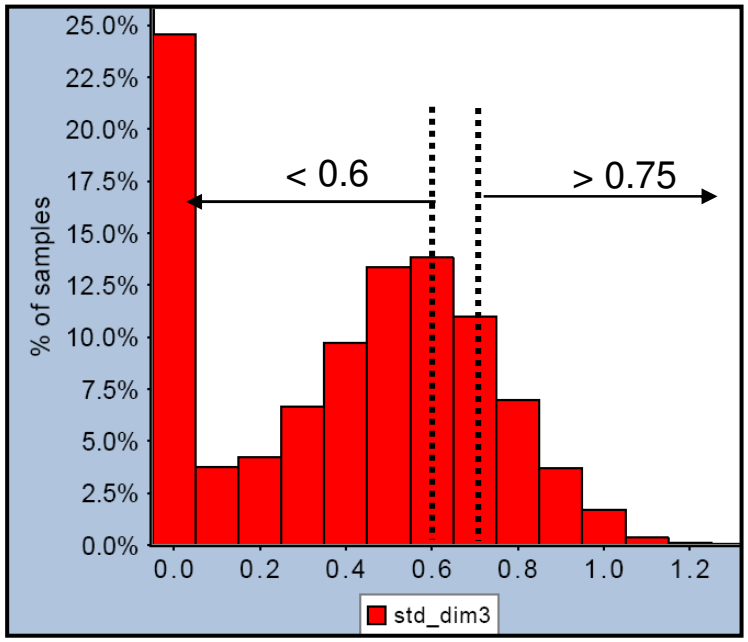
- Shape diversity calculated for three libraries containing 1000 molecules each:
  - CSD fragments
  - PDB fragments
  - Commercial
- CSD and PDB fragments are more shape diverse than commercial library.

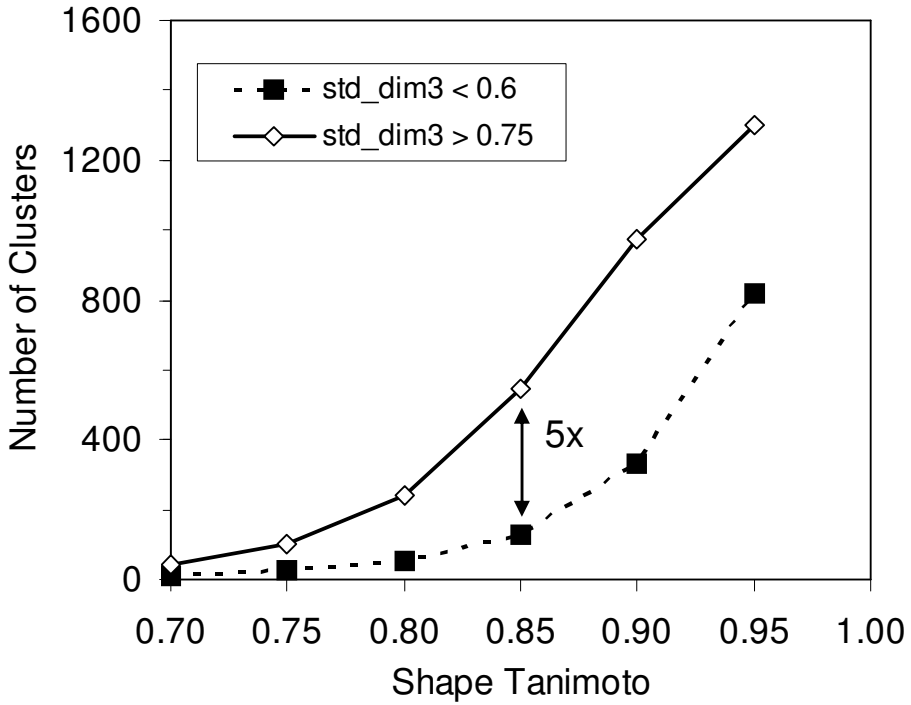


DrugbyDesign

Std_dim3 as a Simple Metric to Increase Shape Diversity

- Shape diversity of fragments is strongly dependent on std_dim3.
  - ZINC was partitioned based on std_dim3 into two sub-libraries of 1500 cmpds each.
  - Shape diversity determined by hierarchical clustering dMax method.

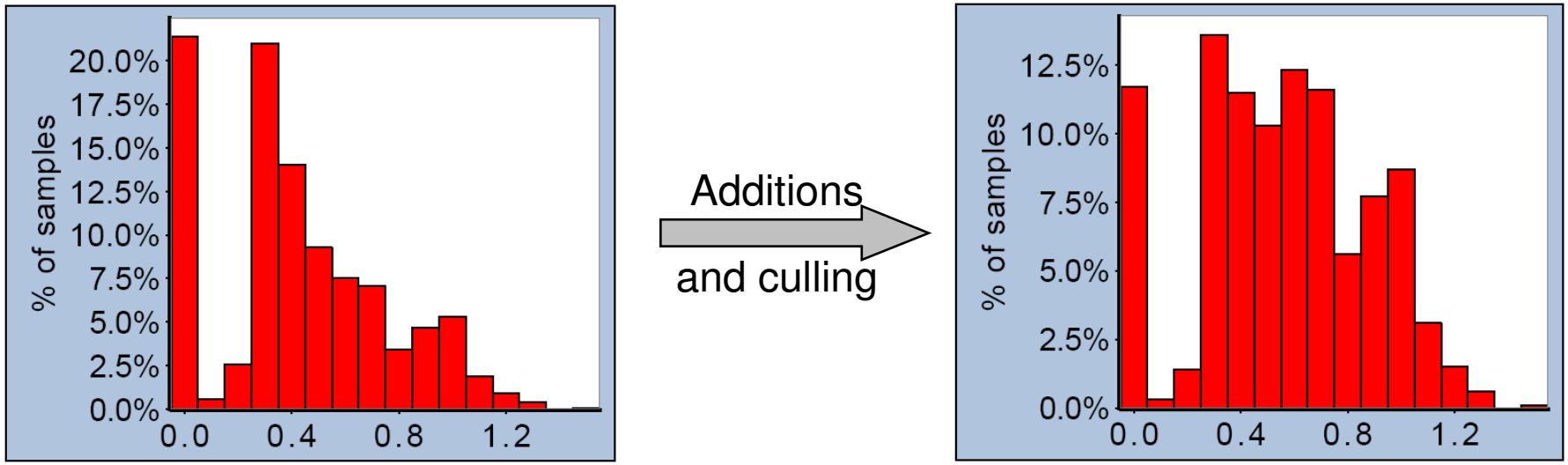




DrugbyDesign

Application of Std_dim3 to Library Design

- For existing libraries, std_dim3 is calculated from:
  - Average for conformational ensemble (Omega)
  - Single conformation (Corina)
  - Single conformation slightly underestimates std_dim3, but results agree +/- 0.2 units for 85% of molecules
- Library profile may then be analyzed and adjusted:
  - New purchases can favor under-represented regions.
  - Over-represented regions may be culled.

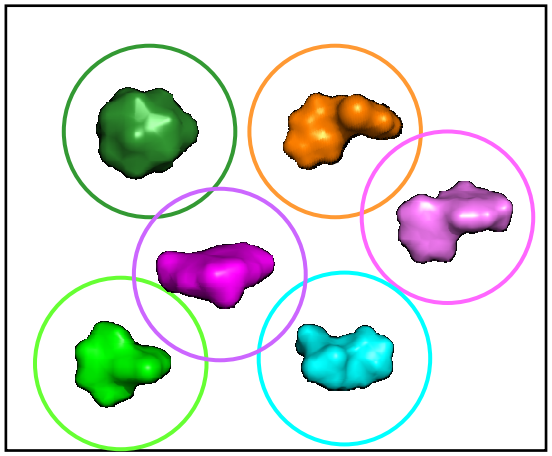


DrugbyDesign

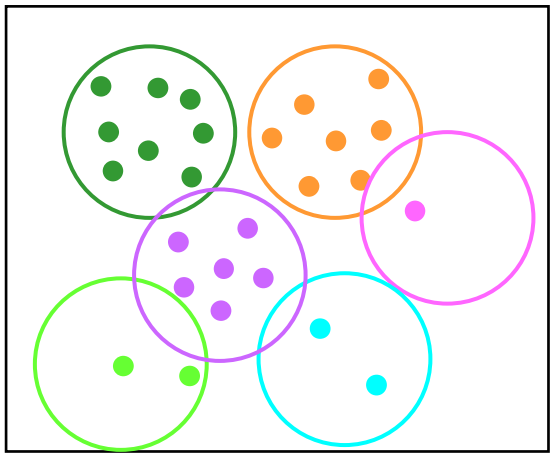
Fragment Library Hole Filling Using Shape

- Need sources of relevant or diverse shapes to define shape space.
  - What is a ‘relevant’ shape?
    - Observed to bind a protein with acceptable ligand efficiency.
    - A portion or subset of a molecule that is part of a protein-ligand complex structure in the PDB estimated to have good ligand efficiency.
  - Small molecule X-ray structures in the CSD provide a source of diverse shapes.
- Example workflow:

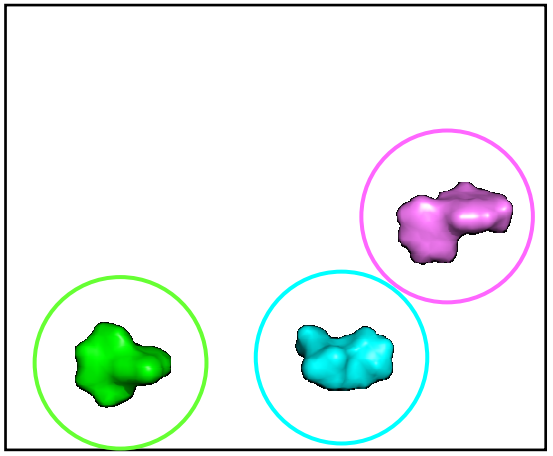
Shape-space Definition



Hits in Current Libary



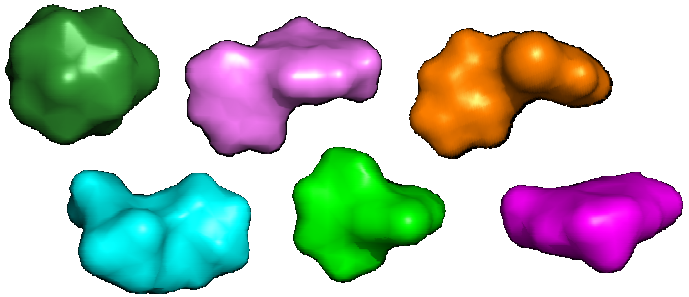
Shape-holes to Be Filled



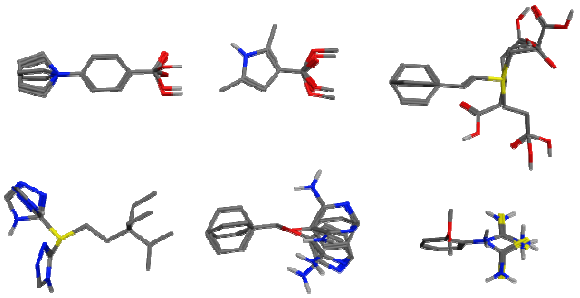
DrugbyDesign

Shape Hole Filling Workflow is Computationally Intensive

1. Evaluate current library for missing shapes.
    a. For each query shape, compare against all molecules/conformers of current library



X



= 10^8 shape comparisons

10^3 shape queries

10^4 molecules x 10 conformations

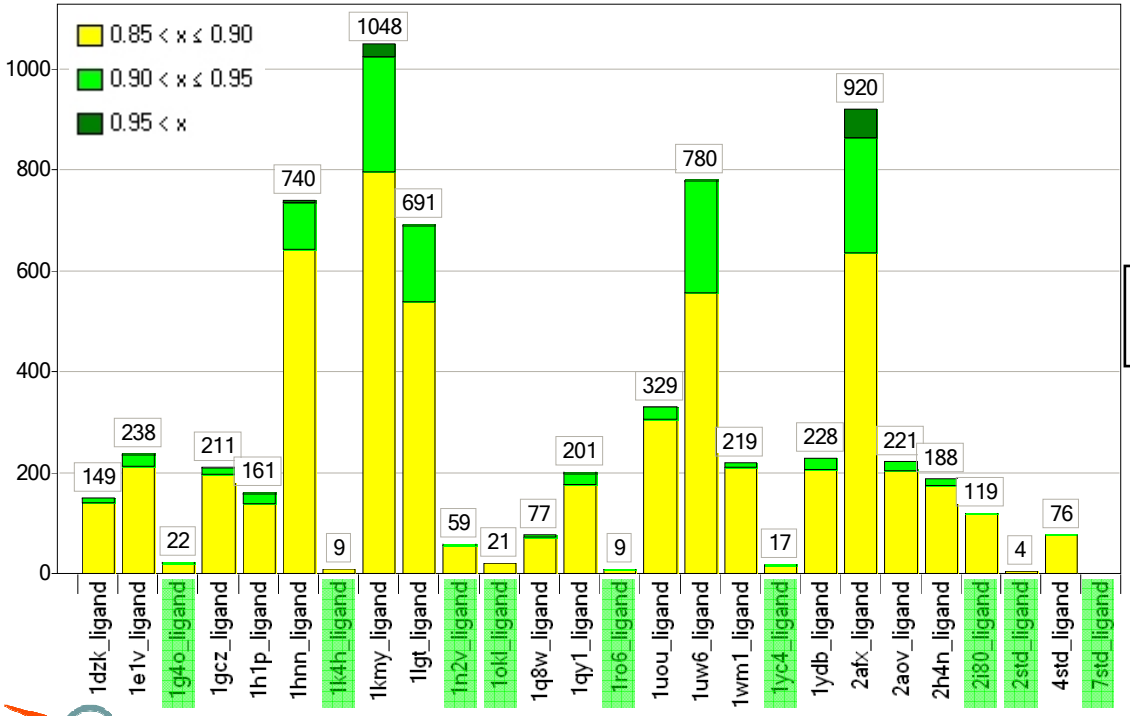
    b. Keep shapes with few matches in the current library.

2. Search commercial or internal compound sources for matches for these new shapes.
    a. Calculation of similar size for searching vendor libraries.
    b. Fewer shape queries but more compounds to search against:

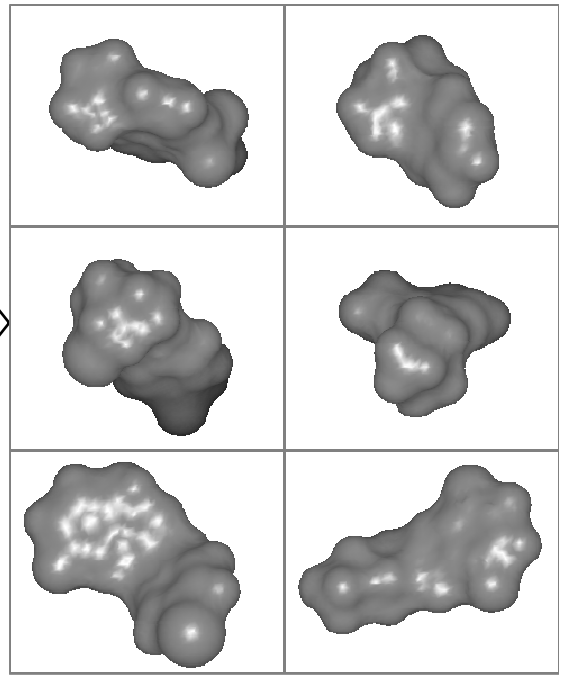
DrugbyDesign

Shapes That Bind Proteins – Mine PDBbind Database

- Database of 1300 protein/ligand complexes deposited in the PDB.
  - Annotated with affinity data.
- Identified 78 possible fragments for follow up based upon filters:
  - Affinity/Efficiency: pKd > 4 & LE > 0.3
  - Properties: 12 < heavy atoms < 20 & std_dim3 >= 0.6



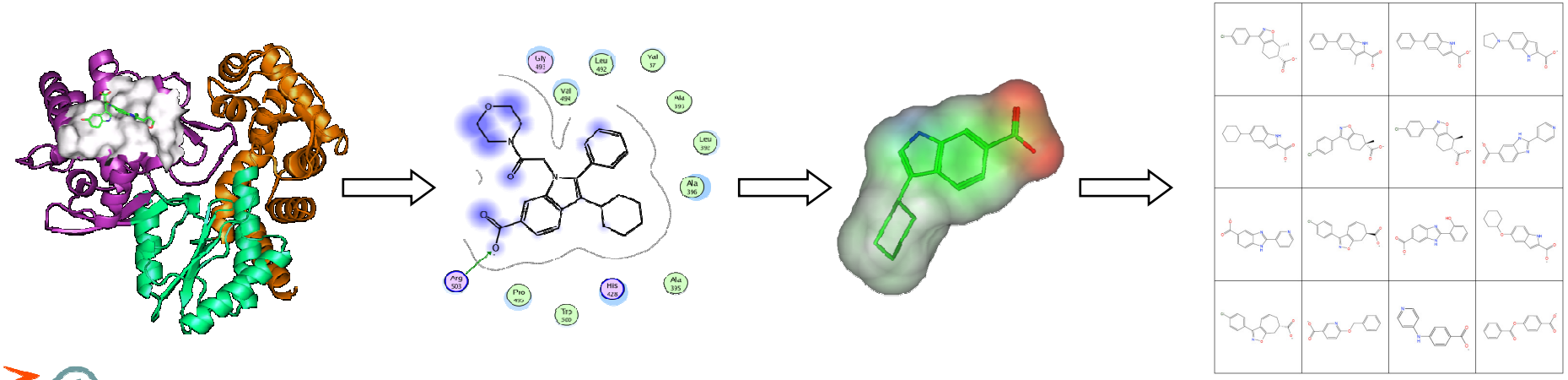
Selected Target Shapes



DrugbyDesign

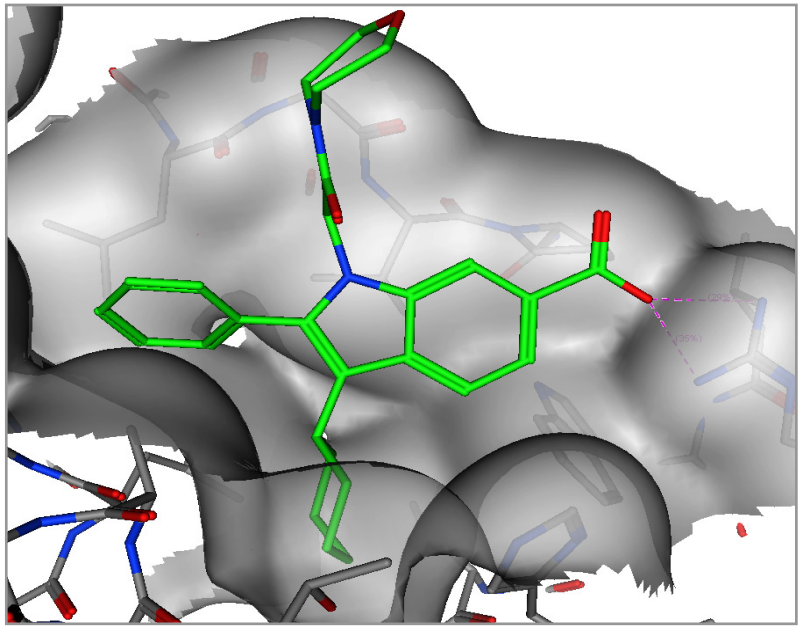
“Molecular Anchor” as Desirable Fragment Shapes

- Use structures of protein-ligand complexes to guide fragment library design.
- Identify “molecular anchors” within these complexes.
  - Subsets of atoms making critical binding site interactions.
- Attempt to identify fragments that match the shape of molecular anchors.
  - Assess coverage of molecular anchors in the current library.
  - Select compounds from commercial sources to fill any holes.
  - Remaining holes in coverage become synthetic targets.

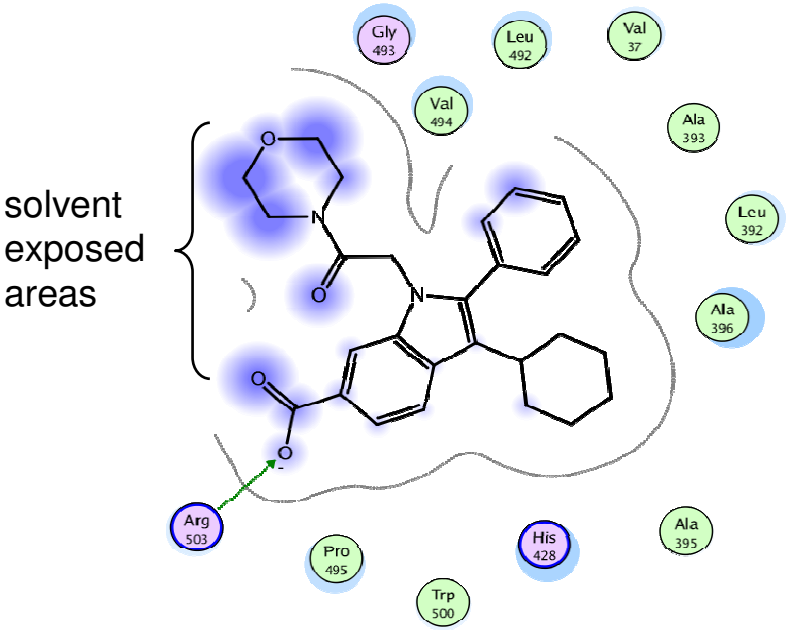


DrugbyDesign

Example of a Molecular Anchor

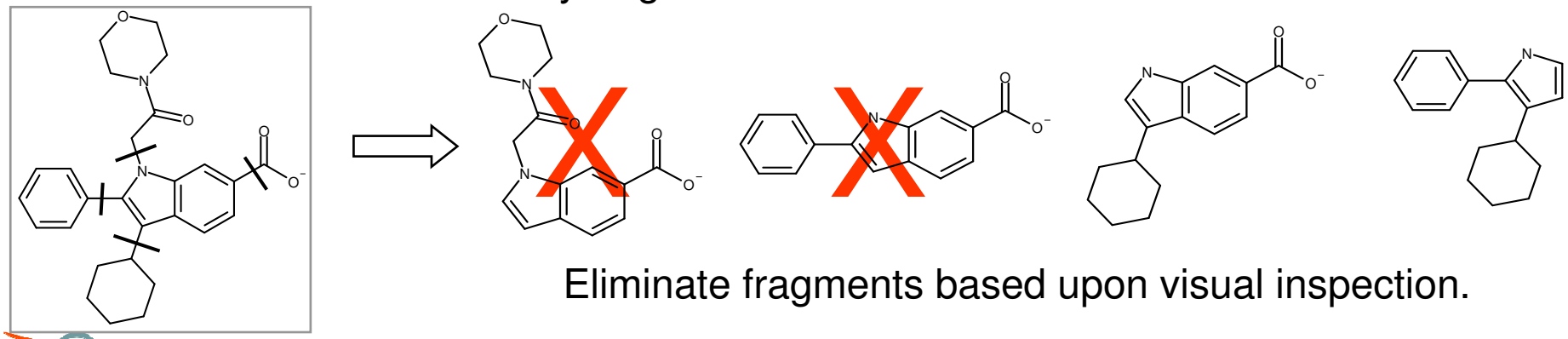


Automated Interaction Map



solvent exposed areas

Possible “molecular anchors” by fragmentation:

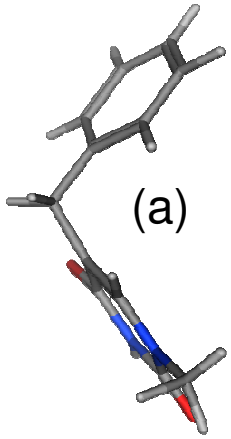


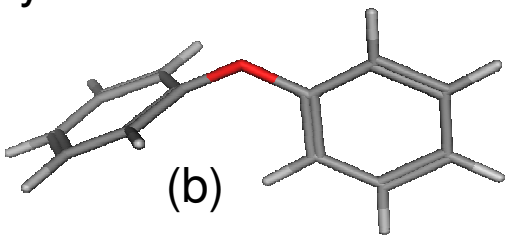
Eliminate fragments based upon visual inspection.

DrugbyDesign

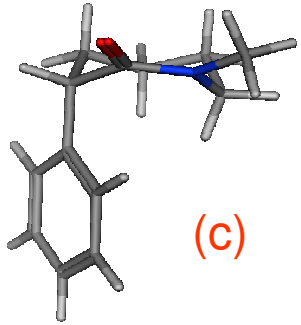
Under-represented Shapes Can be Identified

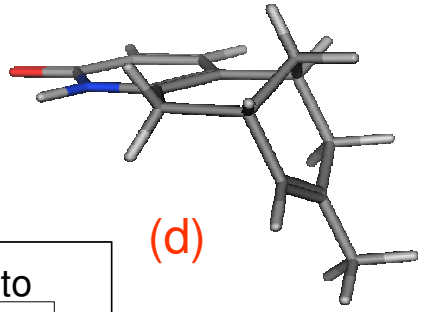
Common motifs such as diphenyl-ether are well represented in a commercial library:

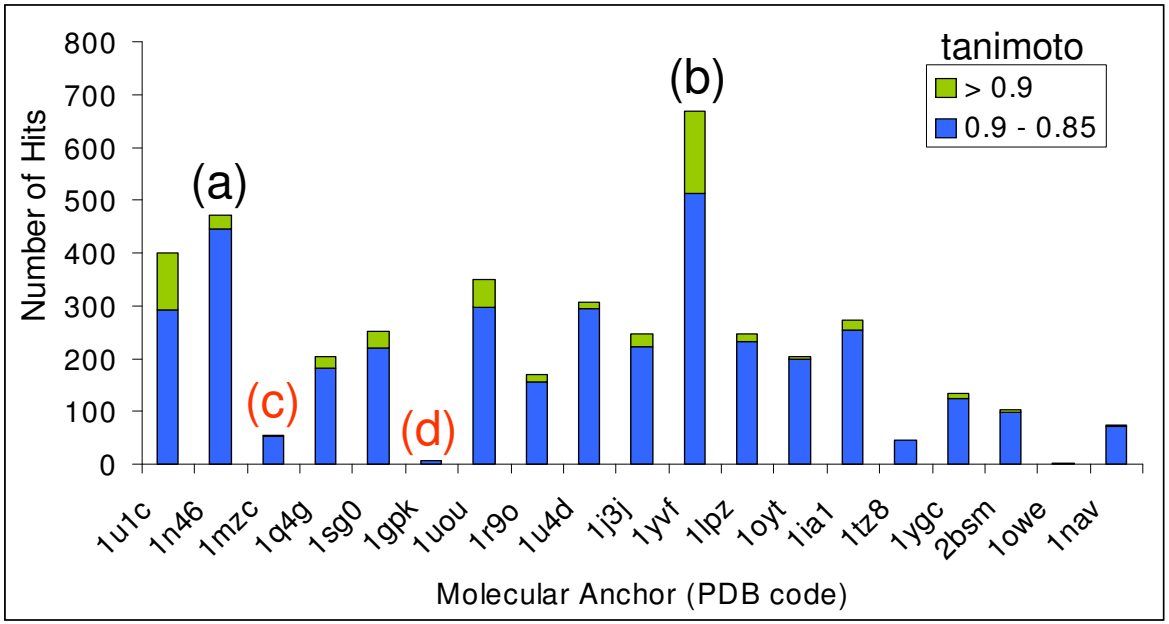




Unusual shapes have low hit rates.

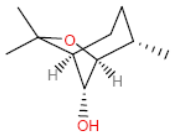
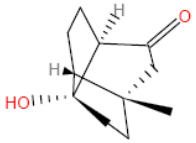
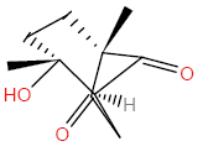
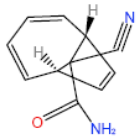
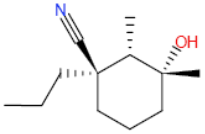
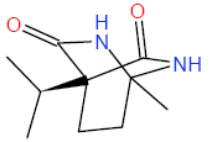
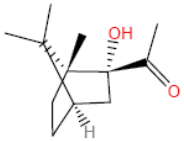
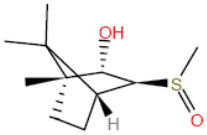
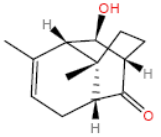
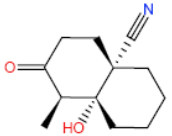
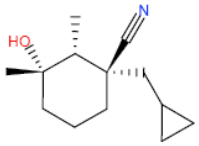
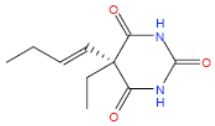
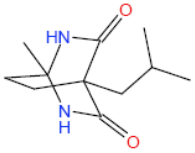
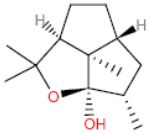
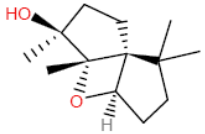
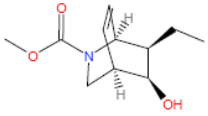
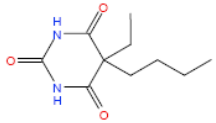
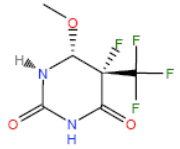
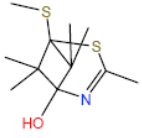
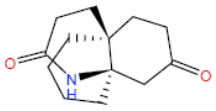
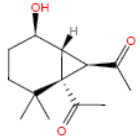
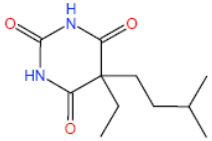
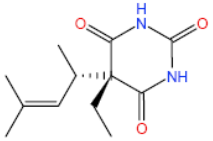
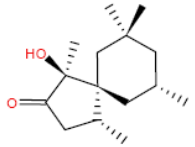
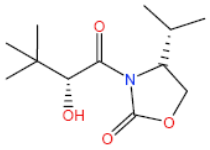
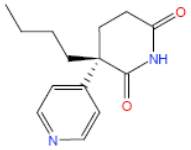
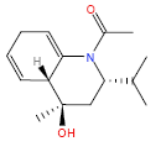
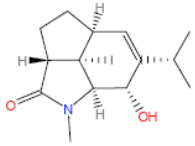
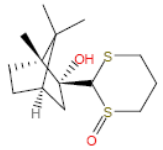
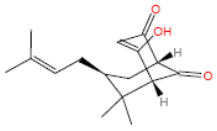


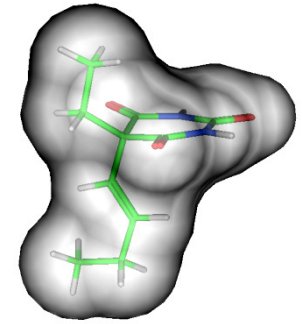


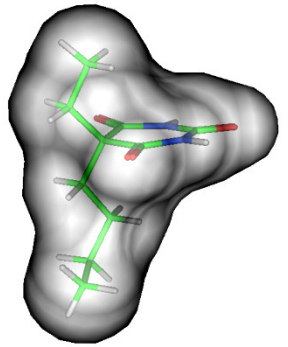


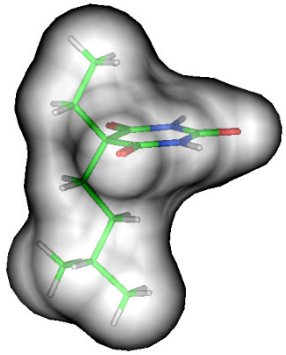
DrugbyDesign

CSD Can Also be Used as a Source for Shapes

 PAZVIW Frequency: 4	 GAXMUN Frequency: 5	 ICUQAZ Frequency: 2	 CAFJID Frequency: 4	 FODLUG Frequency: 3	 BEXNAU Frequency: 5
 NEXNIN Frequency: 3	 VADHIS Frequency: 3	 UCOYIV Frequency: 2	 VEVFAD Frequency: 4	 TEFZIO Frequency: 1	 BEBWUA Frequency: 3
 BEXMOH Frequency: 3	 SAJQEZ Frequency: 4	 GAJDEA Frequency: 5	 MAQHOC Frequency: 3	 ETBBAR01 Frequency: 4	 QEKLEX Frequency: 5
 JUQROC Frequency: 2	 HOFNOF Frequency: 2	 ACOHKT Frequency: 1	 AMBSAM10 Frequency: 2	 JIFRIZ Frequency: 2	 MATZEN Frequency: 5
 DARRAP Frequency: 5	 SIKMOO Frequency: 3	 AYIPIH Frequency: 4	 KIKSUS Frequency: 5	 DTIBOX Frequency: 4	 MOLRIO Frequency: 2

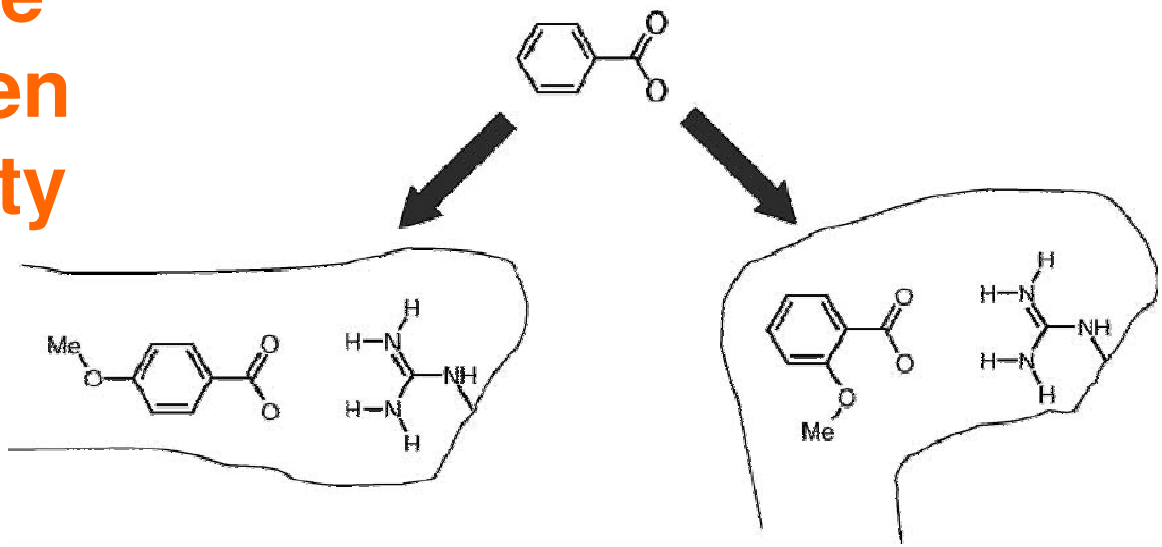






DrugbyDesign

Theory Says: Inverse Relationship Between Molecular Complexity and Hit Rate



- Simple model for ligand-receptor interactions suggests ligands of reduced complexity have a higher probability of binding.[1]
  - Corollary 1: More complexity = More ways to be incompatible with a binding site.
  - Corollary 2: Less complexity = Fewer interactions and weaker binding.

- Oh, no! Does shape diversity add to the complexity of a molecule and therefore *reduce* the observed hit rate?

DrugbyDesign

Figure: Leach, Hann, Burrows, Griffen in "Structure-Based Drug Discovery An Overview"

[1]M. Hann, A. Leach, G. Harper. JCIM 2001, 41, 856-864.

HTS Hit Rate Analysis for Fragments: No Penalty for Shape Diversity

- Retrospective analysis of 150K compounds from HTS library < 300 MW.
  - Each compound assayed in 40-100 screens.
  - 50K “active”: Hit* in one or more assays.
  - 100K “inactive”: Not a hit in any assay

	Inactive		Active	
	Mean	StdDev	Mean	StdDev
# of Assays	65.3	10.7	67.1	10.9
cLogP	2.1	1.6	2.6	1.6
MW	250	35	258	32
PSA	48.8	22.0	46.0	22.6
HB_Acceptors	3.1	1.5	3.0	1.6
HB_Donors	1.1	1.0	1.2	1.0

	Inactive		Active	
	Mean	StdDev	Mean	StdDev
Kier1	14.1	2.3	14.3	2.2
Kier2	6.4	1.6	6.5	1.5
Kier3	3.8	1.5	3.7	1.5
KierA1	11.0	2.1	11.0	1.9
KierA2	4.9	1.3	5.0	1.3
KierA3	2.9	1.3	2.8	1.3
std_dim1	3.3	0.7	3.4	0.7
std_dim2	1.6	0.3	1.6	0.3
std_dim3	0.8	0.3	0.8	0.3

- More grease trends towards more potency.
- Shape metrics do not correlate with HTS hit rates.

*Hit is defined as > 6σ from median of each assay.

DrugbyDesign

Practical Application of Key Learnings

- A typical fragment library will have poor shape diversity.
- The PDB and CSD may be good sources of diverse molecular shapes.
- A ROCS shape tanimoto of 0.85 can be used in library design to identify “shape holes”.
- Shape diversity is one of the guiding principals in fragment library design at Roche.
  - Does it work? Too early to know.
  - Metrics used to measure success.
    - Hit rates for molecular-anchor shapes.
    - Success rate in crystallography.
    - Selection rate by chemists for further optimization.

- Shape diversity does not appear to add complexity that is detrimental to hit rates.

Acknowledgements

Roche Global Fragment Library Workgroup

Caterina Bissantz
Javier DeVicente
Mike Dillon
David Fry
Tobias Gabriel
Tony Giannetti
Takeo Harada
Alexander Mayweg
Hans Maag
Takaaki Miura
Sung-Sau So
Henri Stalder

DrugbyDesign