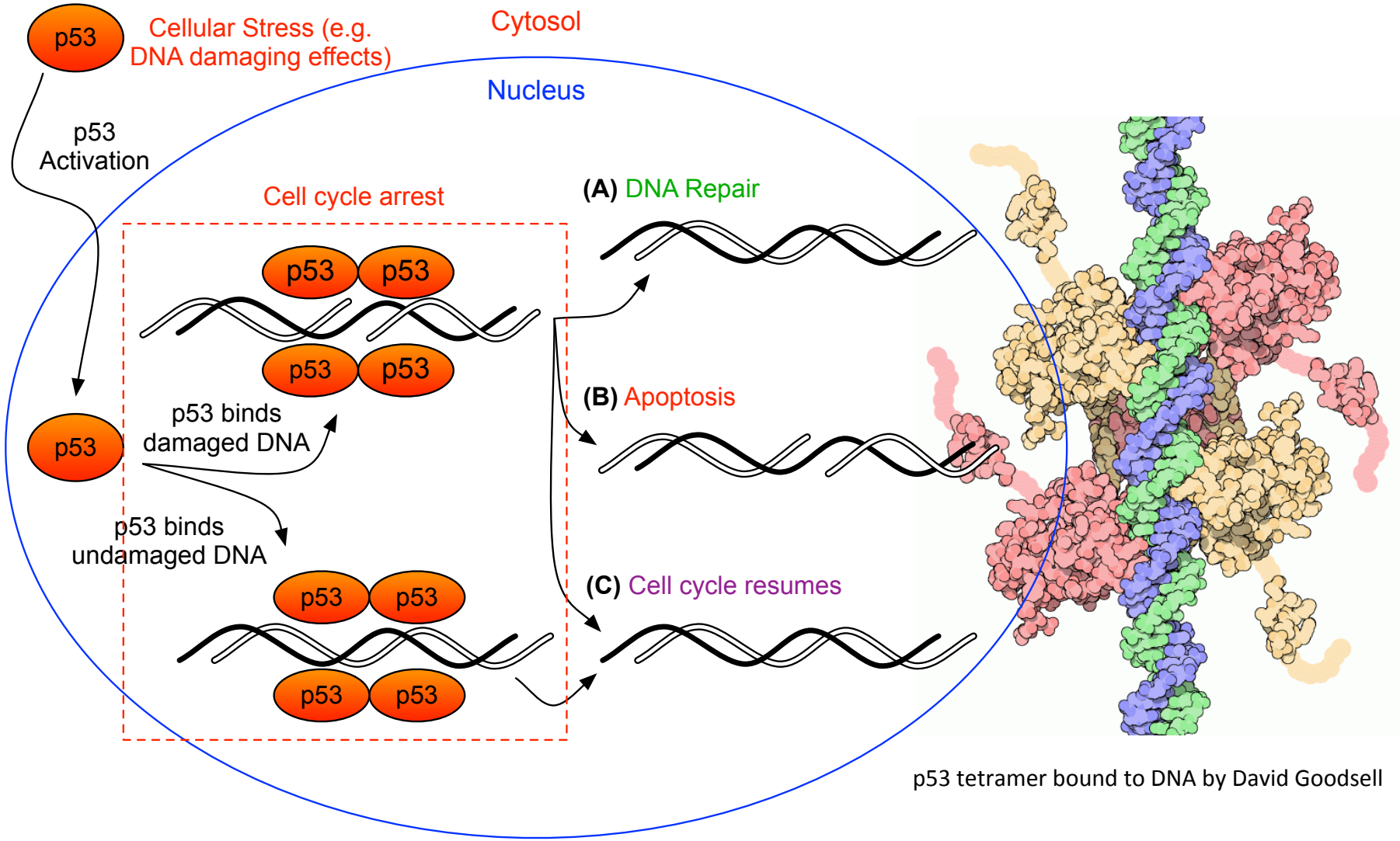
Probing the Chemical Similarity Network of Designed Inhibitors to the p53 binding site on MDM2

George A Pantelopulos, Asanga Bandara, Shanshan Song

PY 895 – Network Science

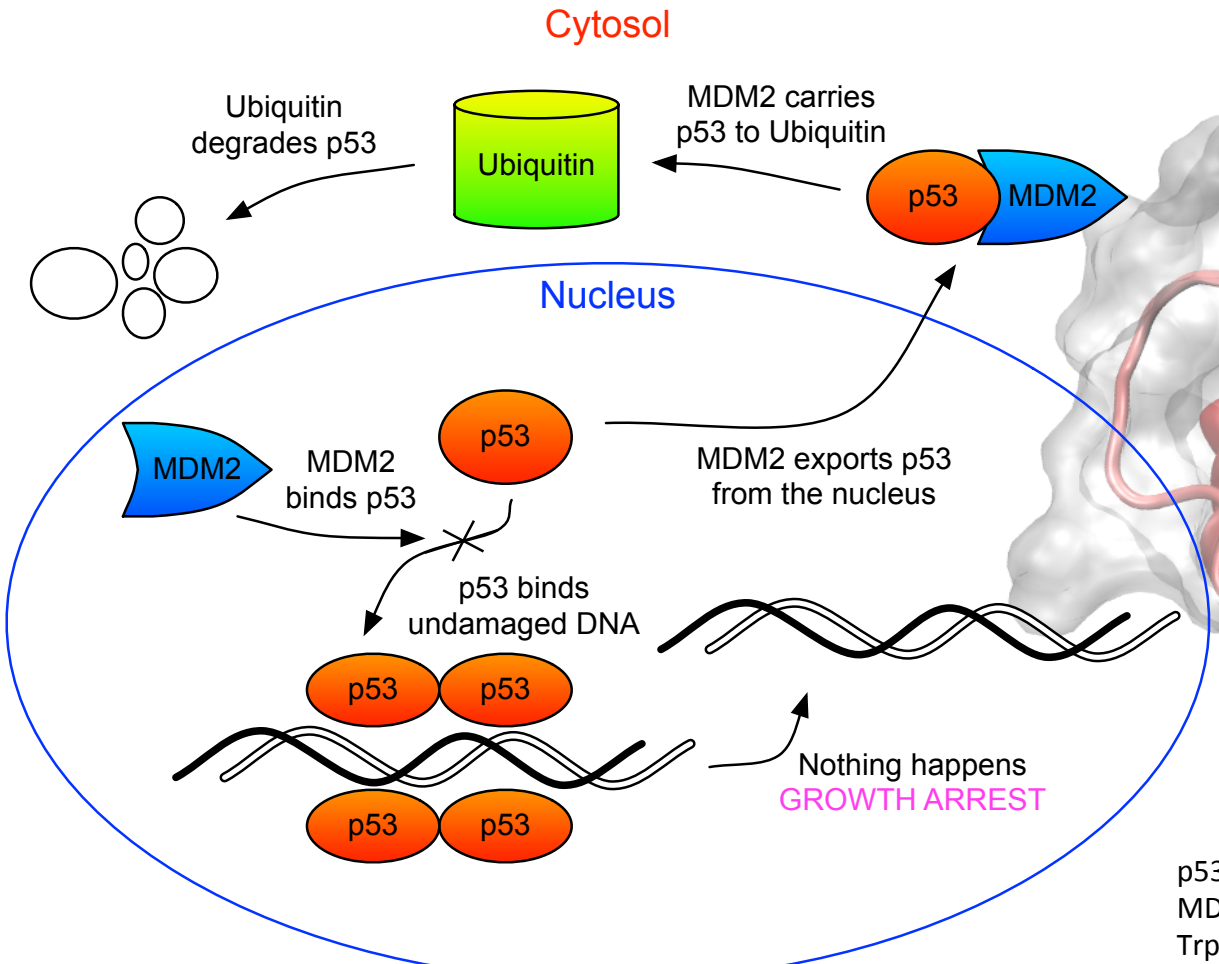
p53 is the Guardian of the Genome

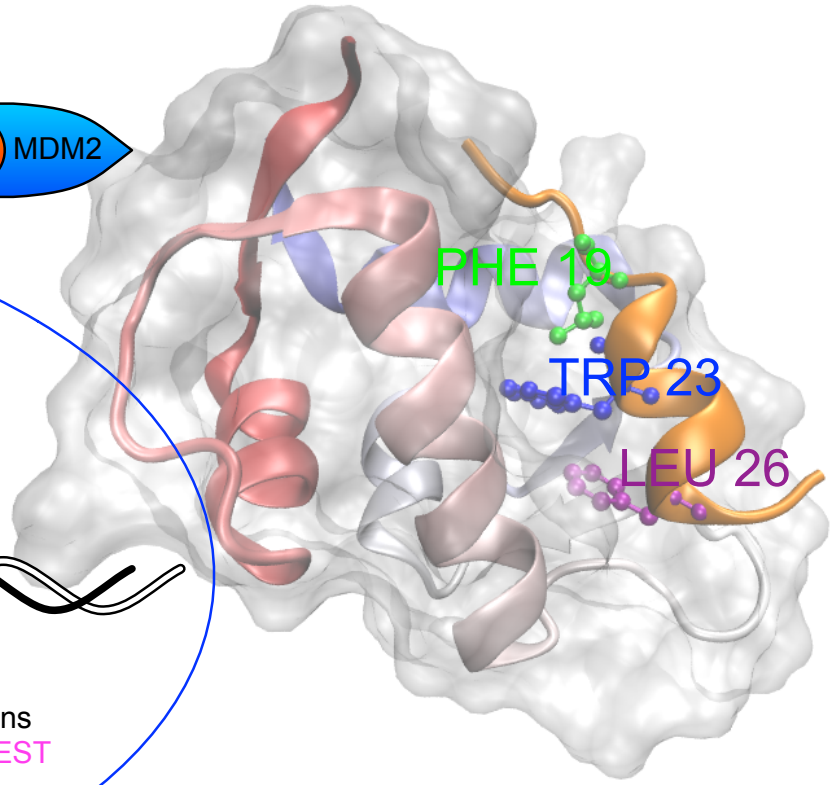


p53 tetramer bound to DNA by David Goodsell

B. Vogelstein, D. Lane, A.J. Levine (2000): Surfing the p53 network. *Nature* **408**, pp. 307-310.

MDM2 antagonizes p53 to promote healthy growth

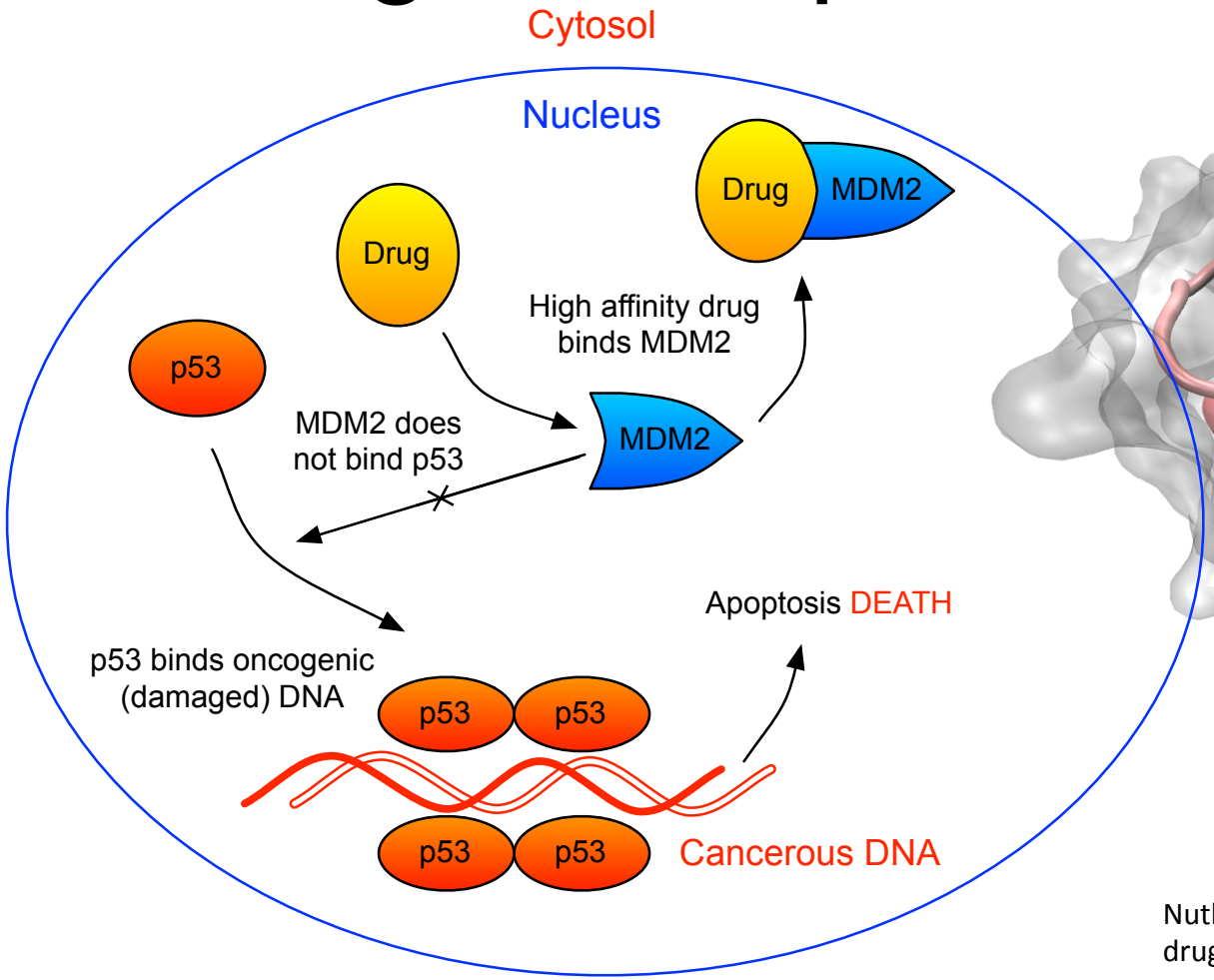


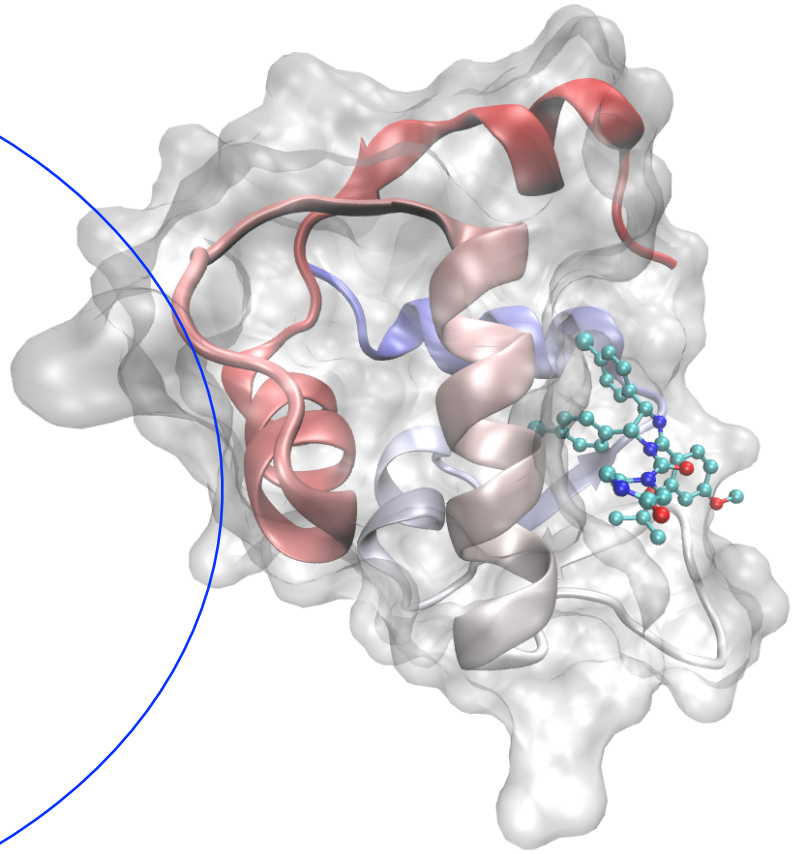


p53 N-terminal helix (orange) bound to MDM2 (PDB: 1YCR) at residues Phe19, Trp23, and Leu26

Kussie, P. H.; Gorina, S.; Marechal, V.; Elenbaas, B.; Moreau, J.; Levine, A. J.; Pavletich, N. P. *Science* **1996**, *274*, 948–953.

MDM2 is an attractive target for anti-cancer drug development





Nutlin-3a bound to MDM2 (PDB: 4HG7). A drug currently in clinical trials.

Anil, B.; Riedinger, C.; Endicott, J. A.; Noble, M. E., The structure of an MDM2-Nutlin-3a complex solved by the use of a validated MDM2 surface-entropy reduction mutant. *Acta crystallographica. Section D, Biological crystallography* **2013,** *69* (Pt 8), 1358-66.

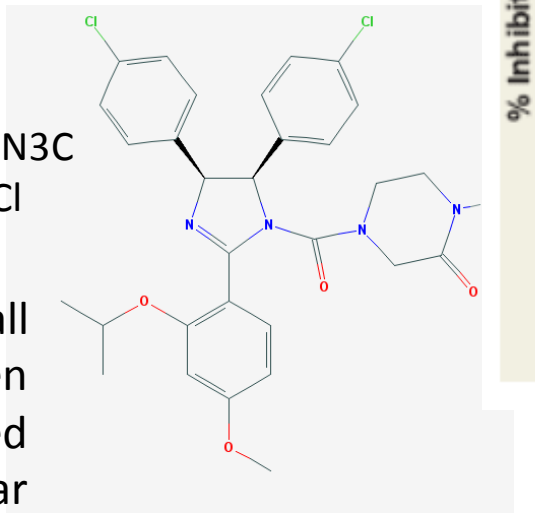
Many MDM2 inhibitor designs have been made, and have measured binding affinities

ding DB has 2,389 entries for designs that bind MDM2

93 of these designs have exactly reported IC_{50}s

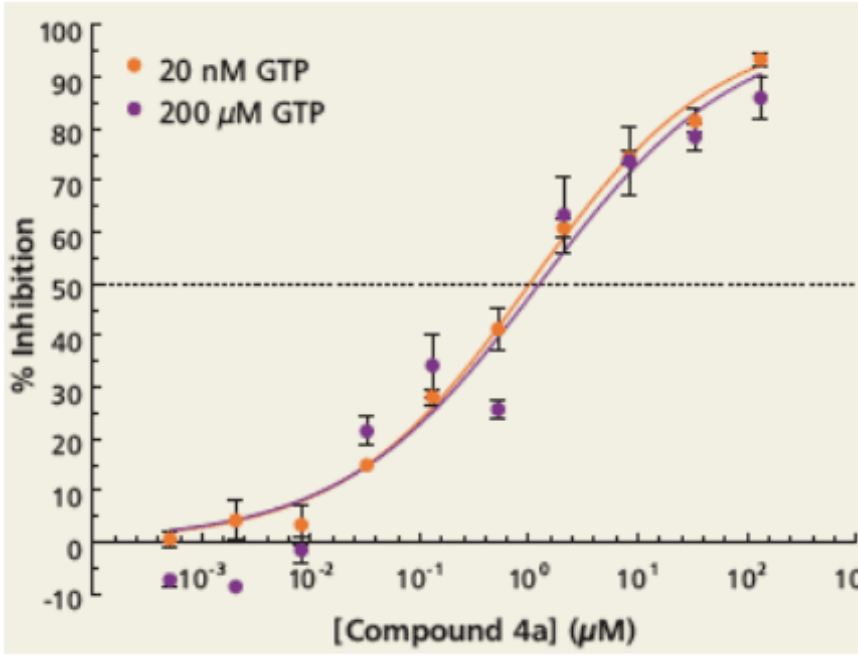
ILES (simplified molecular-input line-entry system) strings ine the 2D structure for each design.

g. the Nutlin-3a SMILES string
C(C)OC1=C(C=CC(=C1)OC)C2=NC(C(N2C(=O)N3C
NC(=O)C3)C4=CC=C(C=C4)Cl)C5=CC=C(C=C5)Cl



ld networks where nodes are small ecules and links are formed between ll molecules of similar structure be used find groups (communities) of similar ll molecules (chemotypes)?

at would such a network look like?



Example IC50 plot.[1] Dotted line represents IC_{50}.

Levin, Jules (2007) Discovery of a novel class of Hepatitis C Polymerase Inhibitor. http://www.natap.org/2007/HCV/101007_01.htm

We can define the similarity between molecules using SMILES strings based on bit-wise correlations!

Given strings that are 2D descriptors of molecular structure, we could
a string length-normalized string-similarity based method…

Tanimoto coefficients (T)! (Jaccard index)

$$T = (A \cap B)/A + B - (A \cap B)$$

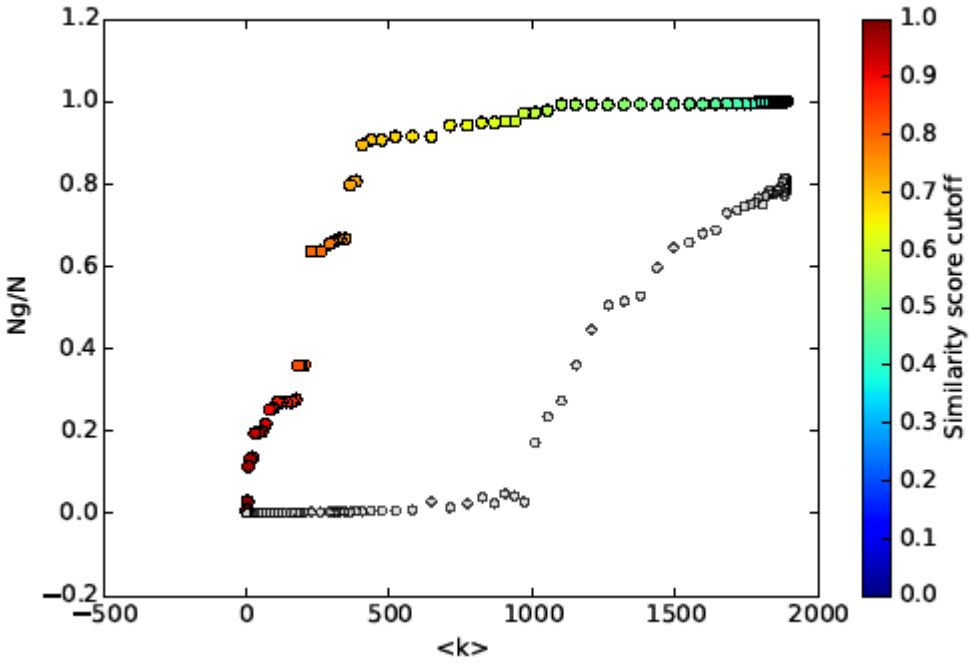
here A is the number of bits that are valued "1" in string A, B is t
mber of bits that are valued "1" in string B, and A ∩ *B* are the number
ts that are valued "1" that intersect between strings A and B.

Tanimoto similarity scores, *T*, vary from 0 to 1.

Connectivity of networks described using different similarity score linking cutoffs can give us a sense of a decent cutoff.

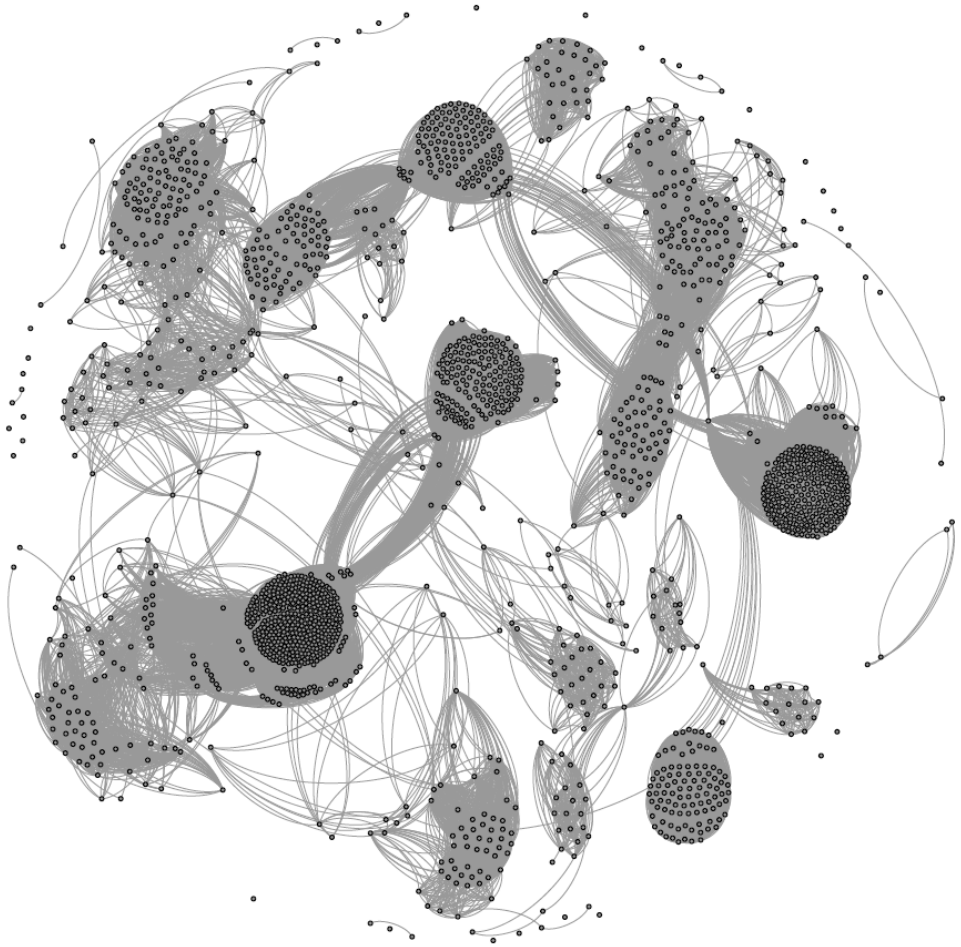
By constructing chemical similarity networks of these MDM2 inhibitors at many different similarity score cutoff values, we get a sense of the connectivity of the network for each cutoff…

A similarity score linking cutoff of 0.65 was selected to ensure a high connectivity while being far from a complete graph and far from above random.

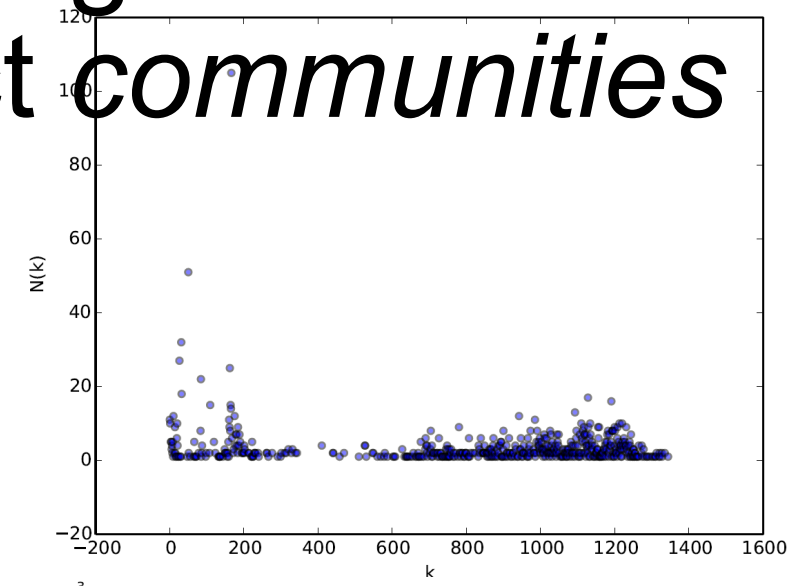


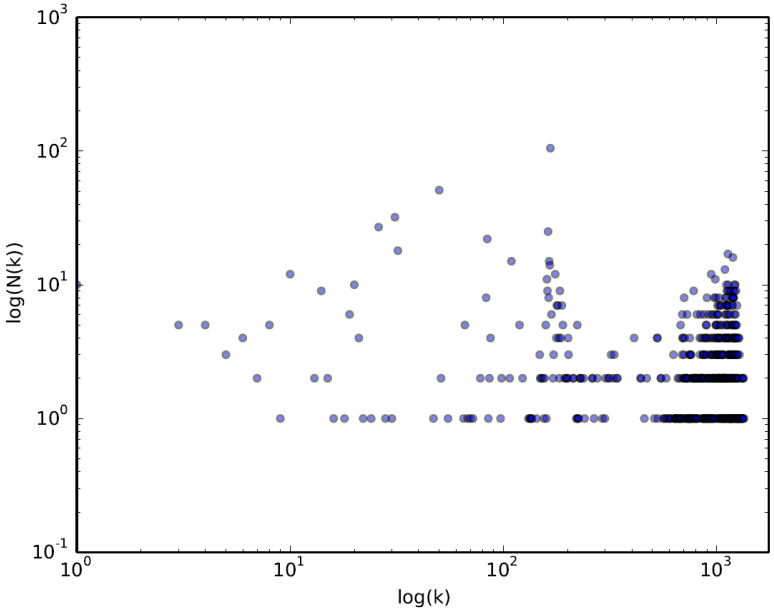
Phase diagram of MDM2 inhibitor chemical similarity networks at different similarity score linking cutoffs. Ng is the number of nodes in the largest connected component, N is the number of nodes in the network, and <k> is the average degree of nodes in the network. ER graphs are constructed at each <k> for comparison, and are represented in gray.

Our resulting network is large and appears to contain distinct *communities*



Representation with Fruchterman-Reingold force-directed drawing shows multiple possible partitions.





We partition our network into distinct communities via the Louvain Method

Optimizes the *modularity*, *Q*, a quantity that measures the density of links within communities of a network.

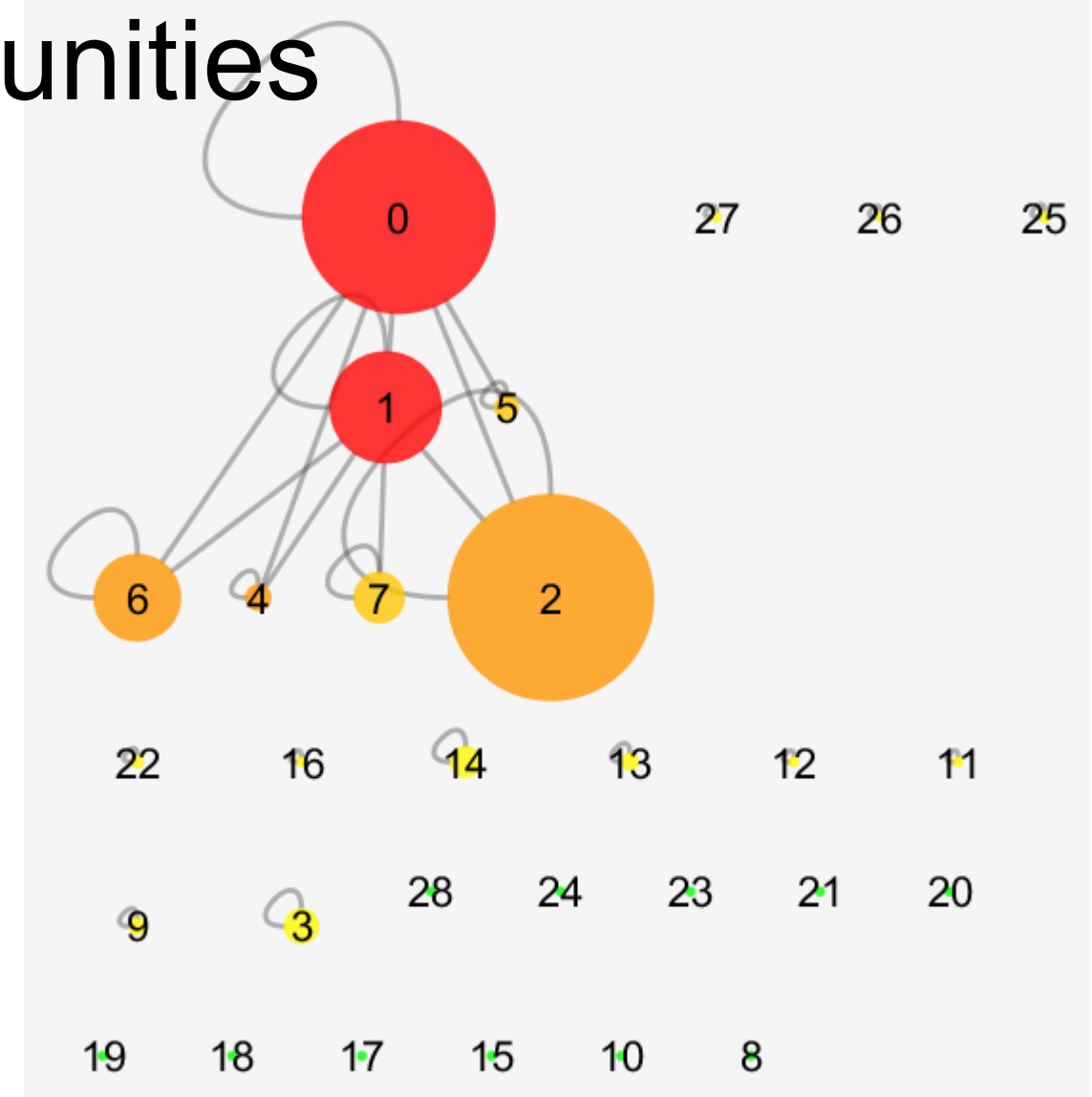
$$Q = 1/2m \sum_{ij} [A_{ij} - k_i k_j / 2m] \delta(c_i, c_j)$$

Where *i and j* are the indices of nodes in the network, m is half the sum of edge weights in the graph (weights = 1 here), c_i and c_j are the communities which each node belongs, and A_{ij} is the edge weight between *i* and *j* (=1 here

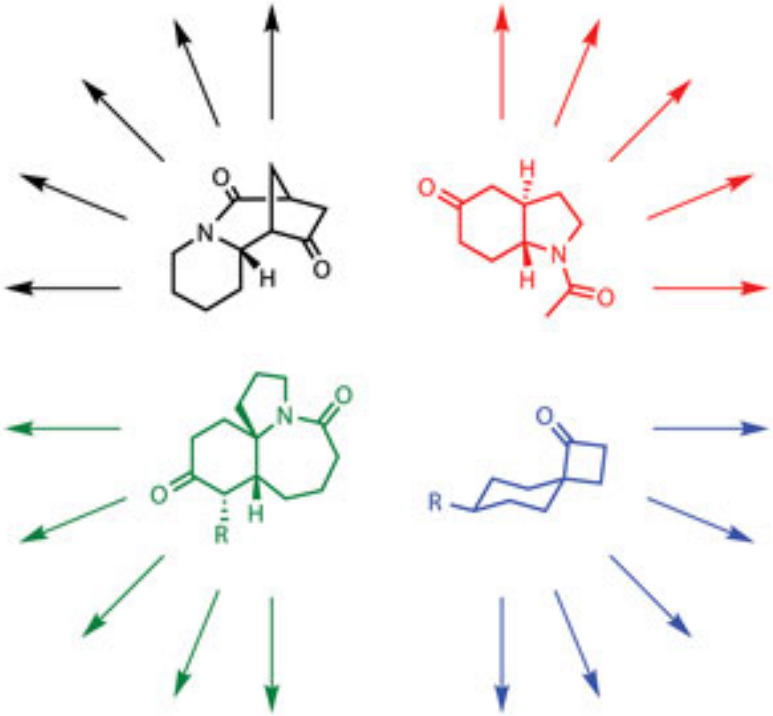
The assignments of nodes to communities is accomplished by whatev discretization of nodes into communities maximizes *Q*.

Our network contains a handful of well-populated communities

Community	Population
0	**602**
1	**267**
2	**657**
3	32
4	21
5	19
6	**166**
7	51
8	1
9	10
10	1
11	2
12	2
13	11
14	27
15	1
16	2
17	1
18	1
19	1
20	1
21	1
22	5
23	1
24	1
25	4
26	2
27	2



Community Representatives



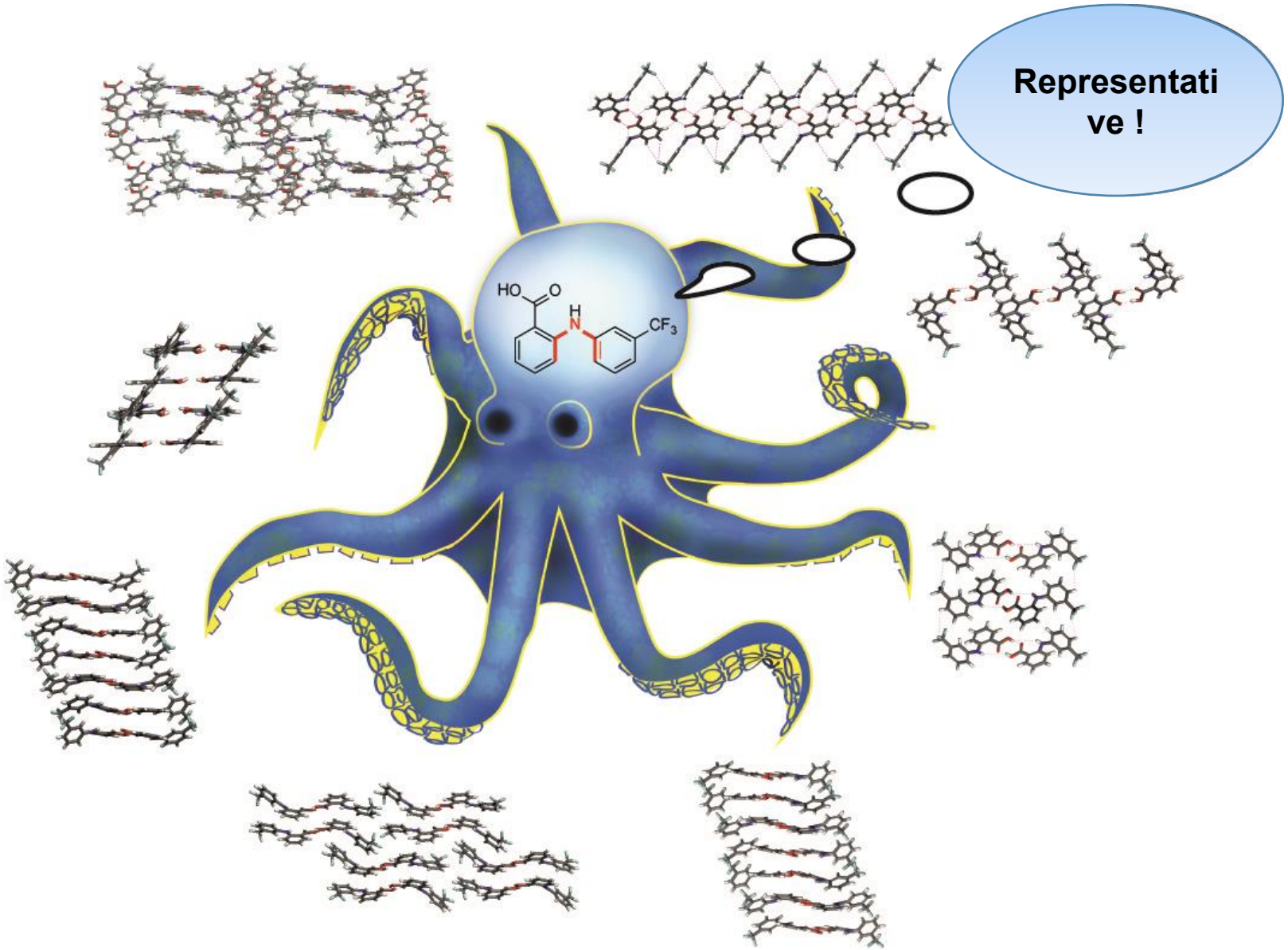
Motivation

Large number of members in a community.

Individual analysis is time consuming.

Chemical intuition is narrow with individual analysis.

Not member molecules but scaffolds (Functional groups, Steric groups) deemed more interesting !


Representati
ve !
HO
O
H
N
CF3

Approach

Nodes with highest links : Representative ?

Links have different weights

Cumulative weighted links

Representative Molecule for a Community

=

Node with the Highest Weighted Degree

nvestigating network and chemical structures n our communities may reveal interesting and useful features…

We briefly investigate network and chemical structures in our four most-populated communities via the following:

1. Visualization with force-directed drawing.
2. Log-log degree distributions
3. Representative molecule structures
4. Similarity maps for other molecules in communities in reference to community representatives

niker and Landrum *Journal of Cheminformatics* 2013, **5**:43
ttp://www.jcheminf.com/content/5/1/43

Journal of Cheminformatics

Similarity Maps

Structure features

- functional group
- 3D structure

Advantage of Conversion

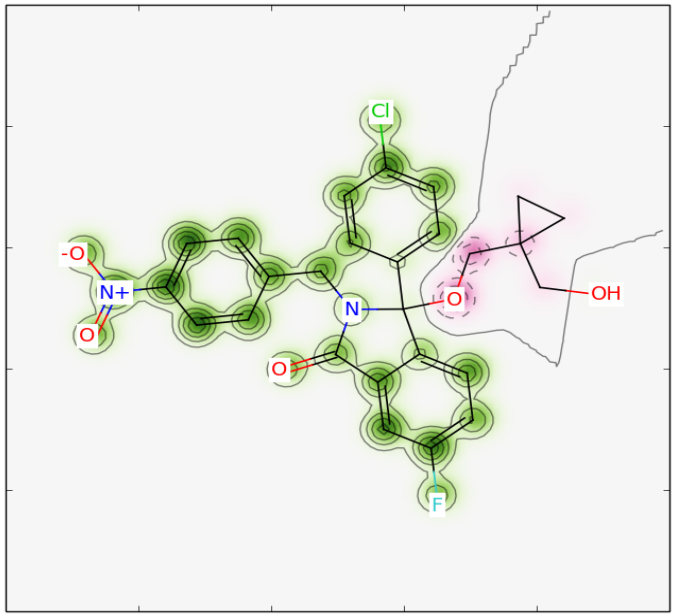
- computationally efficient
- compare similarity

Fingerprints

- bits in vector
- count in vector

Weight -- Similarity

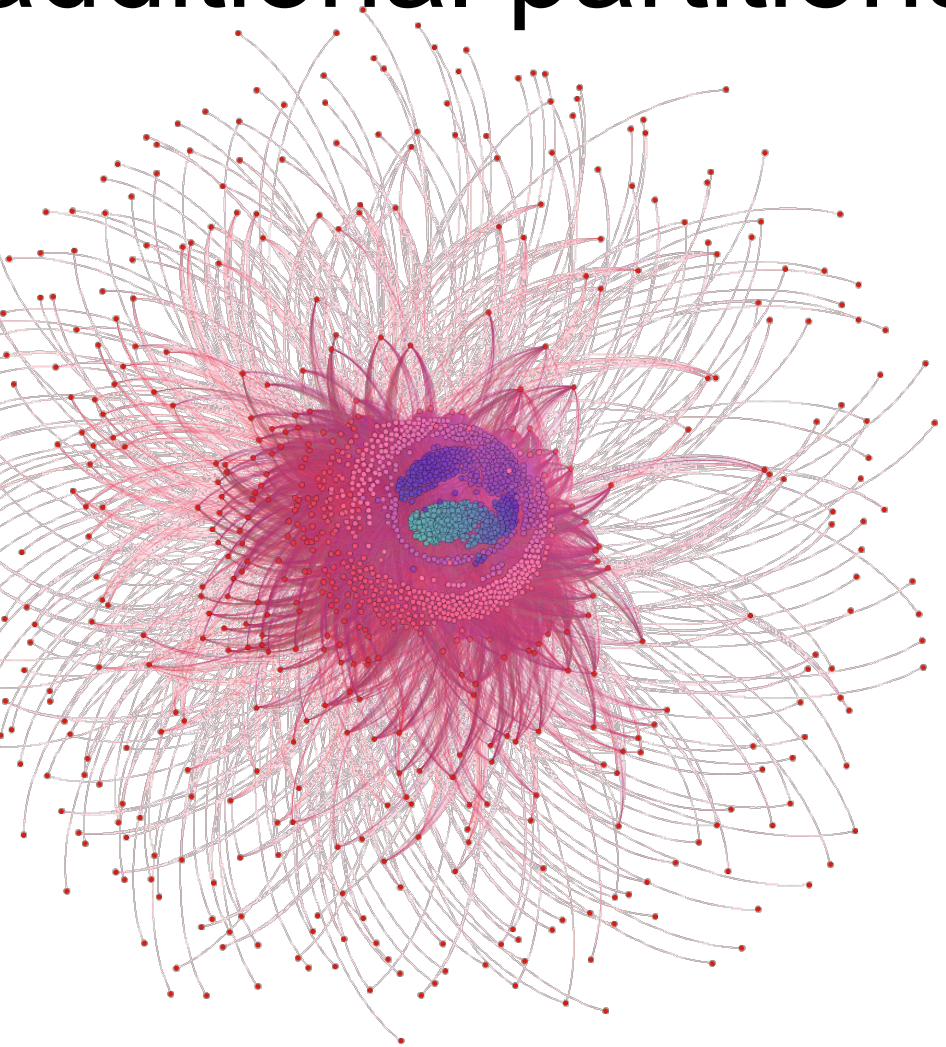
- Remove bits of one atom
- Green: positive difference
- Pink: negative difference
- Grey: no difference

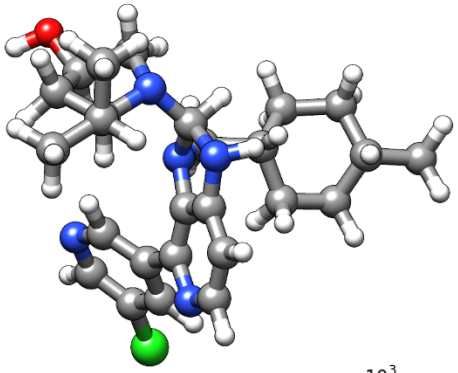


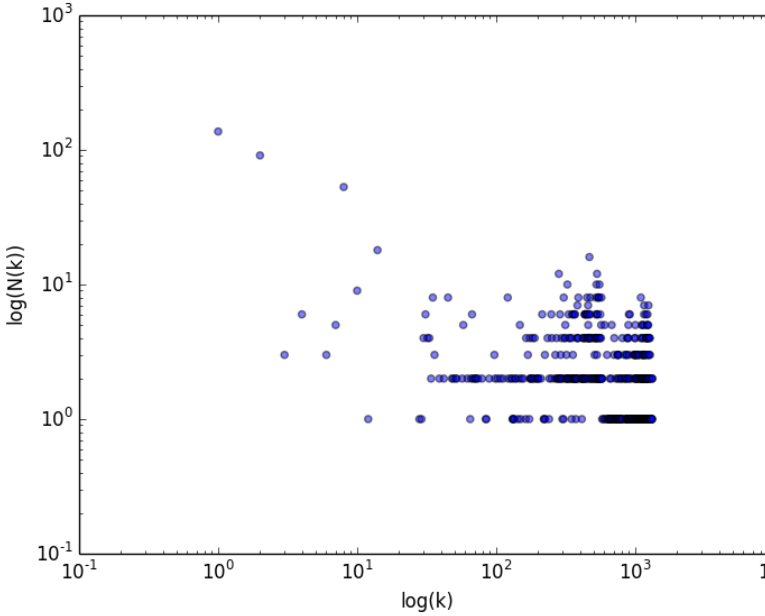
FeatMorgan2

- FEATURE based
- donar,acceptor,aromatic,basic,acid
- CIRCULAR FINGERPRINT
- w/ RADIUS 2 (radius=2)
- as BIT VECTOR (size=1024bit)

Community 0 appears to show no additional partitions

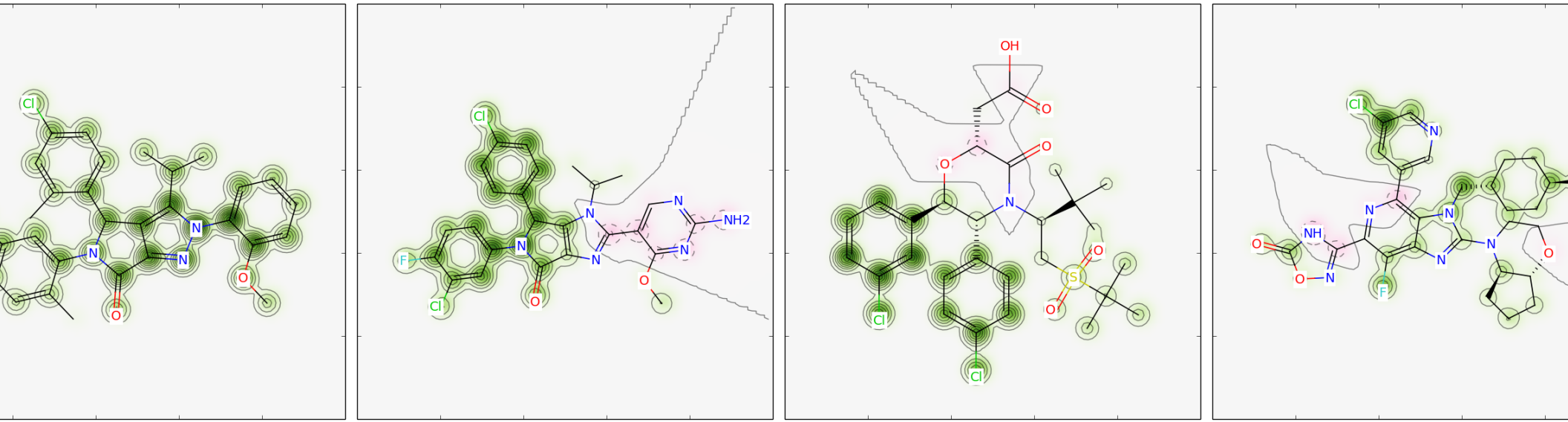






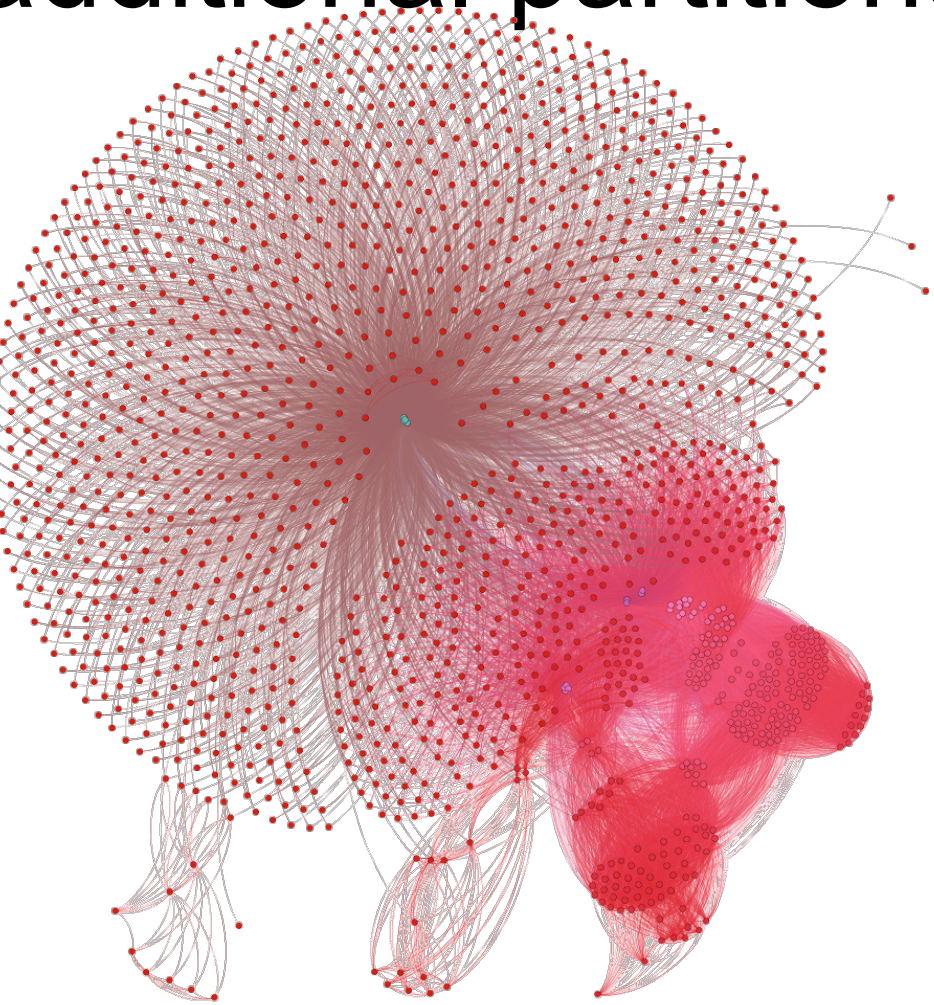
man-Reingold force-directed drawing shows no partitions.

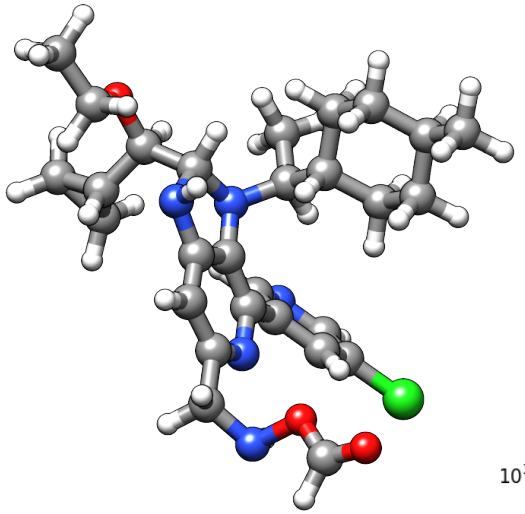
Community 0 similarity depends on aromatic nitrogen-containing scaffold and aromatic chlorine-containing sidechains.

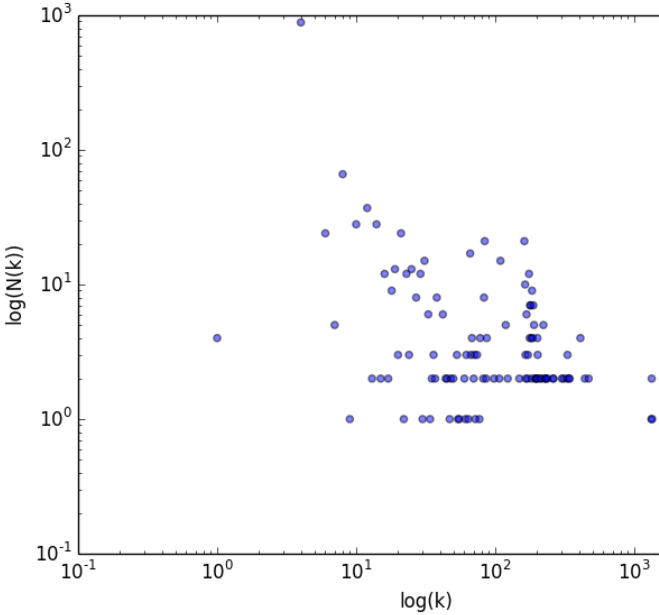


larity maps based on per-atom similarity with the representative molecule. 2nd, 3rd, 4th and 5th highest-larity molecules presented from left to right.

Community 1 appears to contain additional partitions

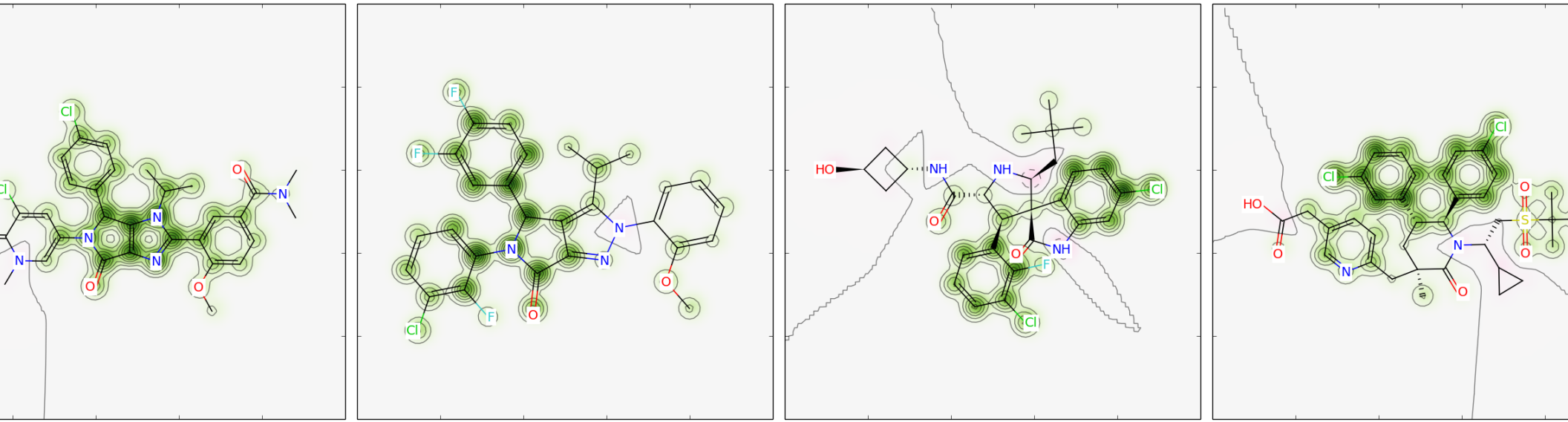






man-Reingold force-directed drawing shows possible further partitions.

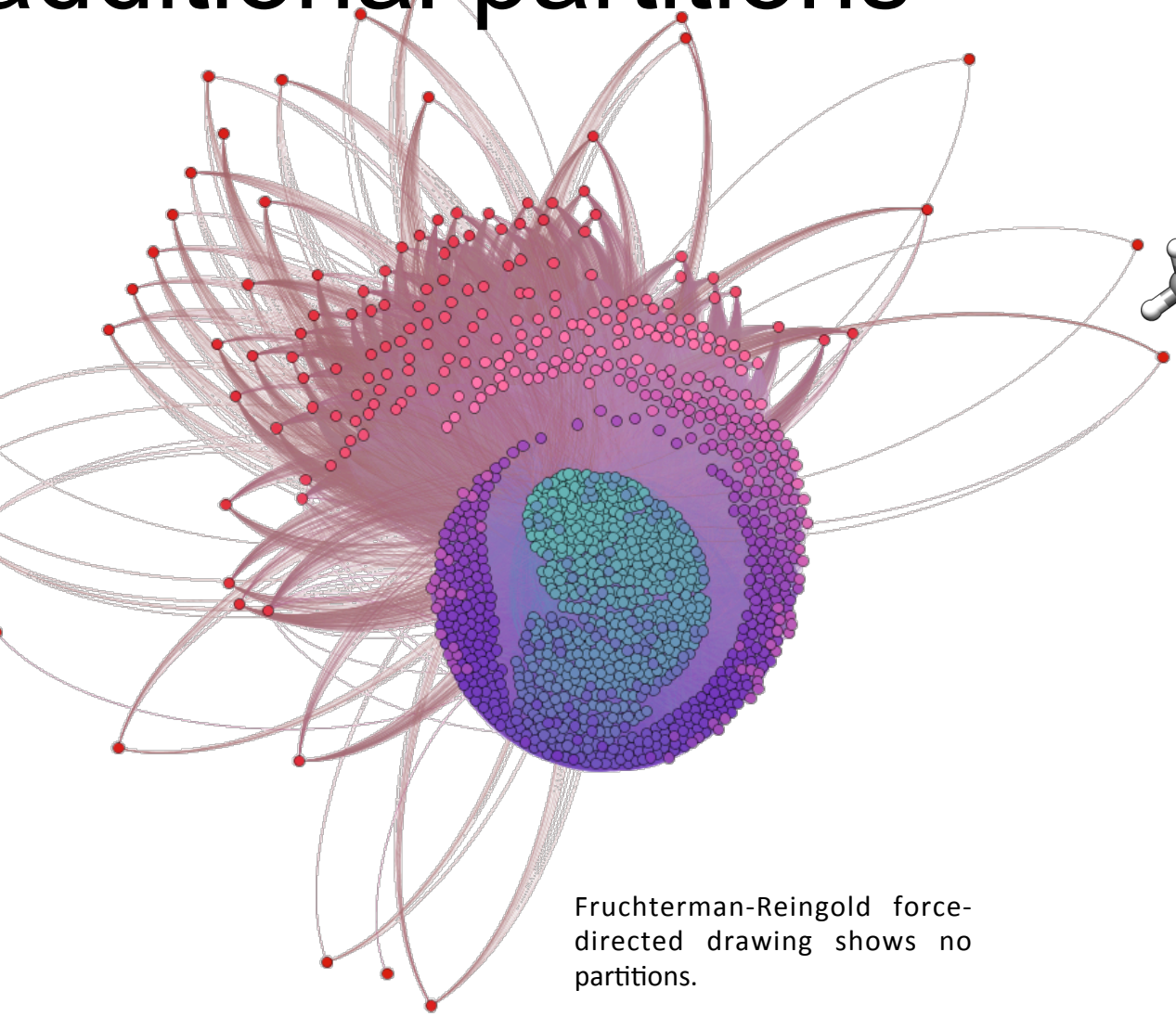
Community 1 similarity depends on aromatic chlorine and fluorine-containing side chains.



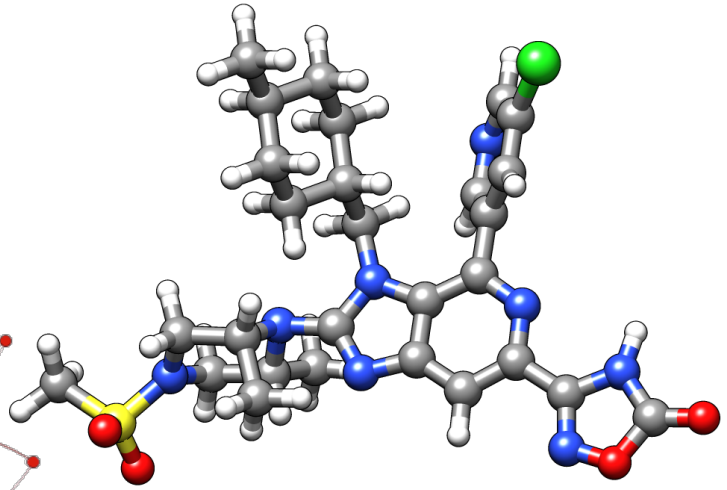
larity maps based on per-atom similarity with the representative molecule. 2nd, 3rd, 4th and most-esentative molecules presented from left to right.

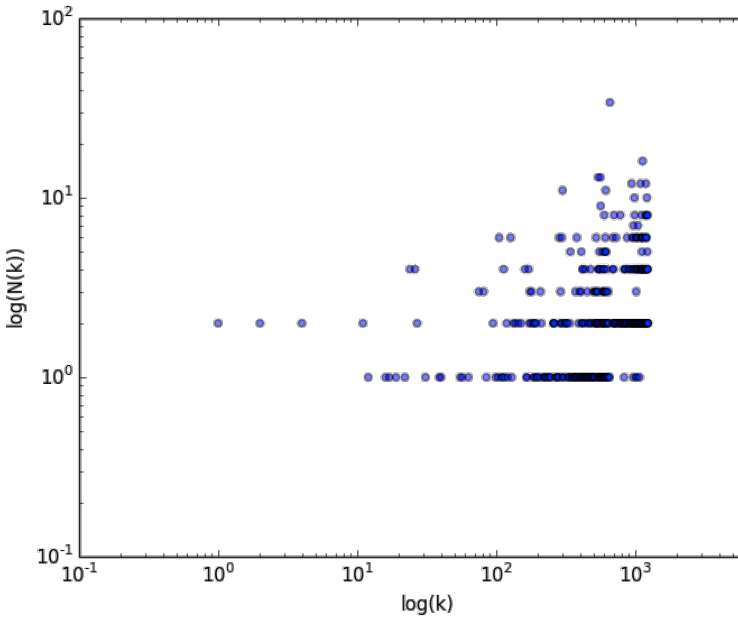
sible that molecules out of the top 3 most-representative molecules might be in need of further partitioning.

Community 2 appears to show no additional partitions

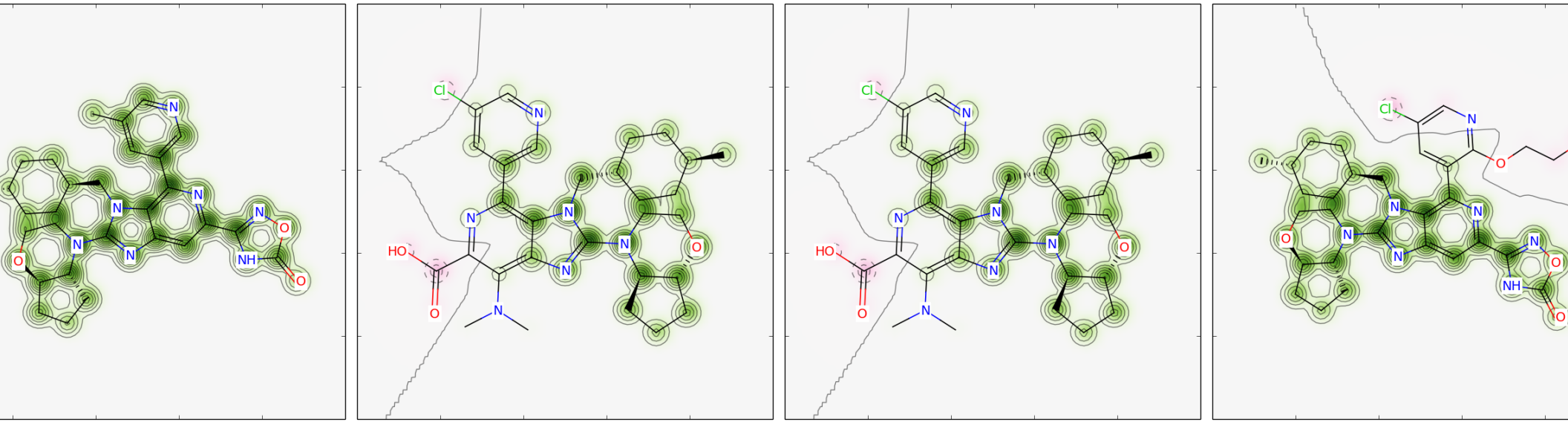


Fruchterman-Reingold force-directed drawing shows no partitions.



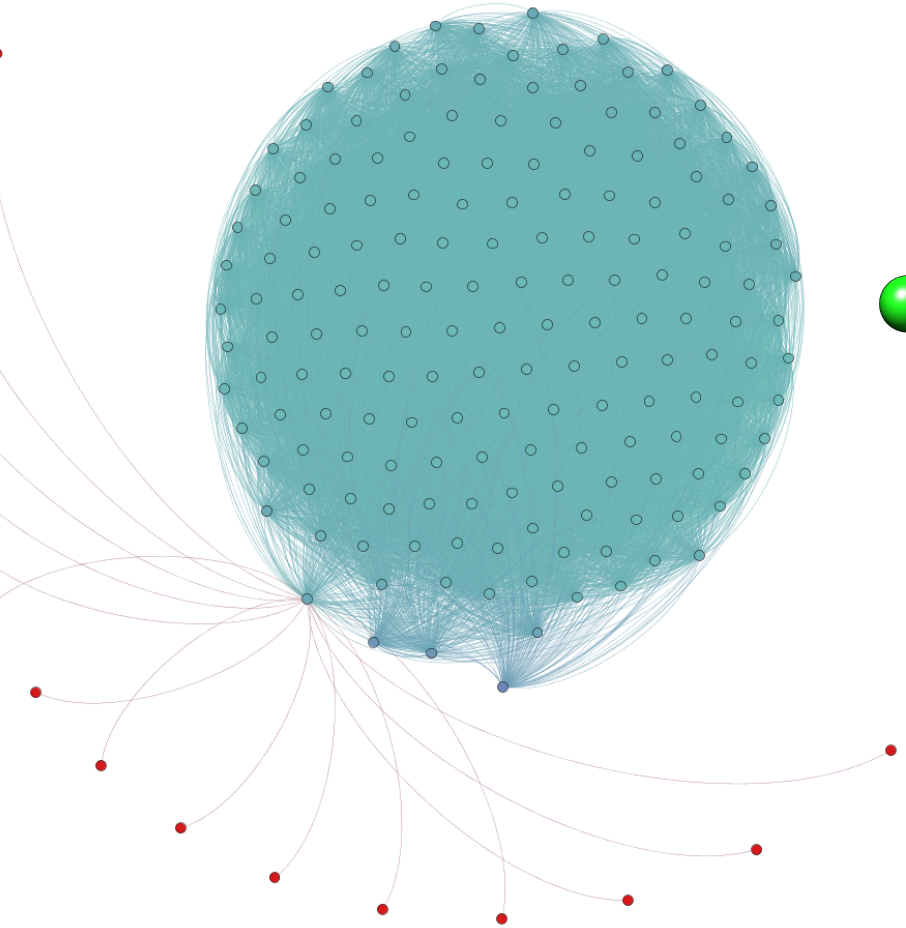


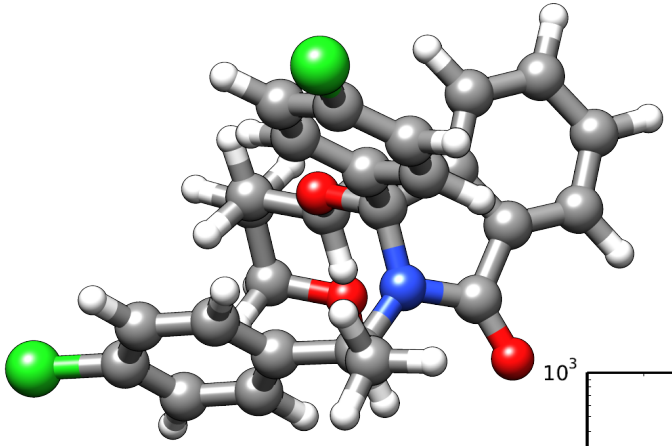
Community 2 is characterized by nitrogen-containing aromatic scaffolds.

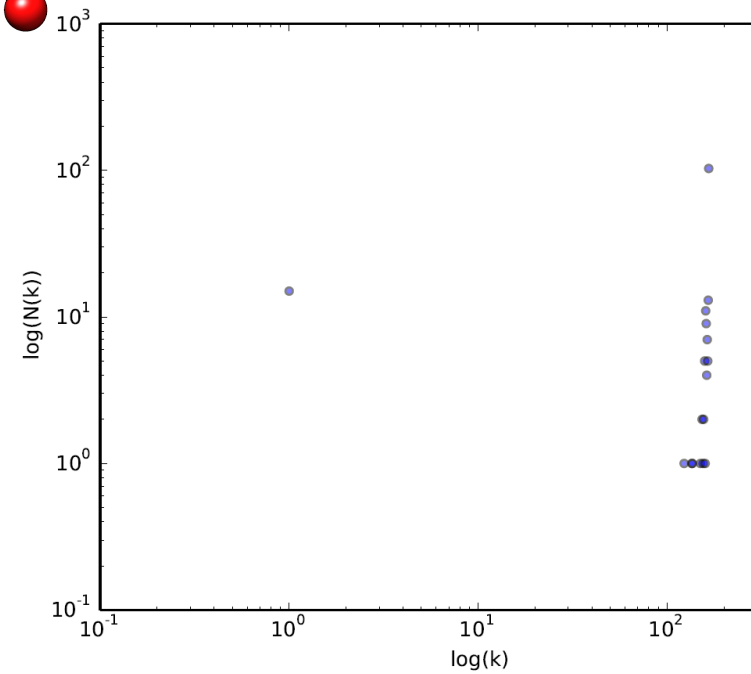


larity maps based on per-atom similarity with the representative molecule. 2nd, 3rd, 4th and 5th most-esentative molecules presented from left to right.

Community 6 appears to show no additional partitions, but a handful of outliers

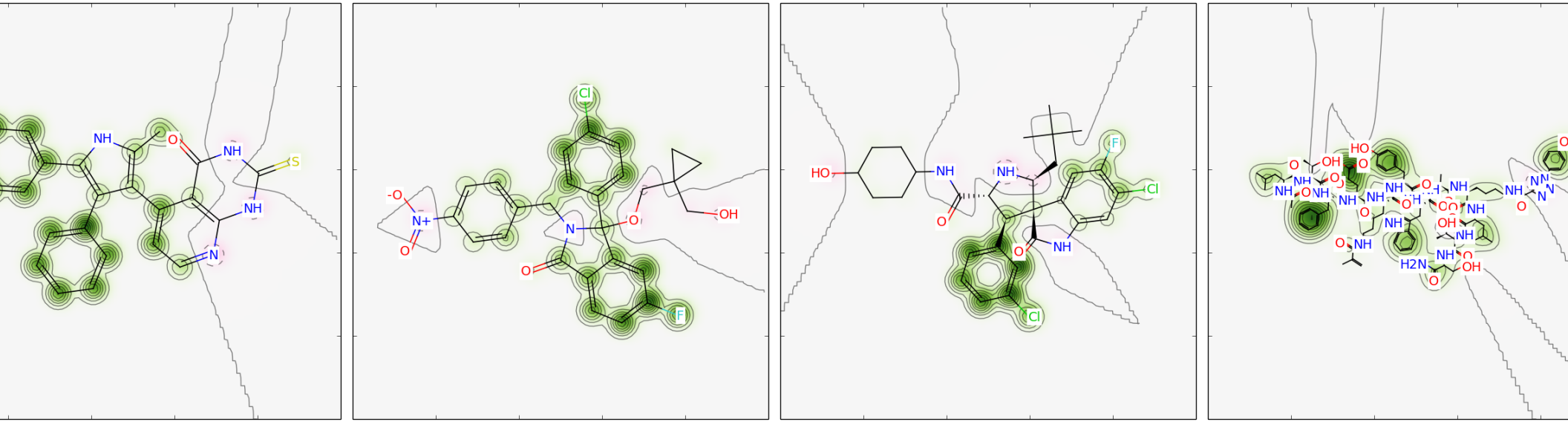






rman-Reingold force-directed drawing shows no partitions – shows outliers.

Community 6 contains well-defined aromatic side chains.



larity maps based on per-atom similarity with the representative molecule. 2nd, 3rd, 4th and 5th most-esentative molecules presented from left to right.

Future work(?)

- Building a small molecule inhibitor network where links are weighted b similarity score and there is no similarity score cutoff – would th communities be any more well-defined?
- Using a different algorithm for partitioning molecules into communities?
- Characterizing small molecules within communities in terms of their IC_{50} and unique molecular structure.
- Generating semi-random new 2D molecular designs that would fall int our communities. Binding affinity could be approximately predicted usin molecular docking to the p53 binding site on MDM2... Can we generat potent new designs algorithmically using our communities?