3D Similarity; Methods and Algorithms

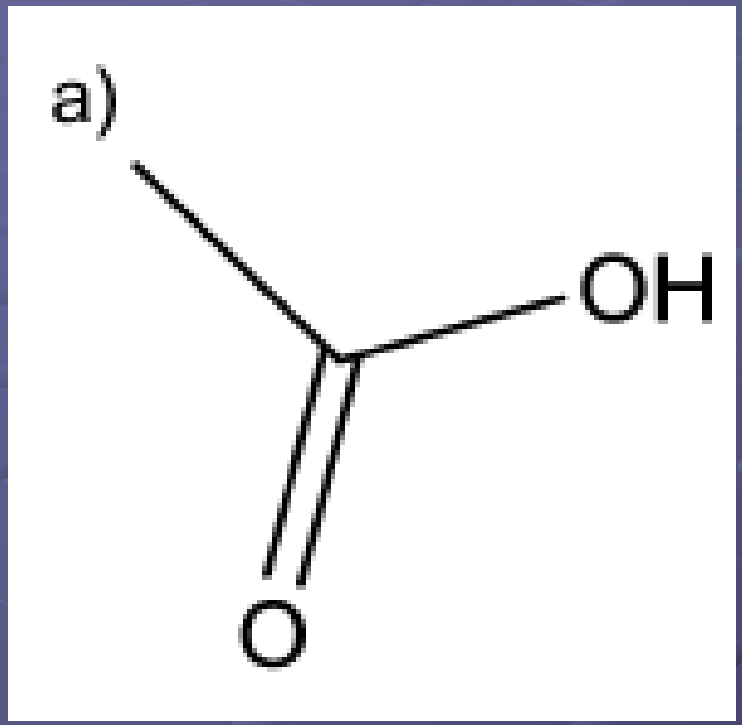
Oleg Ursu PhD student
Mircea Diudea Professor Dr.

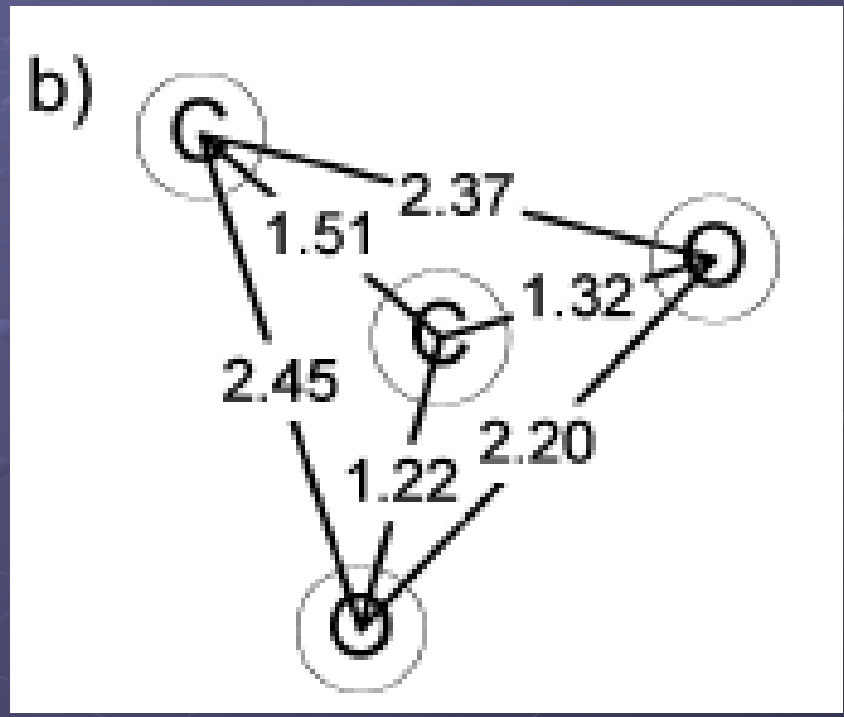
Contents

- *Introduction*
- *Distance Geometry and Bound Smoothing*
- *Clique detection*
- *Validation study*

- *New method for 3D similarity searching in chemical databases*
- *Applications – virtual screening and lead discovery in rational drug design*
- *Higher accuracy in quantifying the degree of structural similarity than other widely used methods based on bit-strings representation of the molecular structure*

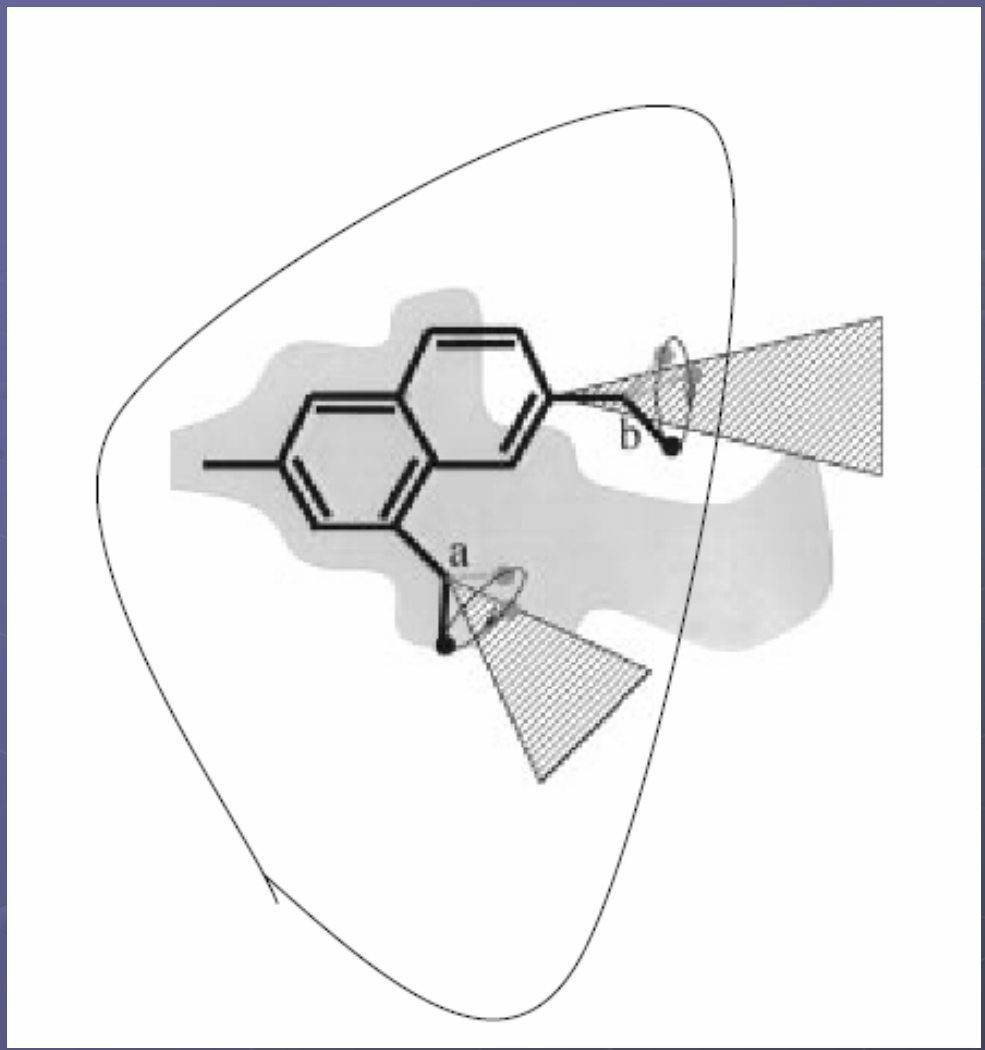
Molecular structures are represented as weighted, undirected, labeled graphs





In (b) all edges are weighted by the corresponding geometric distance between pairs of atoms in acetic acid

- *Molecules usually exhibit some degree of conformational flexibility – distances between atoms are not fixed*
- *Generate all suitable conformations and compare them to determine the similarity? – not a very efficient approach and does not guarantee that the optimal similarity has been identified*
- *Used approach – Distance Geometry*



Conformational flexibility example

- *Distance geometry establishes the maximum and minimum possible distances between all pairs of atoms in a molecule by using bound smoothing procedure*
- *Given an initial set of upper and lower bounds between all pairs of atoms in a molecule Bound Smoothing serves to eliminate geometric inconsistencies by iteratively sharpening the upper and lower bounds.*
- *How do we get the initial set of upper and lower bounds?*

- *Constrains defined by a set of simple rules known to each chemist*
- *Two types of constrains – distance and chirality constrains*
- *Distance constrains are the upper and lower bounds on the interatomic distances*
- *Chirality constrains are imposed by asymmetric centers in the molecule*

Covalent distance constrains impose constrains between covalent bonded pairs of atoms – defined by the low of cosines

$$d_{13}^2 = d_{12}^2 + d_{23}^2 - 2d_{12}d_{23}\cos(\theta)$$

$$d_{13}^2 = l_{13}^2 + \left(u_{13}^2 + l_{13}^2\right)\sin^2\left(\frac{\theta}{2}\right)$$

Vicinal distance constrains impose constrains on 1,4 distance or cis/trans bounds

$$d_{14}^2 = l_{14}^2 + \left(u_{14}^2 - l_{14}^2\right)\sin^2\left(\frac{\varphi}{2}\right)$$

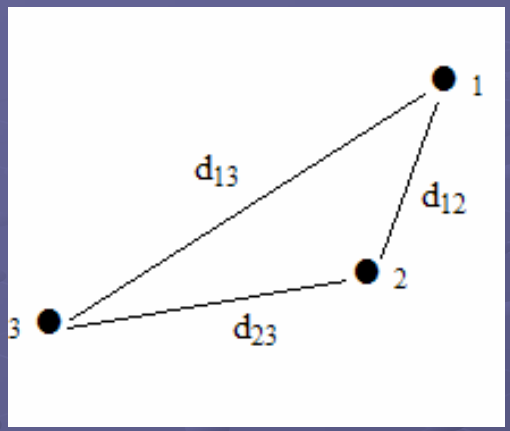
Chirality constrains are expressed in terms of the Cartesian coordinates by the sign of the determinant

$$\chi_{1234} = \mathrm{sgn}\left(\det \begin{bmatrix} 1 & 1 & 1 & 1 \\ x_1 & x_2 & x_3 & x_4 \\ y_1 & y_2 & y_3 & y_4 \\ z_1 & z_2 & z_3 & z_4 \end{bmatrix} \right)$$

Steric distance constrains set lower bounds to the sum of the van der Waals radii

$$l_{ij} = r_i + r_j$$

- ***Triangle bound smoothing – triangle inequality***



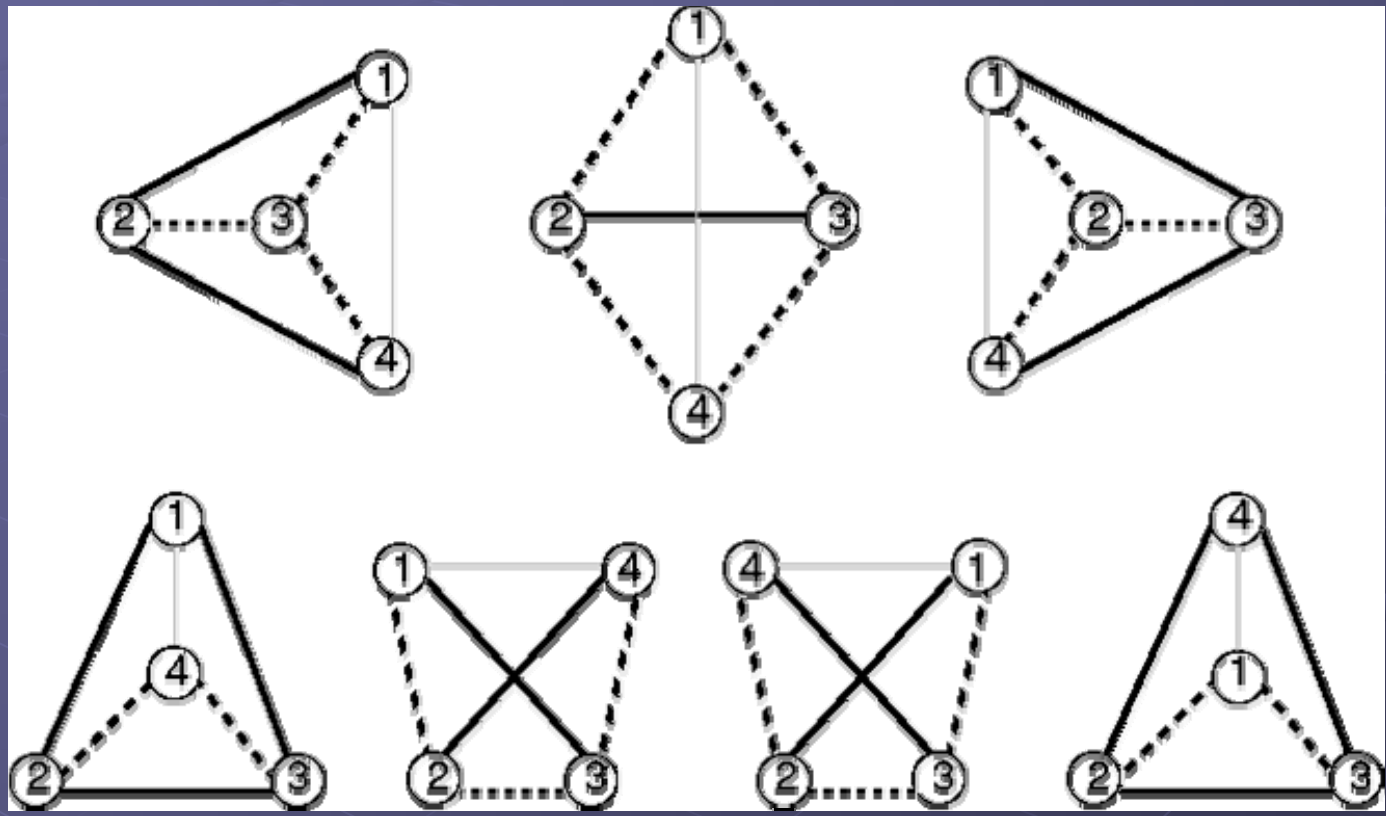
$$d_{ij} \leq d_{ik} + d_{jk}$$

The upper and lower bounds derived from the triangle inequality

$$u_{ij} \leq u_{ik} + u_{jk}$$

$$l_{ij} \leq l_{ik} + l_{jk}$$

Tetrangle bound smoothing – tetrangle bound smoothing



Tetrangle inequality limits

Tetrangle inequality limits expressed in terms of Cayley-Menger determinants

$$0 \le CM\left(d_{12},\ldots,d_{34}\right) = \det\begin{pmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & d_{12}^2 & d_{13}^2 & d_{14}^2 \\ 1 & d_{12}^2 & 0 & d_{23}^2 & d_{24}^2 \\ 1 & d_{13}^2 & d_{23}^2 & 0 & d_{34}^2 \\ 1 & d_{14}^2 & d_{24}^2 & d_{34}^2 & 0 \end{pmatrix}$$

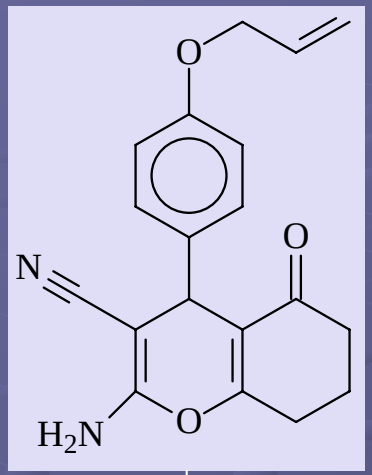
Upper tetrangle inequalities

$$\begin{aligned} & 0 \le CM\left(l_{12},u_{13},u_{14},u_{23},u_{24},u_{34}\right), \\ \text{or} \quad & 0 \le CM\left(u_{12},l_{13},l_{14},u_{23},u_{24},u_{34}\right), \\ \text{or} \quad & 0 \le CM\left(u_{12},u_{13},u_{14},l_{23},l_{24},u_{34}\right), \end{aligned}$$

Lower tetrangle inequalities

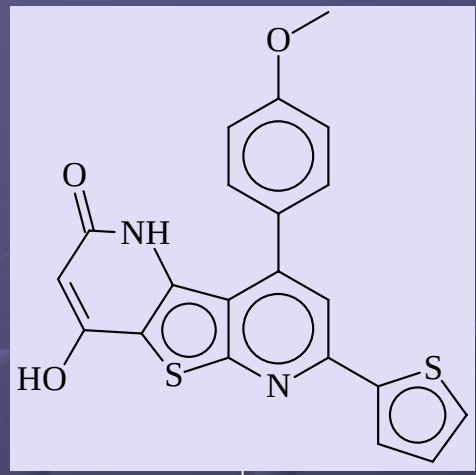
$$\begin{aligned} & 0 \le CM\left(u_{12},u_{13},l_{14},l_{23},u_{24},l_{34}\right), \\ \text{or} \quad & 0 \le CM\left(u_{12},l_{13},u_{14},u_{23},l_{24},l_{34}\right), \\ \text{or} \quad & 0 \le CM\left(l_{12},l_{13},u_{14},l_{23},u_{24},l_{34}\right), \\ \text{or} \quad & 0 \le CM\left(l_{12},u_{13},l_{14},u_{23},l_{24},l_{34}\right), \end{aligned}$$

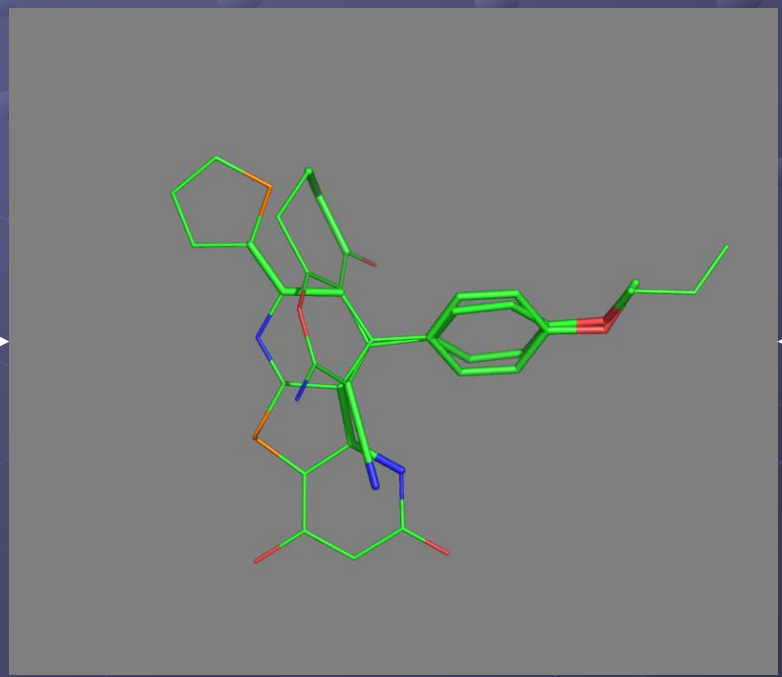
3D maximum overlapping structure example



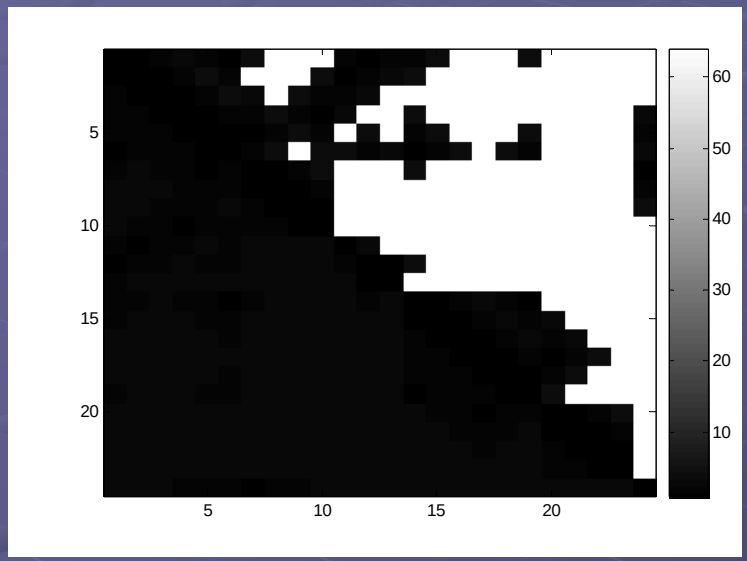
$$S_{12} = \frac{|G_{12}|}{|G_1| + |G_2| - |G_{12}|}$$

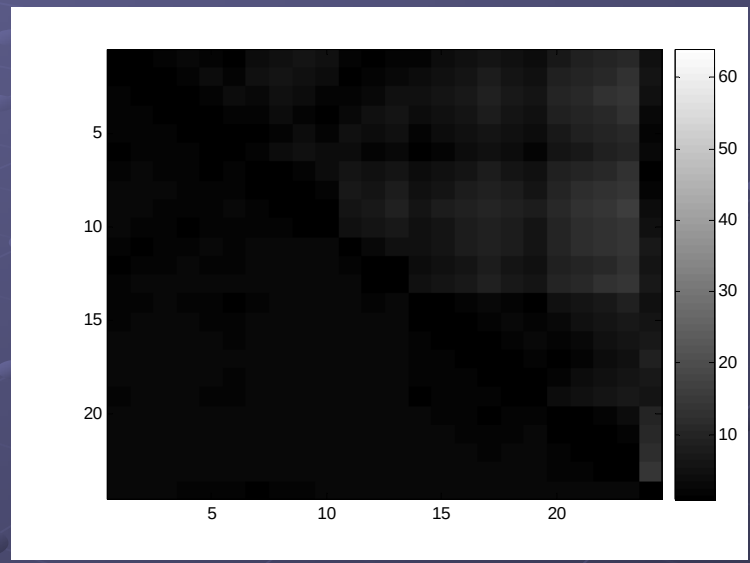
$$S_{12} = \mathbf{0.40541}$$





Distance bound smoothing example, upper (lower) diagonal values indicate upper (lower) bound distances, respectively

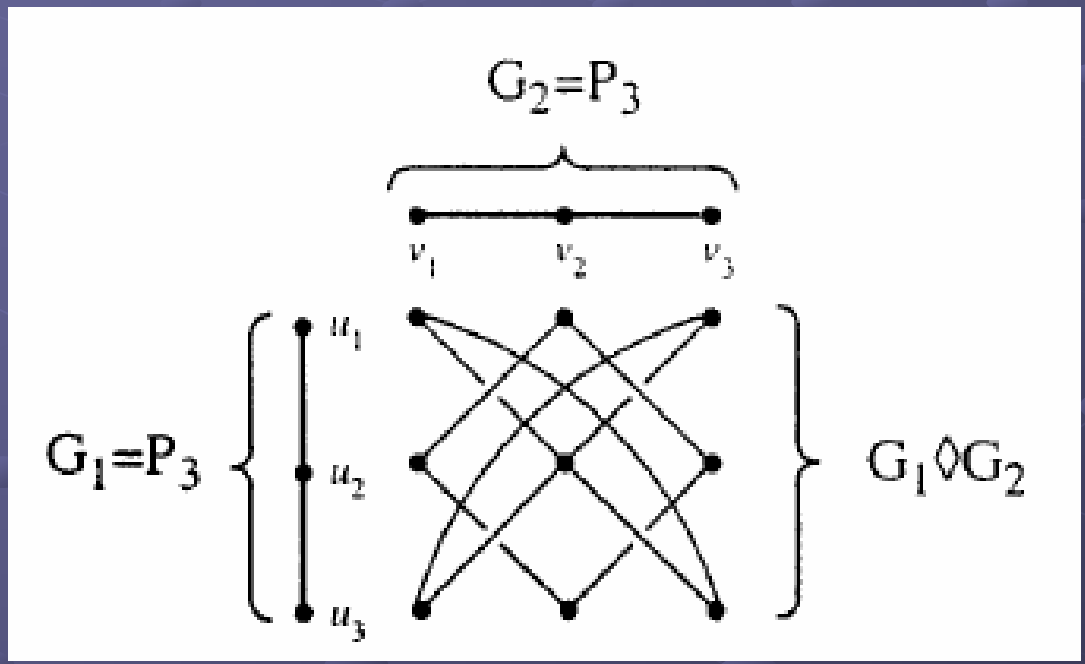




Before bound smoothing

After bound smoothing

The problem of finding Maximum Common Subgraph (MCS) can be reduced to one of determining the maximum Clique in a compatibility graph



$G_1 \Diamond G_2$

Denotes modular product graph or compatibility graph[1]

[1]Barrow, H.; Burstall, R. Subgraph Isomorphism, Matching Relational Structures and Maximal Cliques. Inf. Proc. Lett. 1976, 4, 83-84.

Modular product graph definitions

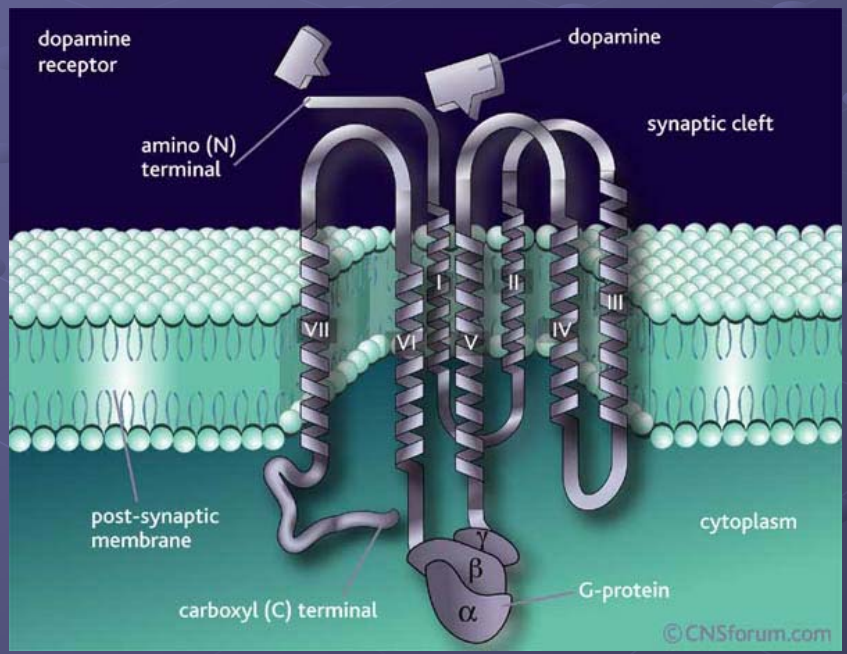
- The modular product graph of two graphs G_1 and G_2 is

  defined on the vertex set $V(G_1 \Diamond G_2)=V(G_1)xV(G_2)$
- A vertex (u_i, v_i) exists in $V(G_1 \Diamond G_2)$ if and only if $w(u_i)=w(v_i)$ where $w(u_i)$ and $w(v_i)$ denote vertex labels in chemical graph or atom types in a molecular structure
- Two vertices (u_i, v_i) and (u_j, v_j) in the modular product graph are adjacent whenever

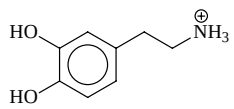
$$[L[u_i, u_j], \mathrm{U}[u_i, u_j]] = [\mathrm{L}[v_i, v_j], \mathrm{U}[v_i, v_j]]$$

Validation study

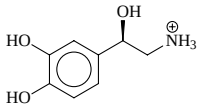
- *Data set of 17 dopamine receptors antagonists*
- *Dopamine receptors in the brain are important in modulating motor, endocrine, and emotional functions*



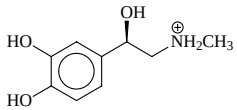
- *5 types of receptors*
- *The structures in the given dataset bind to D1/D5 receptor family*



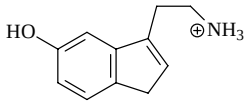
DOPAMINE



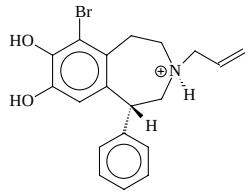
NORADRENALINE



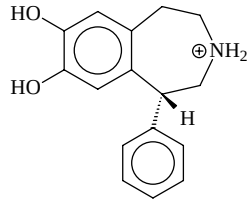
ADRENALINE



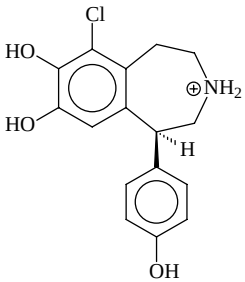
SEROTONIN



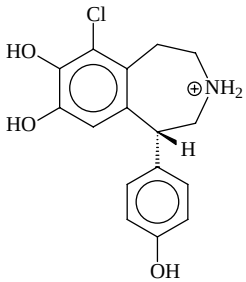
R(+)6-Br-APB



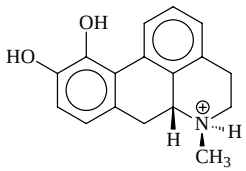
SKF38393



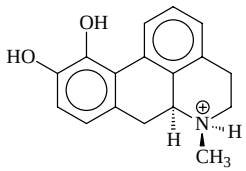
(S) SKF82526



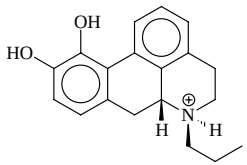
(R) SKF82526



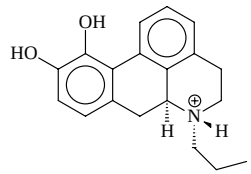
R-(-)-APOMORPHINE



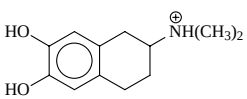
S-(+)-APOMORPHINE



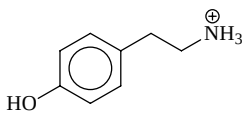
R-(-)-NPA



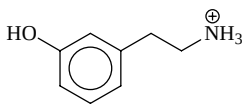
S-(+)-NPA



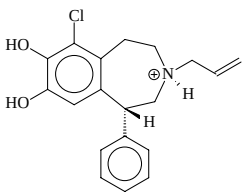
(+)6,7-ADTN



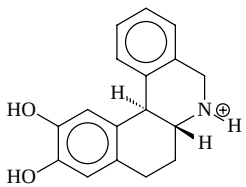
PARA-TYRAMINE



META-TYRAMINE



Cl-APB



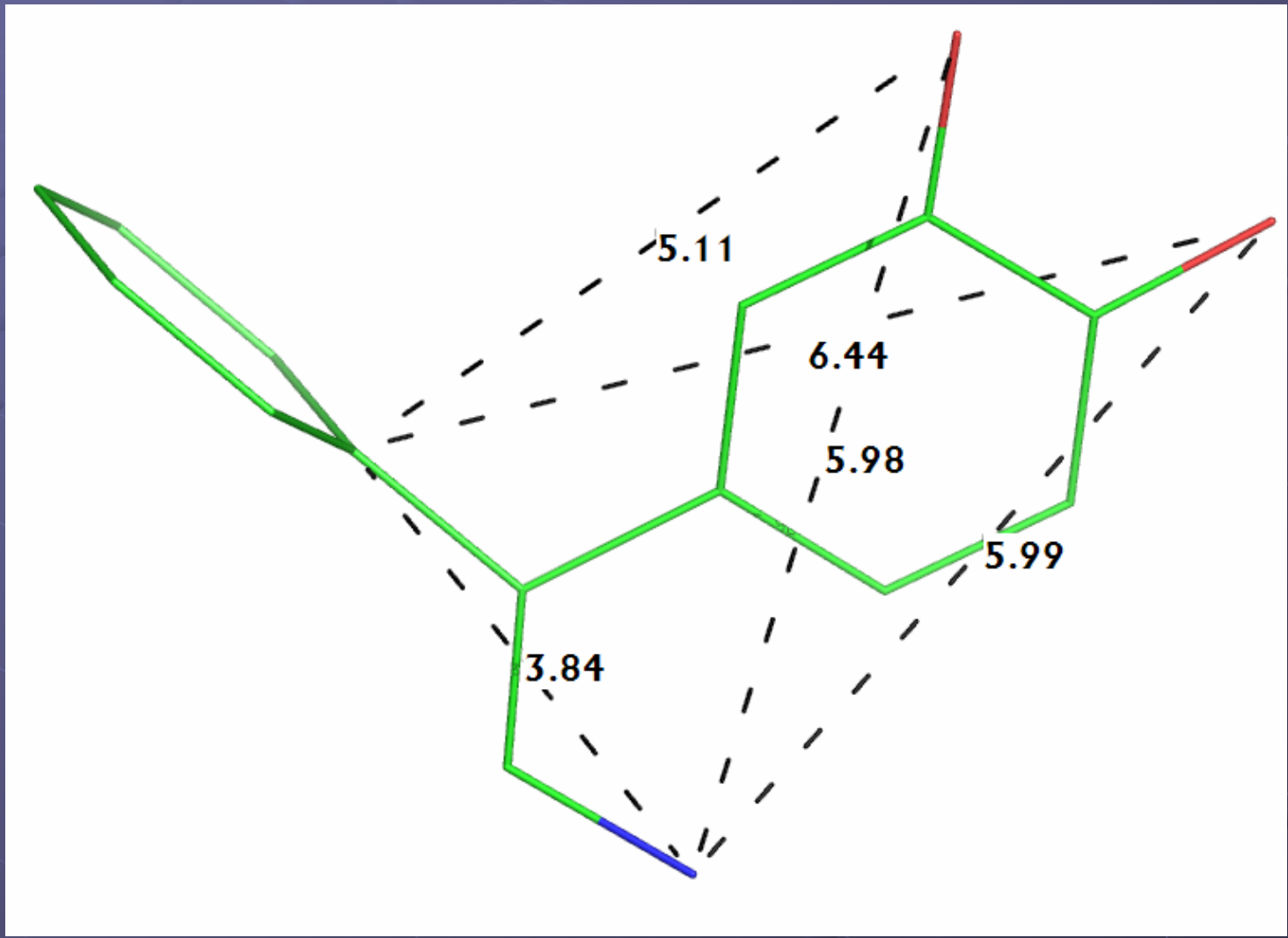
DIHYDREXIOINE

Mi-Youn Kim Brusniak; Pearlman R.; Kim A.; Richard E.; Comparative Molecular Field Analysis-Based Prediction of Drug Affinities at Recombinant D1A Dopamine Receptors, J. Med. Chem., 1996, **39**, 850-859.

Drug affinities and similarity calculations results

molecule	$\log(1/K_d)$	Similarity
(+)-6,7-ADTN	-3.66	0.7500
adrenaline	-4.74	0.4444
Cl-APB	-1.92	0.9907
DHX	-3.08	0.8102
dopamine	-3.39	0.6759
m-tyramine	-4.68	0.5833
noradrenaline	-4.69	0.6759
p-tyramine	-5.59	0.5880
(R)-apomorfine	-2.83	0.7685
(R)-(-)-NPA	-3.26	0.7546
(R)-(+)-6-Br-APB	-2.58	0.9861
(S)-(+)-apomorfine	-3.08	0.7639
S-(+)-NPA	-3.72	0.7639
(S)-SKF82526	-3.26	0.7778
serotonin	-3.99	0.6806
SKF38393	-2.18	0.9861
(R)-SKF-82526*	-1.45	1.0000

Proposed pharmacophore map for dopamine receptor antagonists based on similarity calculations results



Conclusions

- *New efficient method for calculation of intermolecular 3D similarity with applications in virtual screening and lead discovery*
- *Accounts for conformational flexibility*
- *More accurate comparing to other methods for virtual screening like bit strings based methods*
- *Similarity searching in chemical databases of non trivial size*