

Integrating Mathematics and Machine Learning for Drug Design

Guowei Wei

Mathematics

Biochemistry & Molecular biology

Michigan State University

<http://www.math.msu.edu/~wei>

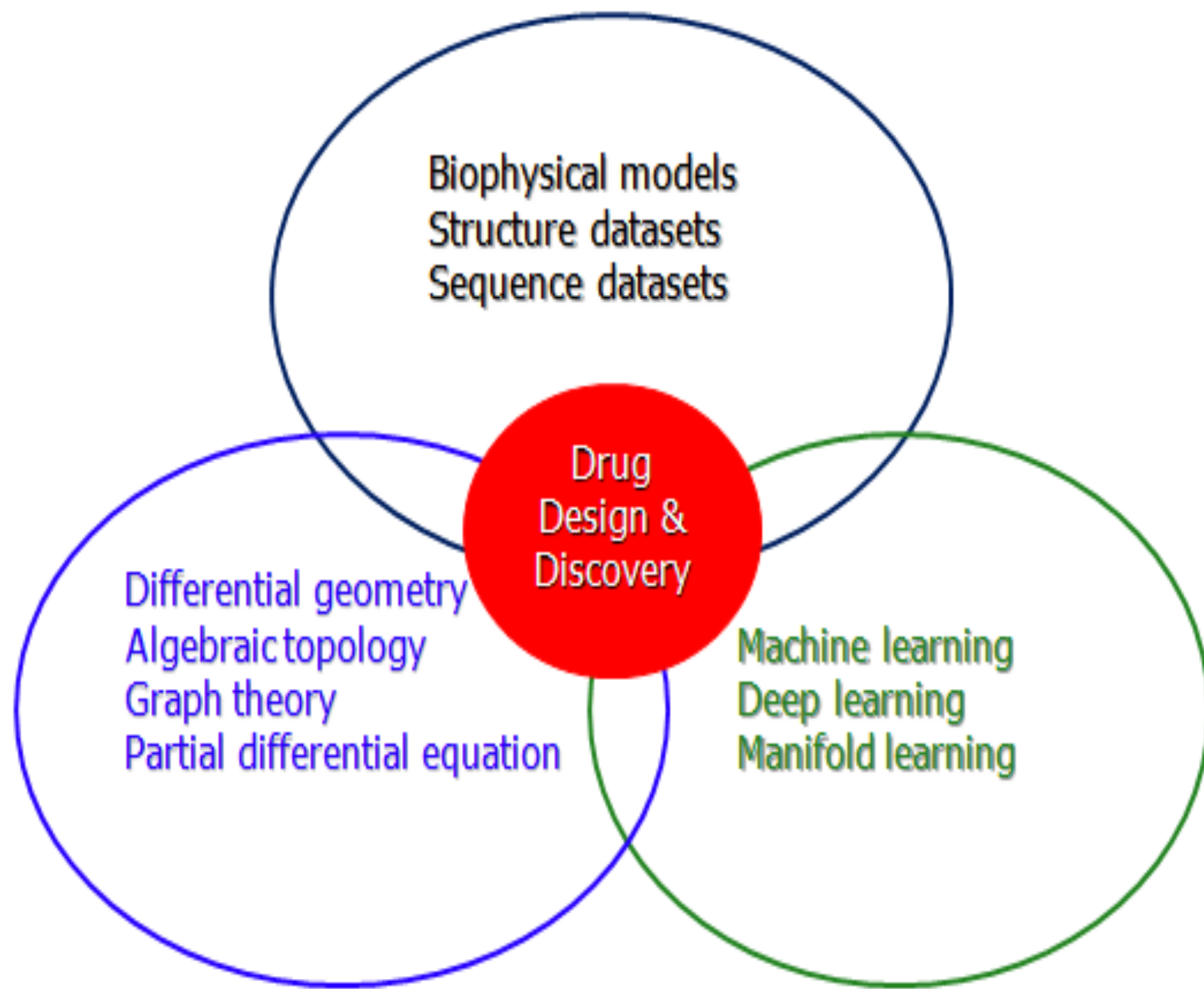
D3R 2018 Workshop

February 22-23, La Jolla

Grant support:

NSF, NIH, MSU and BMS



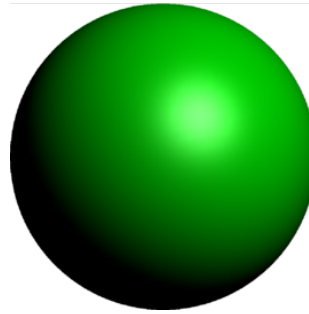


Classical topological objects

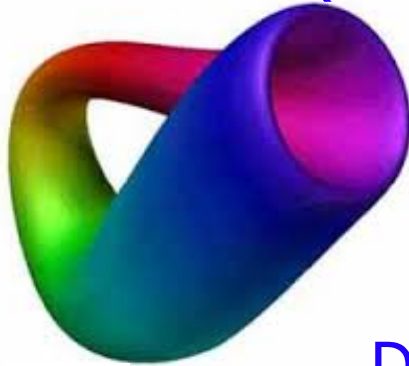
Möbius Strips (1858)



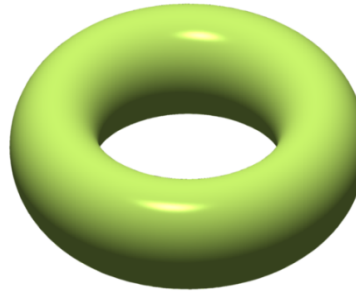
Sphere



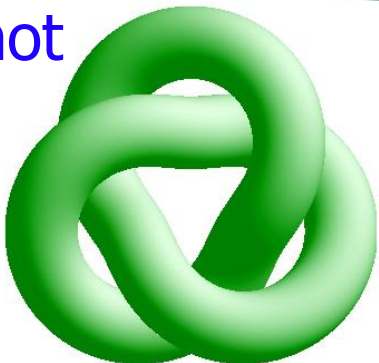
Klein Bottle (1882)



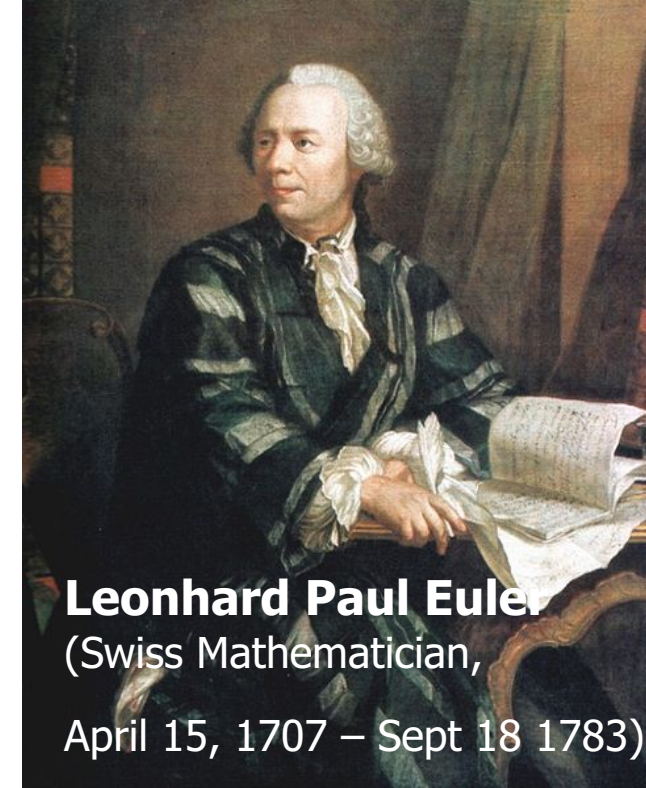
Torus



Trefoil
Knot



Double Torus

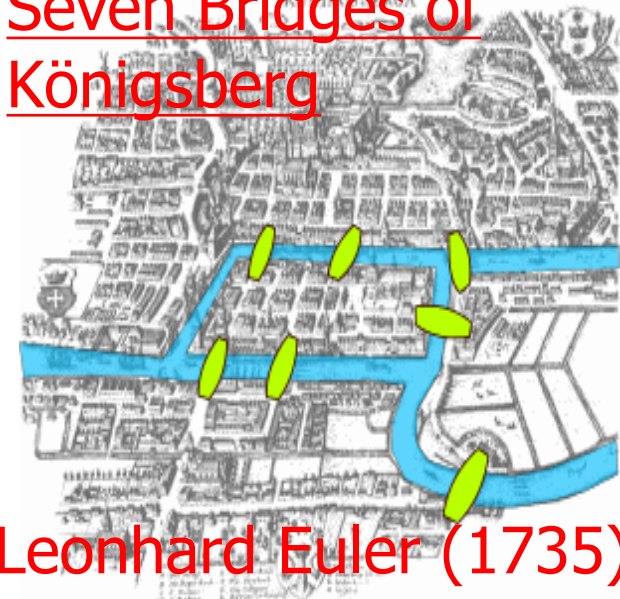


Leonhard Paul Euler

(Swiss Mathematician,

April 15, 1707 – Sept 18 1783)

Seven Bridges of
Königsberg



Leonhard Euler (1735)

Topological invariants: **Betti** numbers

β_0 is the number of connected components.

β_1 is the number of tunnels or circles.

β_2 is the number of cavities or voids.

Point

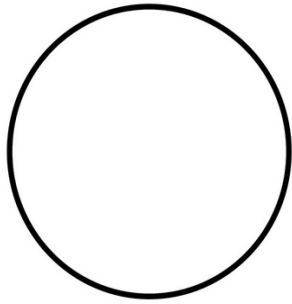


$$\beta_0 = 1$$

$$\beta_1 = 0$$

$$\beta_2 = 0$$

Circle

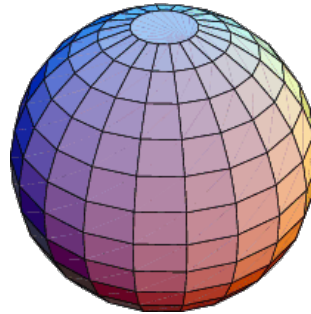


$$\beta_0 = 1$$

$$\beta_1 = 1$$

$$\beta_2 = 0$$

Sphere

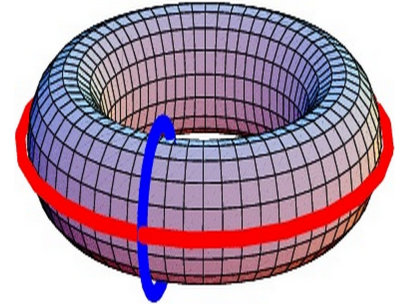


$$\beta_0 = 1$$

$$\beta_1 = 0$$

$$\beta_2 = 1$$

Torus



$$\beta_0 = 1$$

$$\beta_1 = 2$$

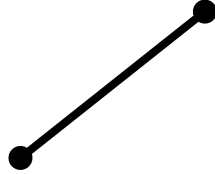
$$\beta_2 = 1$$

Vietoris-Rips complexes of planar point sets

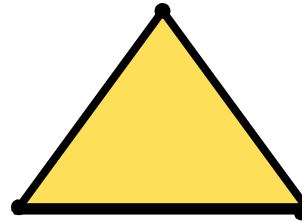
Simplexes:



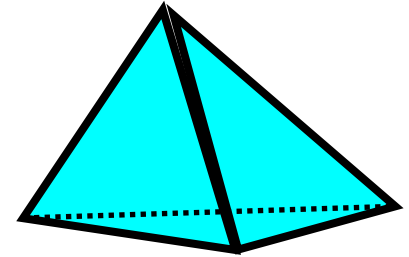
0-simplex



1-simplex

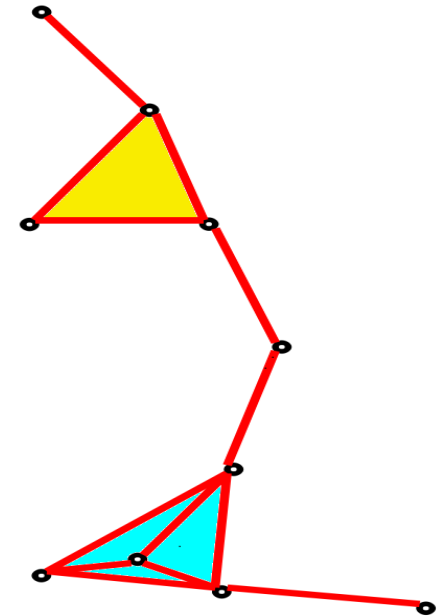
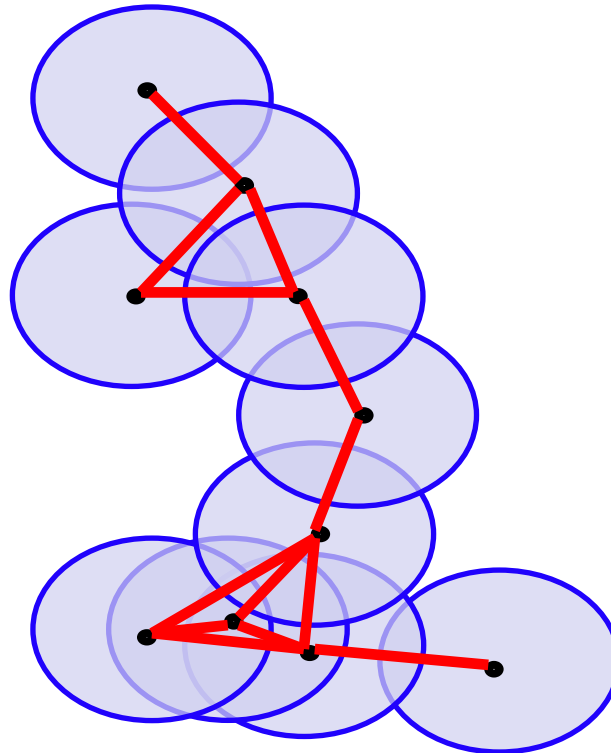
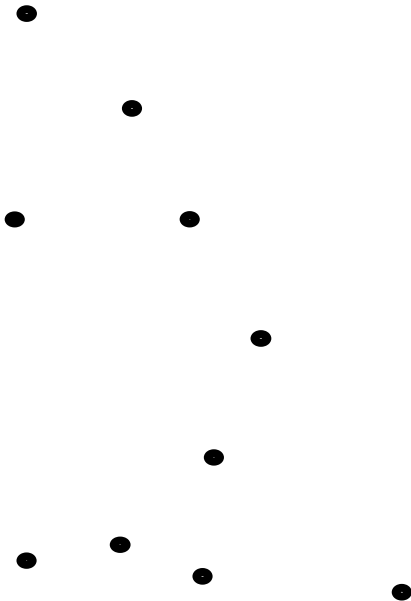


2-simplex



3-simplex

Simplicial complexes of ten points:



Topological modeling - Persistent homology

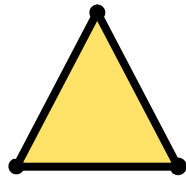
Simplexes:



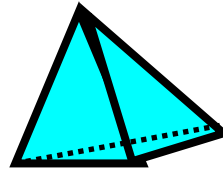
0-simplex



1-simplex



2-simplex



3-simplex

k-chain:

$$\sum_i c_i \sigma_i^k$$

Chain group: $C_k(K, \mathbb{Z}_2)$

Boundary operator:

$$\partial_k \sigma^k = \sum_{i=0}^k (-1)^i \{v_0, v_1, \dots, \widehat{v_i}, \dots, v_k\}$$

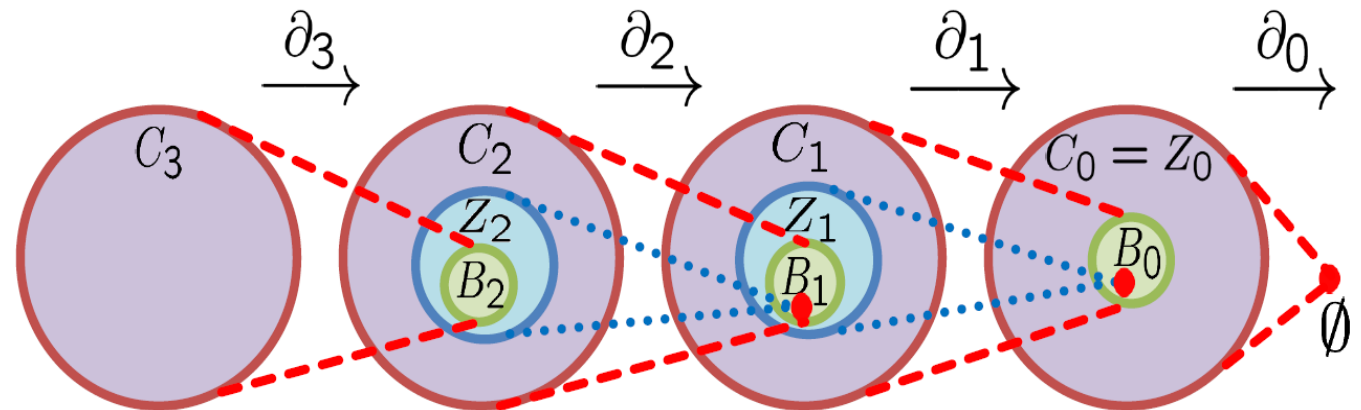
$$Z_k = \text{Ker } \partial_k$$

$$B_k = \text{Im } \partial_{k+1}$$

$$H_k = Z_k / B_k$$

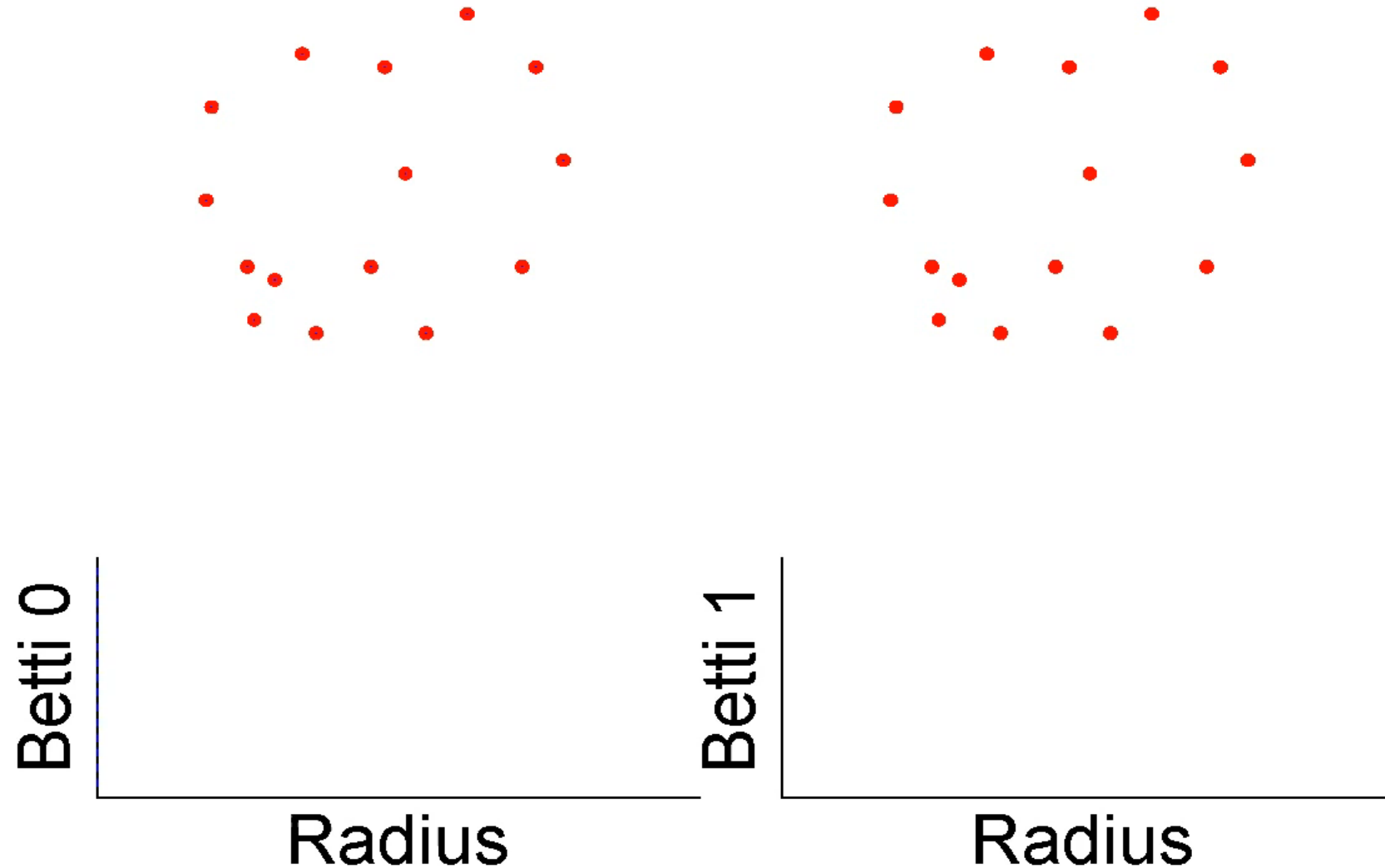
$$\beta_k = \text{Rank}(H_k)$$

Frosini and Nandi (1999),
Robins (1999),
Edelsbrunner, Letscher and Zomorodian
(2002), Edelsbrunner and Harer, (2007)
Kaczynski, Mischaikow and Mrozek (2004),
Zomorodian and Carlsson (2005),
Ghrist (2008),
.....

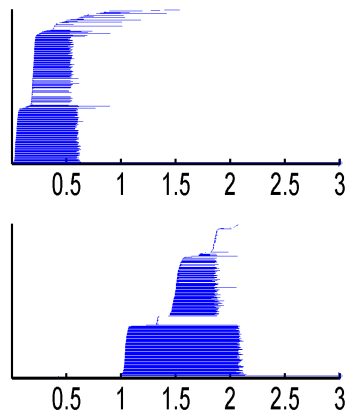
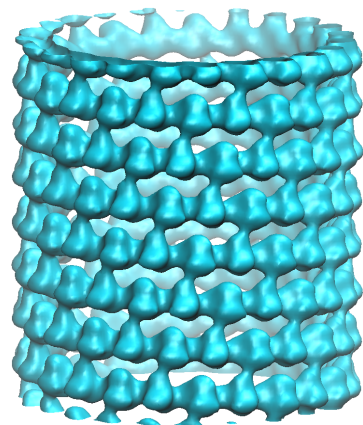
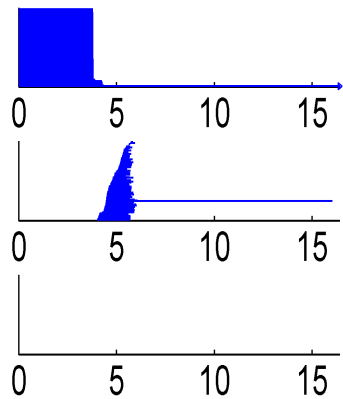
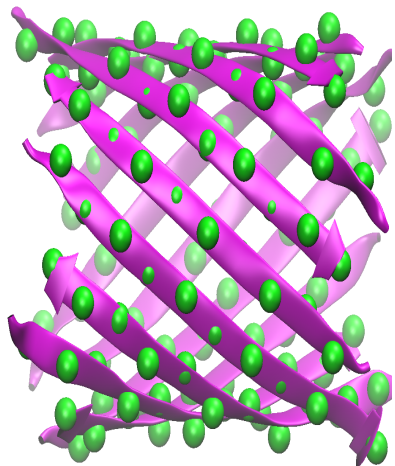
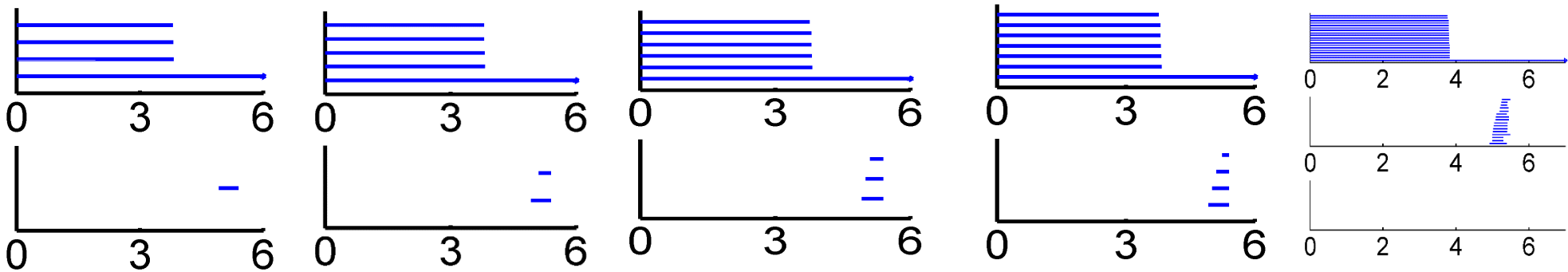
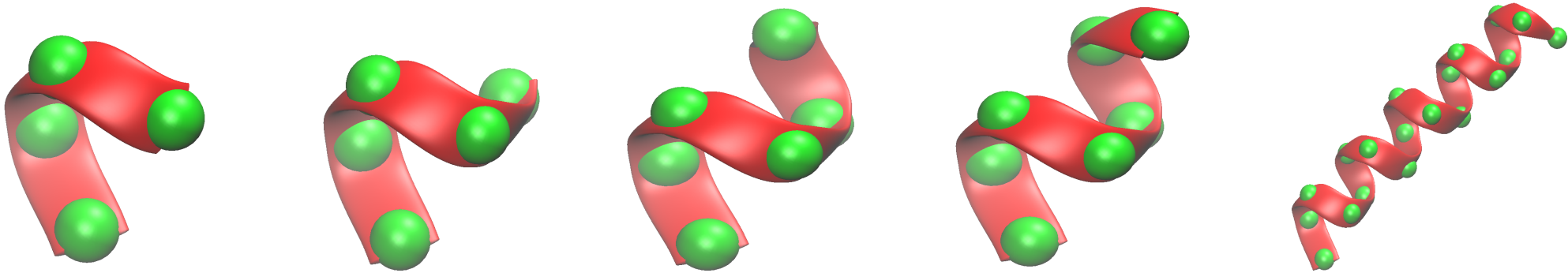


Filtration, Vietoris-Rips complexes, and persistent barcodes

(Xia, Wei, 2014)

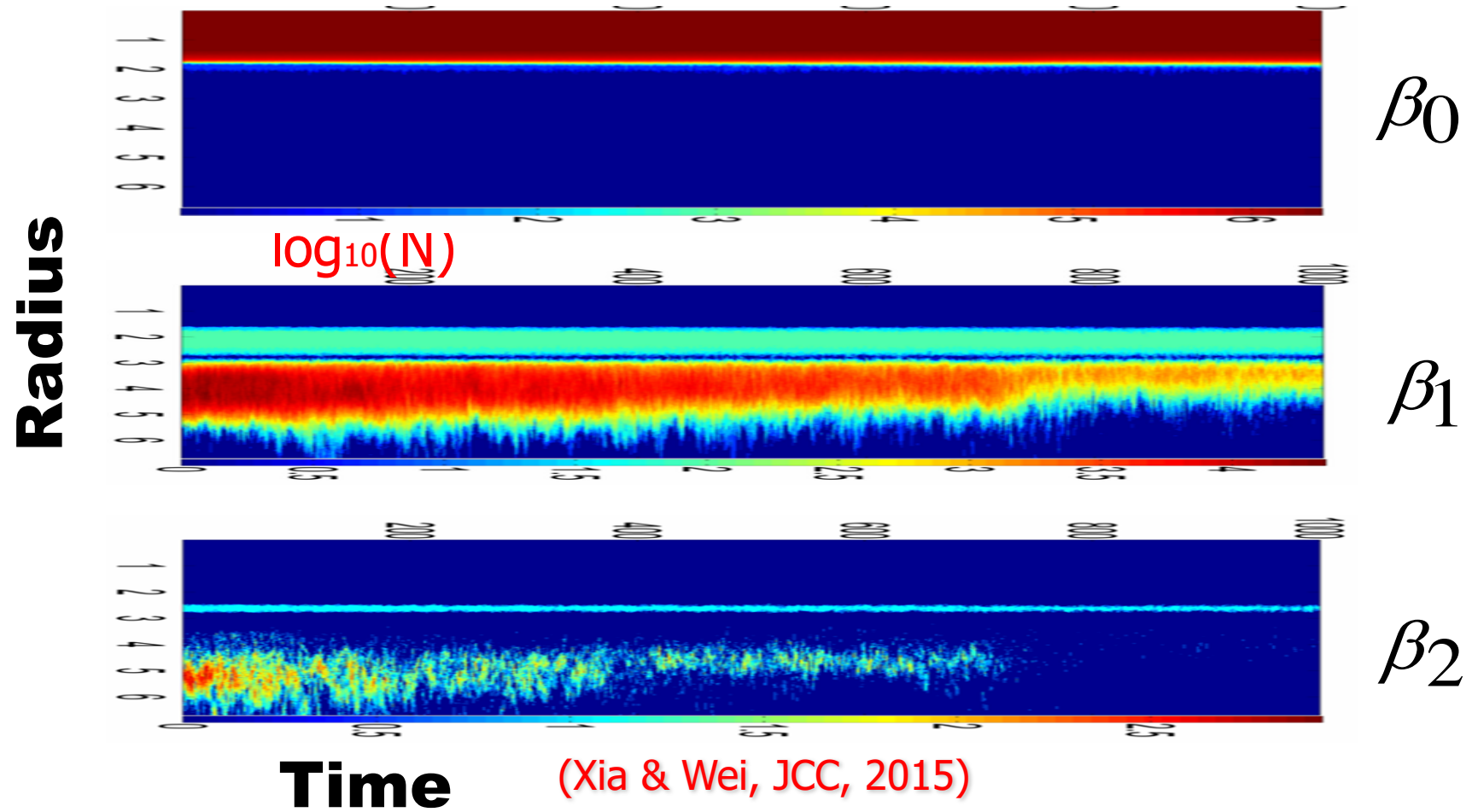
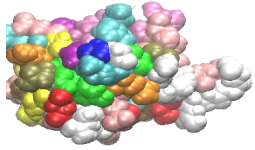


Topological fingerprints of an alpha helix



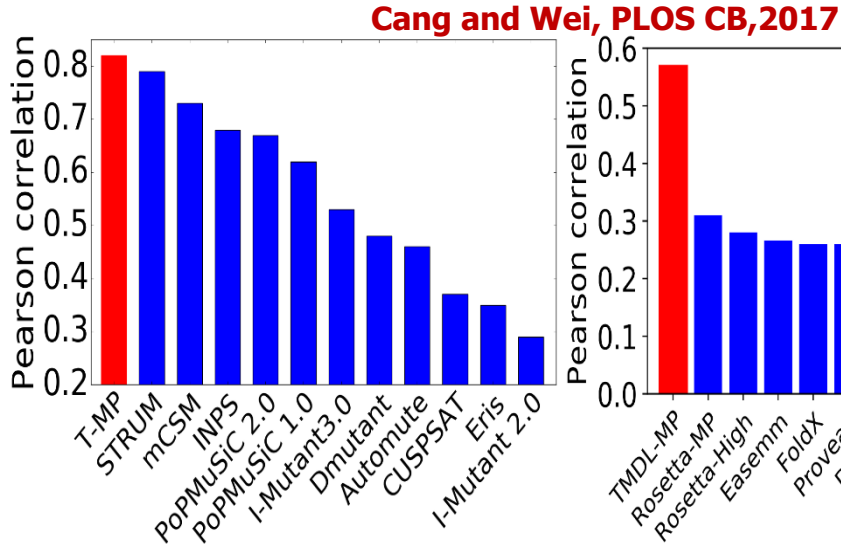
**(Xia & Wei,
IJNMBE,
2014, 2015)**

2D persistence in protein 1UBQ unfolding

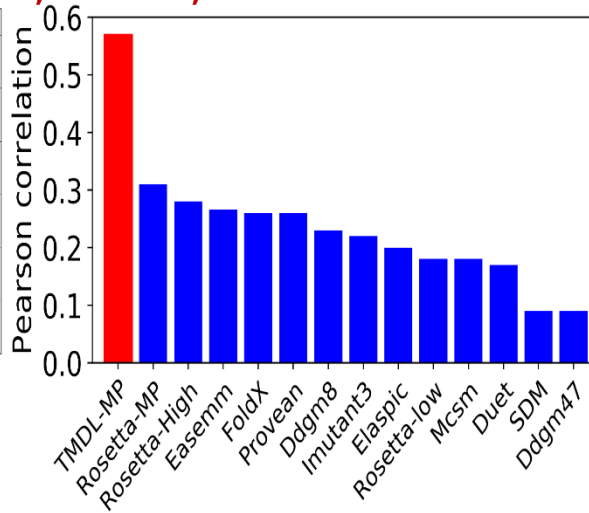


Topological learning based predictions

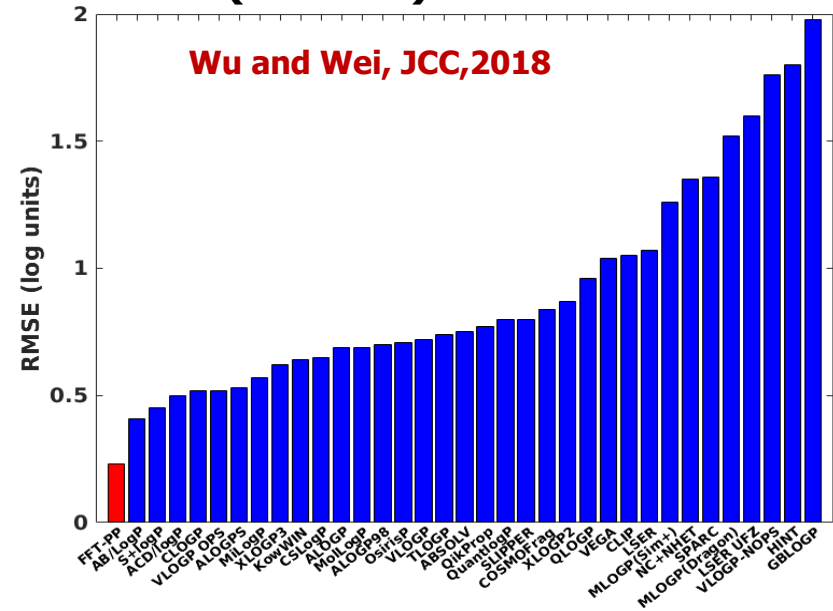
Prediction correlations for 2648 mutations on globular proteins



Prediction correlations for 223 mutations on membrane proteins

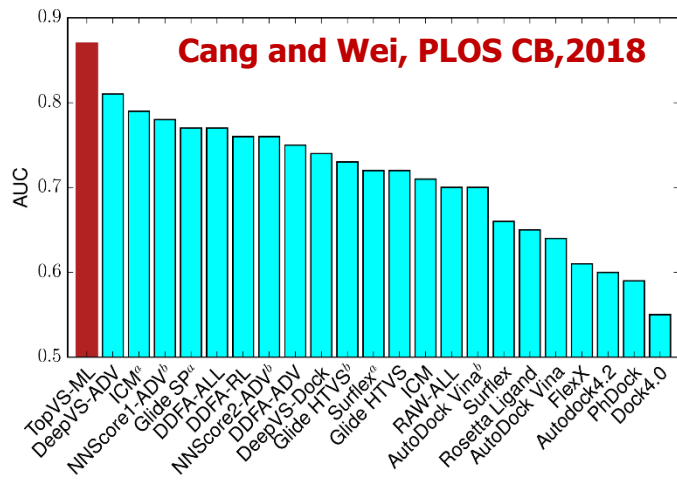


Prediction RMSD of LogP (Star set)

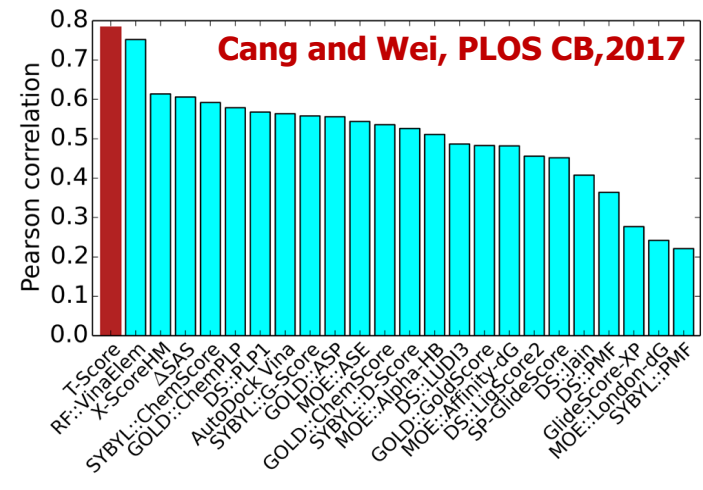


Classification of ligands & decoys

DUD database 128,374 protein-ligand/decoy pairs



Binding affinity prediction of PDBBind v2013 core set of 195 complexes



Drug Design & Discovery Resource (D3R) Grand Challenge 2

Given: Farnesoid X receptor (FXR) and 102 ligands

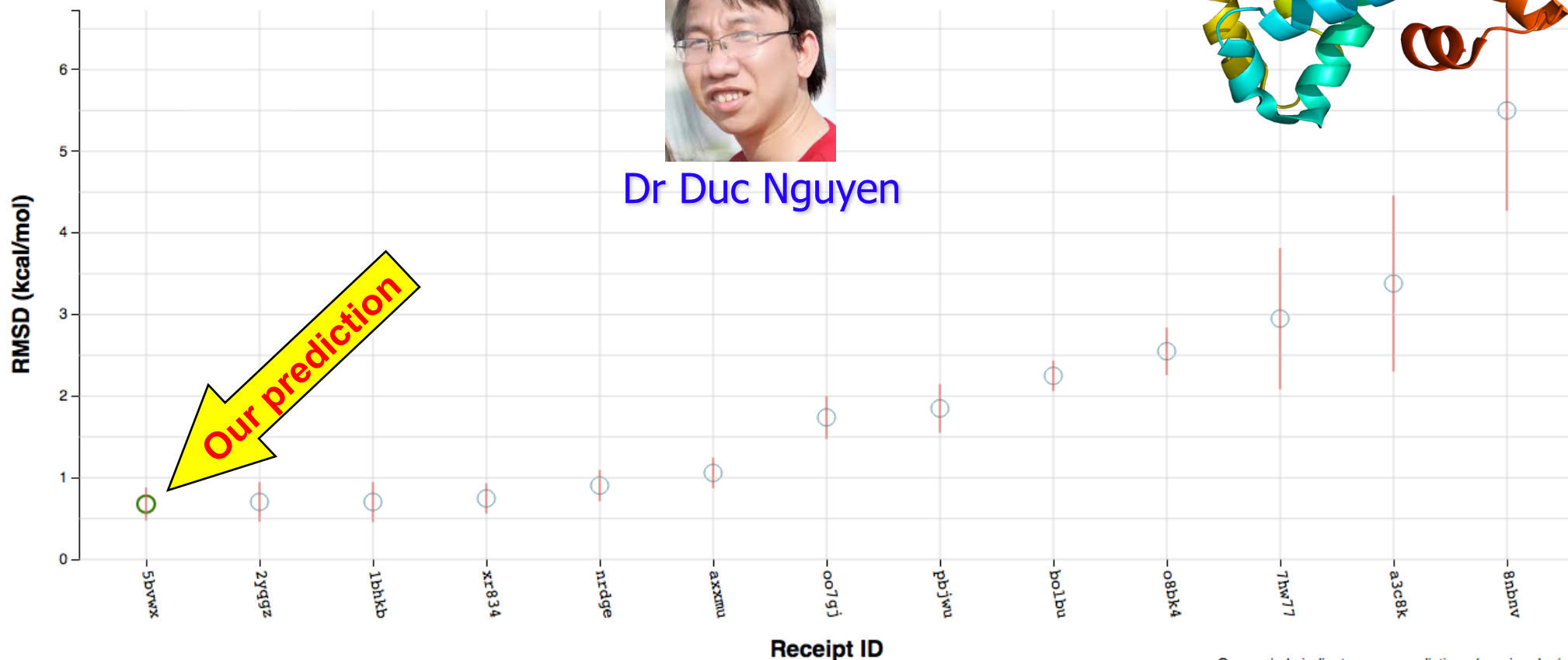
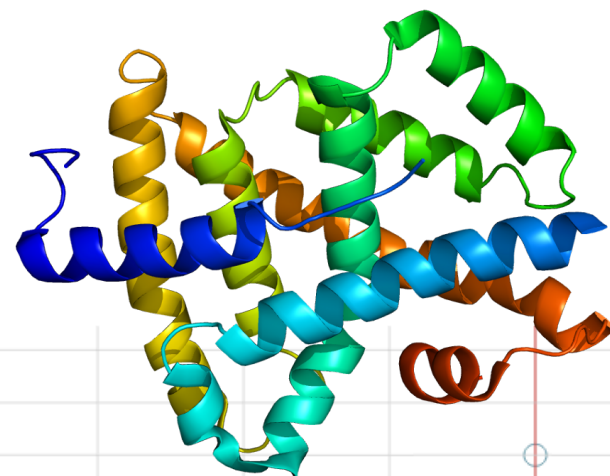
Tasks: Dock 102 ligands to FXR, and compute their poses, binding free energies and energy ranking

Grand Challenge 2

Free Energy Set 1 (Stage 1) - RMSD

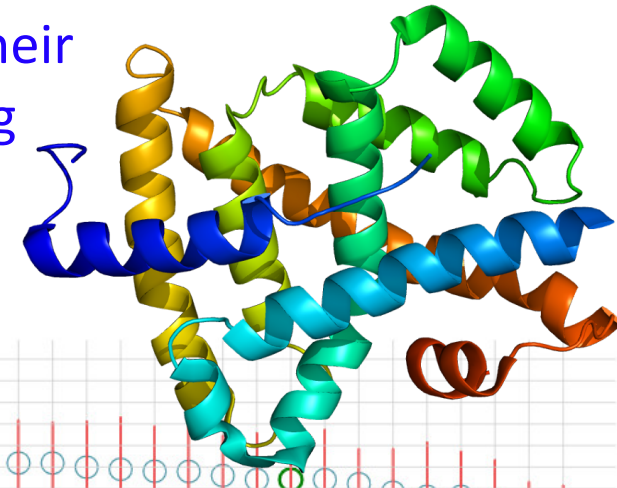


Dr Duc Nguyen

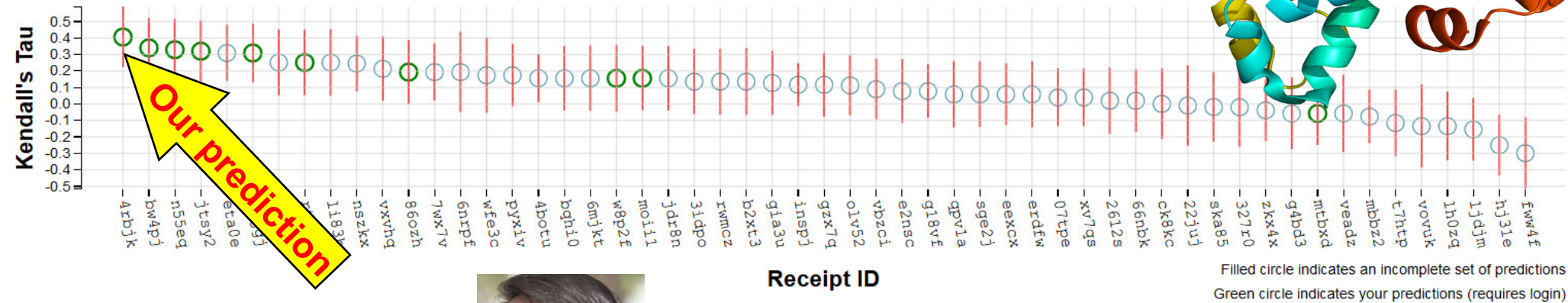


Green circle indicates your predictions (requires login)

Tasks: Dock 102 ligands to FXR, and compute their poses, binding free energies and energy ranking

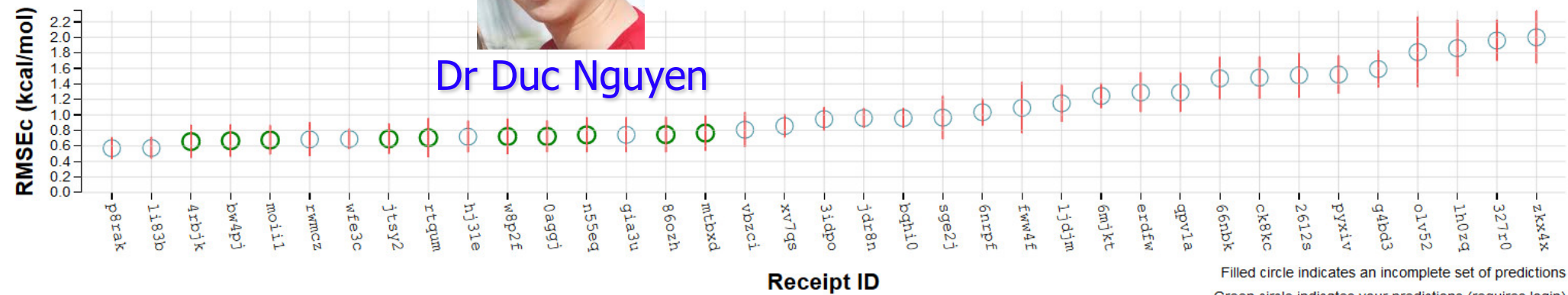


Free Energy Set 1 (Stage 2) - Kendall's Tau

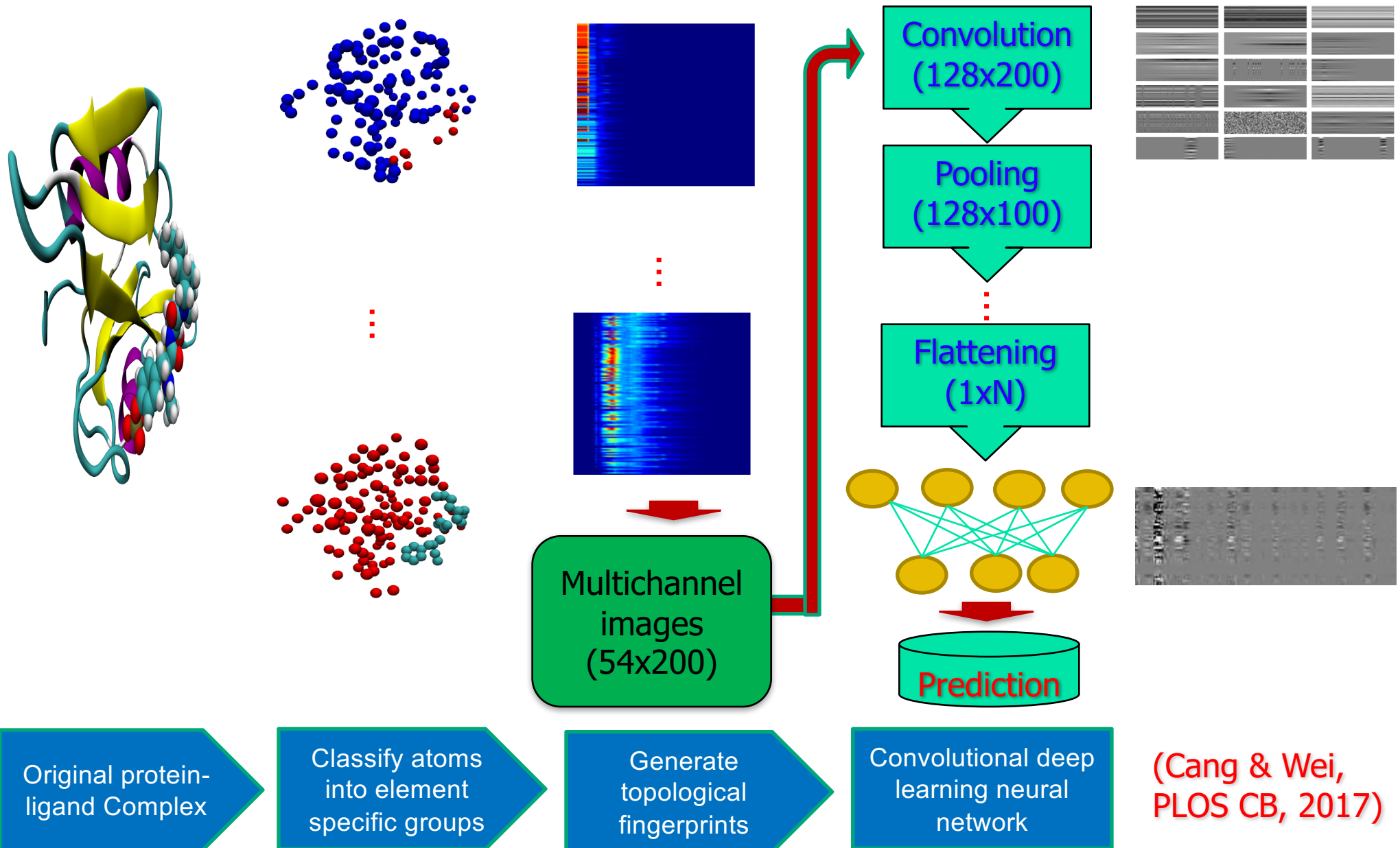


Dr Duc Nguyen

Free Energy Set 1 (Stage 2) - RMSEc



Topological convolutional deep Learning architecture



D3R Grand Challenge 3

Preliminary Evaluations, Subject to Revision and Refinement

Cathepsin Stage 1 Pose Predictions (partials)

Scoring (partials)

Free Energy Sets

Cathepsin Stage 1B Pose Prediction

Cathepsin Stage 2 Scoring (partials)

Free Energy Sets



Zixuan Cang



Dr Duc Nguyen

VEGFR2 Scoring (partials) **JAK SC2** Scoring (partials) **p38-α** Scoring (partials)

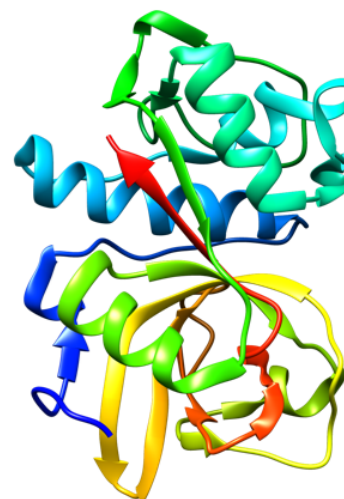
JAK SC3 Free Energy Sets

TIE2 Scoring (partials)

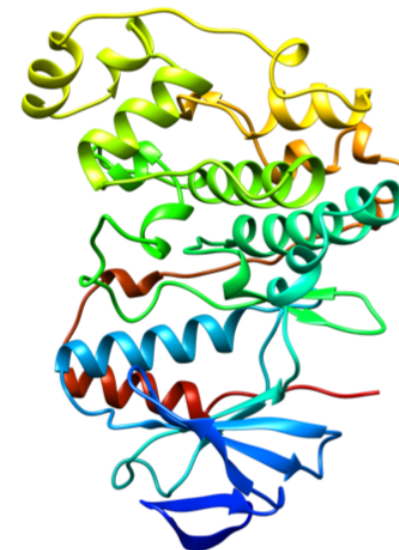
Free Energy Set 1

Free Energy Set 2

ABL1 Scoring



Cathepsin S



Kinase: p38-α