Lecture 7: Topology Based Modeling and Analysis

Guowei Wei

Mathematics

Michigan State University

https://users.math.msu.edu/users/wei/

NSF-CBMS Conference on Mathematical Molecular Bioscience and Biophysics

University of Alabama

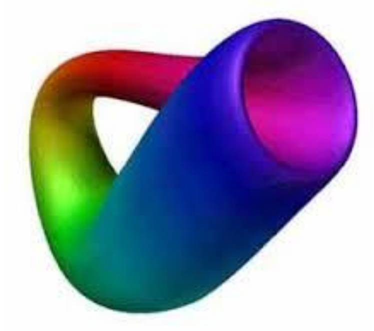
Tuscaloosa, May, 13-17, 2019

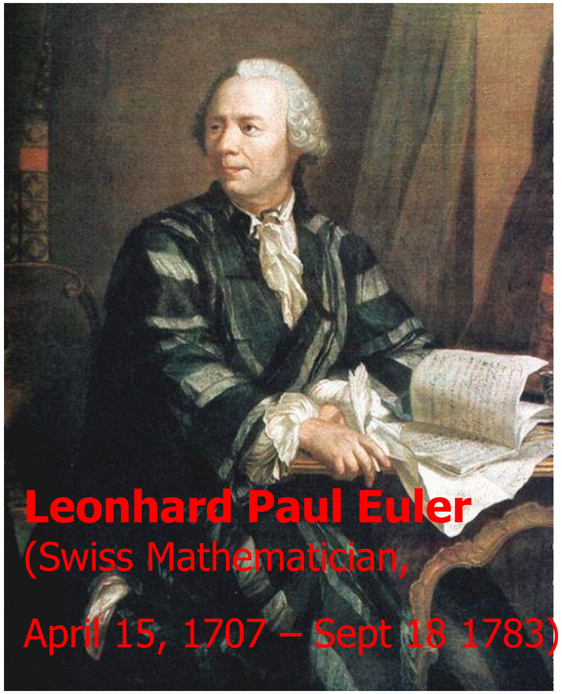
Grant support: NSF, NIH, MSU, BMS, and Pfizer

Classical Topology

Möbius Strips (1858)

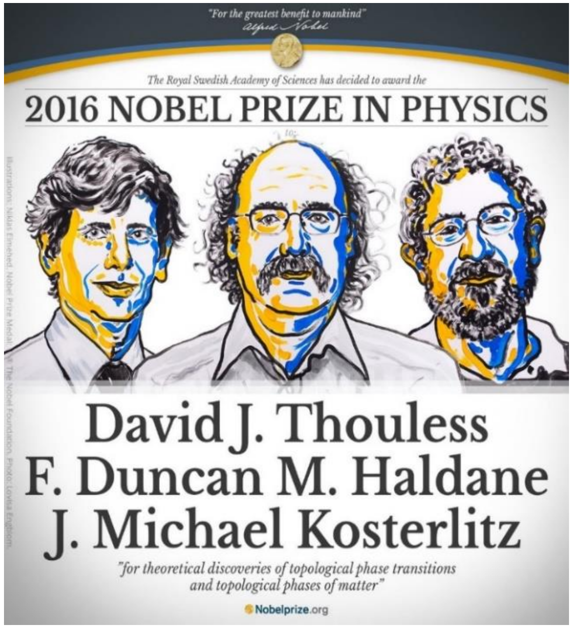
Klein Bottle (1882)



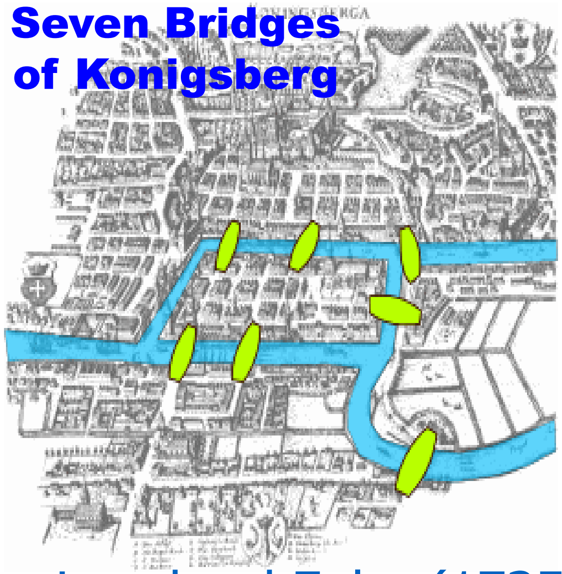


Torus

Double Torus

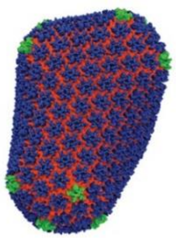


Augustin-Louis Cauchy,
Ludwig Schläfli,
Johann Benedict Listing,
Bernhard Riemann, and
Enrico Betti

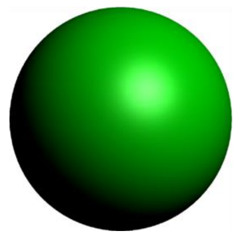


Leonhard Euler (1735)

Topological simplification

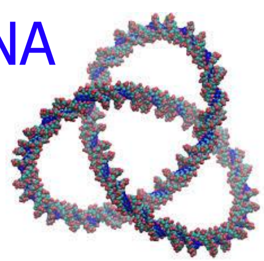


HIV 4.2 million atoms

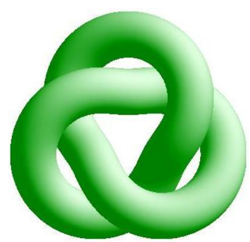


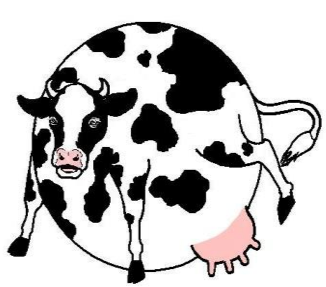
Mug ⇔ **Doughnut**

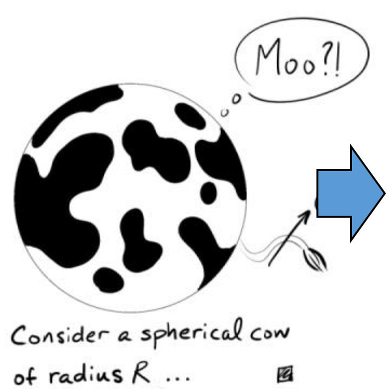
DNA

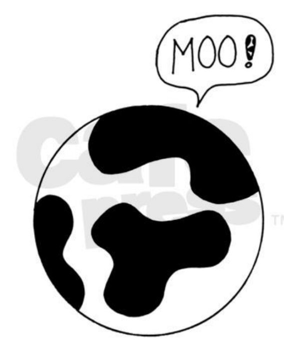


Trefoil Knot

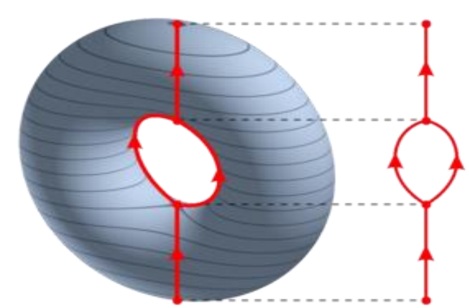








Morse theory



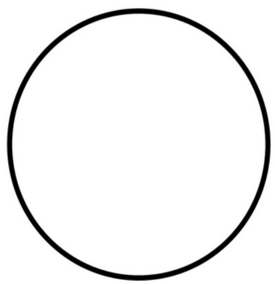
Topological invariants: Betti numbers

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

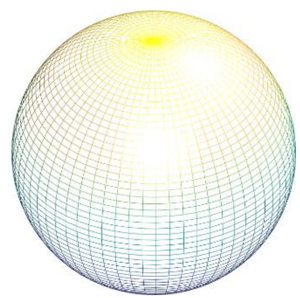
$\beta_0 = 1$
$\beta_1 = 0$
$\beta_2 = 0$

Circle



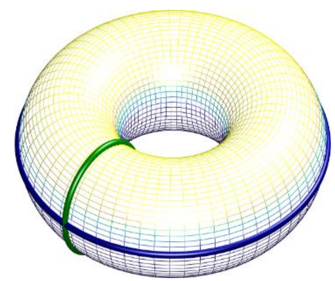
$\beta_0 = 1$
$\beta_1 = 1$
$\beta_2 = 0$

Sphere



$\beta_0 = 1$
$\beta_1 = 0$
$\beta_2 = 1$

Torus



$\beta_0 = 1$
$\beta_1 = 2$
$\beta_2 = 1$

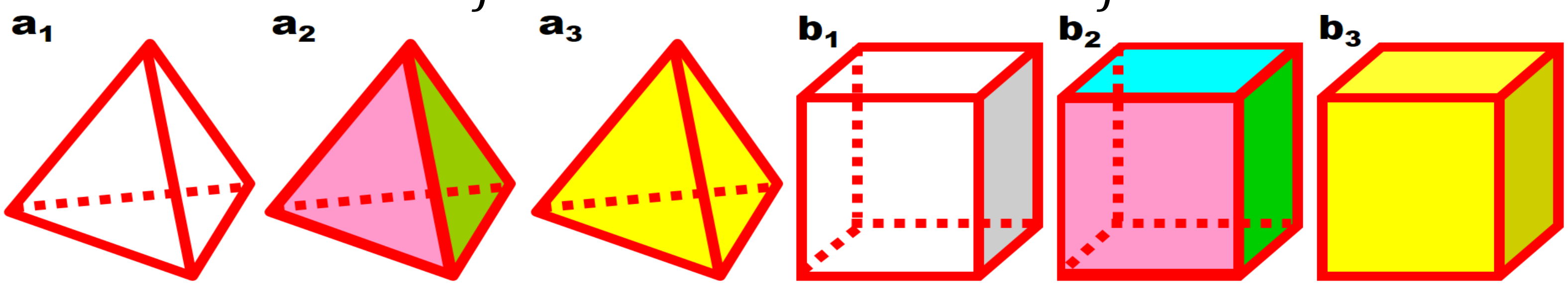
2-holed torus

$\beta_0 = 1$
$\beta_1 = 4$
$\beta_2 = 1$

Topological invariants: Euler characteristic

Euler characteristic (χ) realtes to Betti numbers $\left(\beta_j\right)$

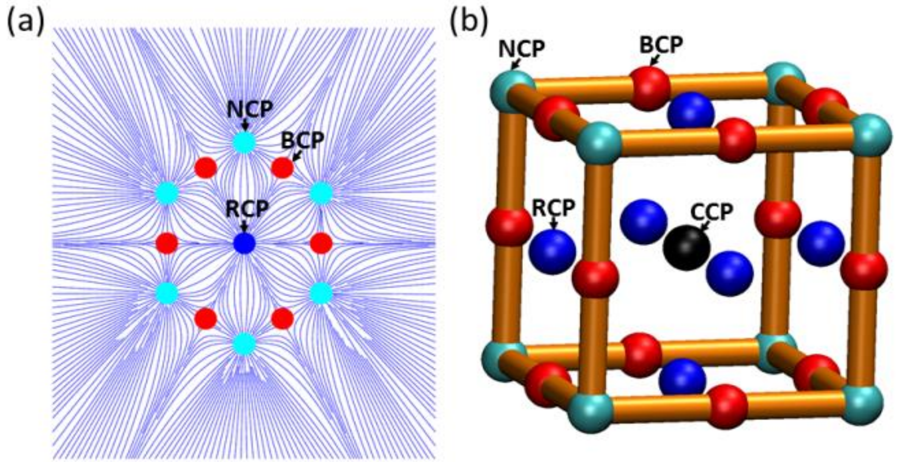
$$\chi(K) = \sum_{j}^{k} (-1)^j \operatorname{rank} C_j(K) = \sum_{j}^{k} (-1)^j \beta_j(K)$$

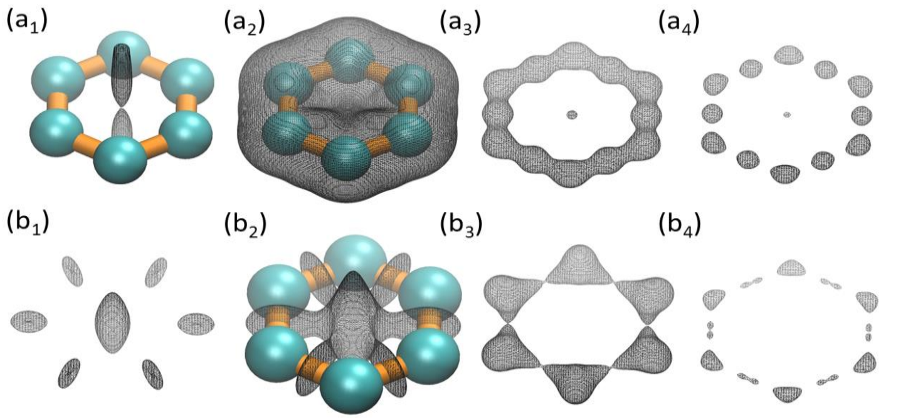


Simplex	V	E	F	C	β_0	β_1	β_2	χ
a_1	4	6	0	0	1	3	0	-2
a_2	4	6	4	0	1	0	1	2
a_3	4	6	4	1	1	0	0	1
b_1	8	12	0	0	1	5	0	-4
b_2	8	12	6	0	1	0	1	2
b_3	8	12	6	1	1	0	0	1

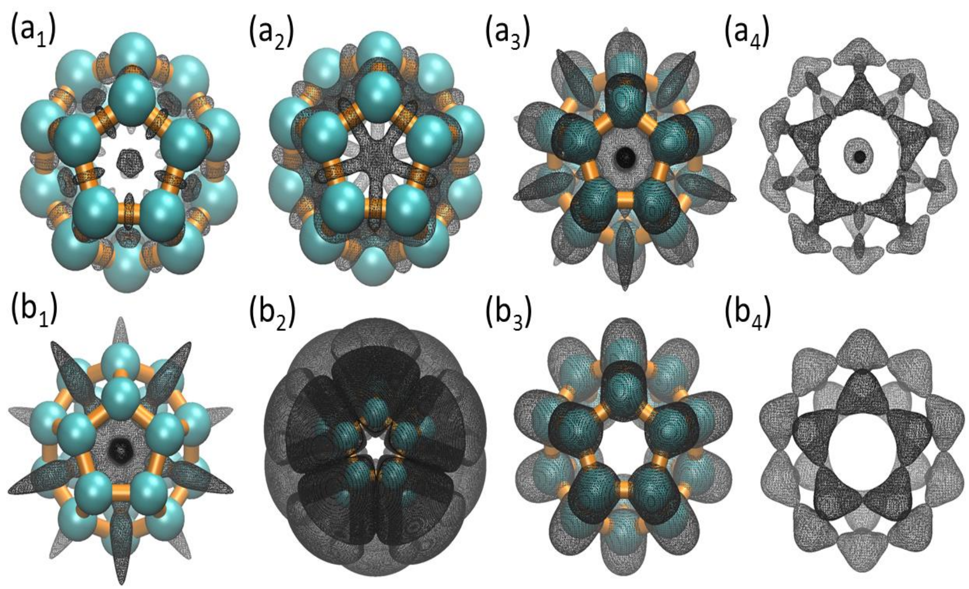
Topological modeling

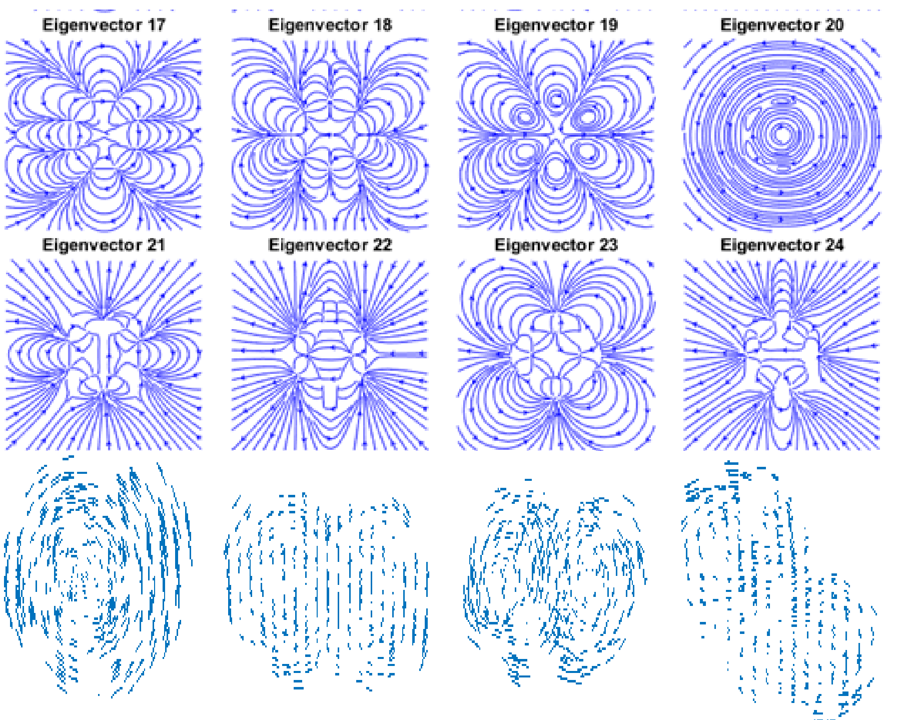
Poincare-Hopf index, Morse theory, and Conley index





(Xia, Wei, arXiv, 2016)





Opportunities, challenges, and promises

Opportunities from topological methods:

❖New approach for big data characterization and classification.

❖Dramatic reduction of dimensionality and data size.

❖Applicable to a variety of fields.

Challenges with topological methods:

➢Geometric methods are often inundated with too much structural detail.

➢Topological tools incur too much reduction of original geometric information.

➢Topology is hardly used for quantitative prediction.

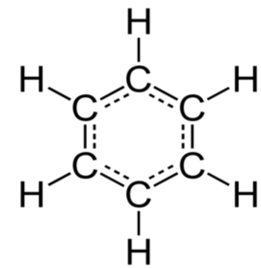
Promises from persistent homology:

✓Embeds geometric information in topological invariants.

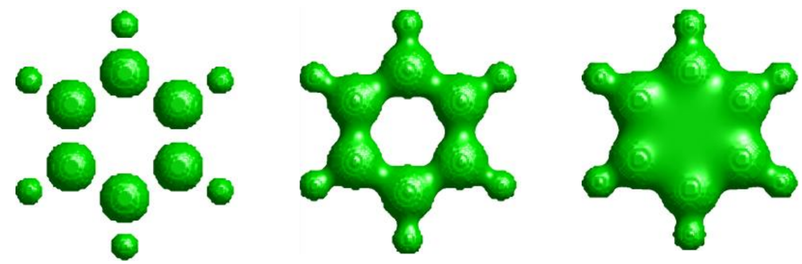
✓Bridges the gap between geometry and topology.

Persistent homology answers following questions

What is the topology of a benzene (C_6H_6)?

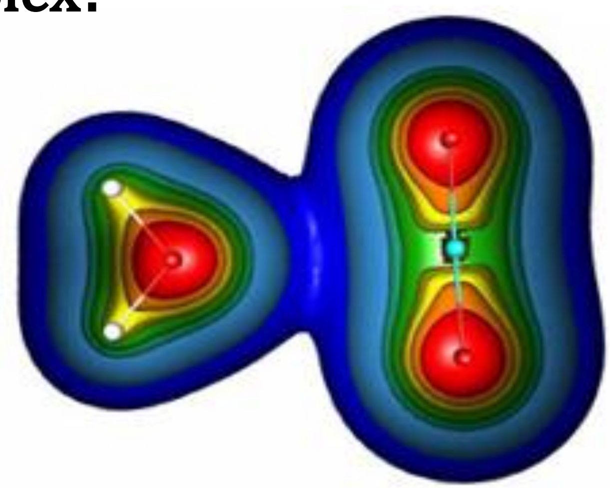


Level sets generated by Laplace-Beltrami flows:



What is the topology of a $H_2O - CO_2$ complex?

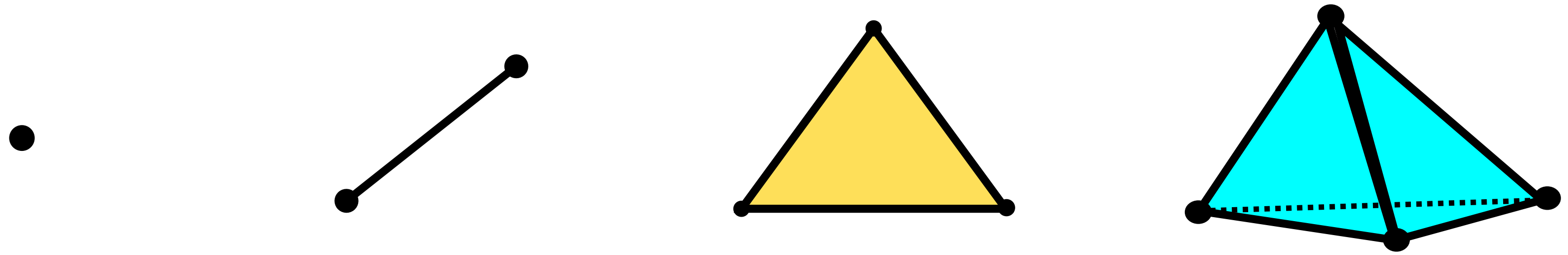
Electron density level sets computed by using quantum mechanics:



H_2O CO_2

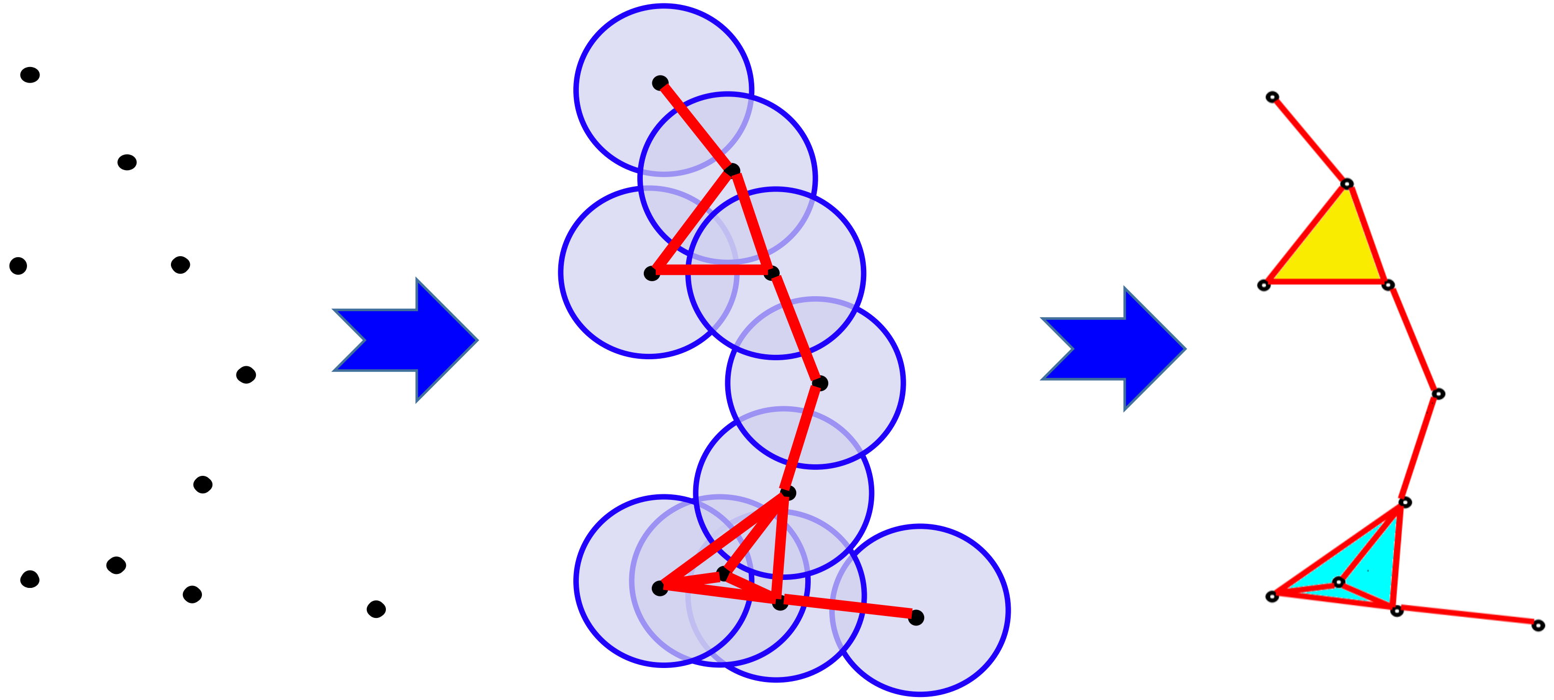
Vietoris-Rips complexes of planar point sets

Simplexes:



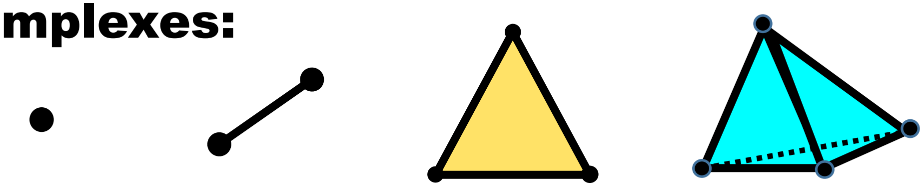
0-simplex *1*-simplex *2*-simplex *3*-simplex

Simplicial complexes of ten points:



Persistent homology

Simplexes:



***0*-simplex** ***1*-simplex** ***2*-simplex** ***3*-simplex**

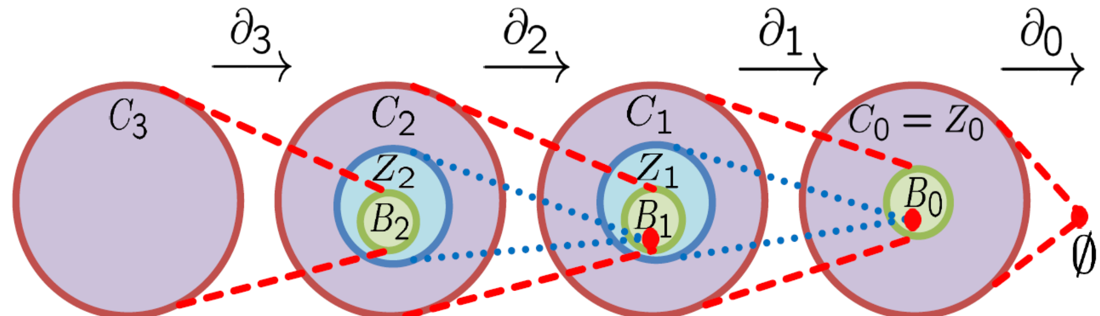
Frosini and Nandi (1999), Robins (1999), Edelsbrunner, Letscher and Zomorodian (2002), Zomorodian and Caelsson (2005), Edelsbrunner and Harer, (2007) Kaczynski, Mischaikow and Mrozek (2004), Ghrist (2008), ……

k-chain: $K = \left\{\sum_j c_j \sigma_j^k\right\}$

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

Boundary operator:

$$\partial_k \sigma^k = \sum_{j=0}^{k} (-1)^j \{v_0, v_1, \ldots, \widehat{v_j}, \ldots, v_k\}$$

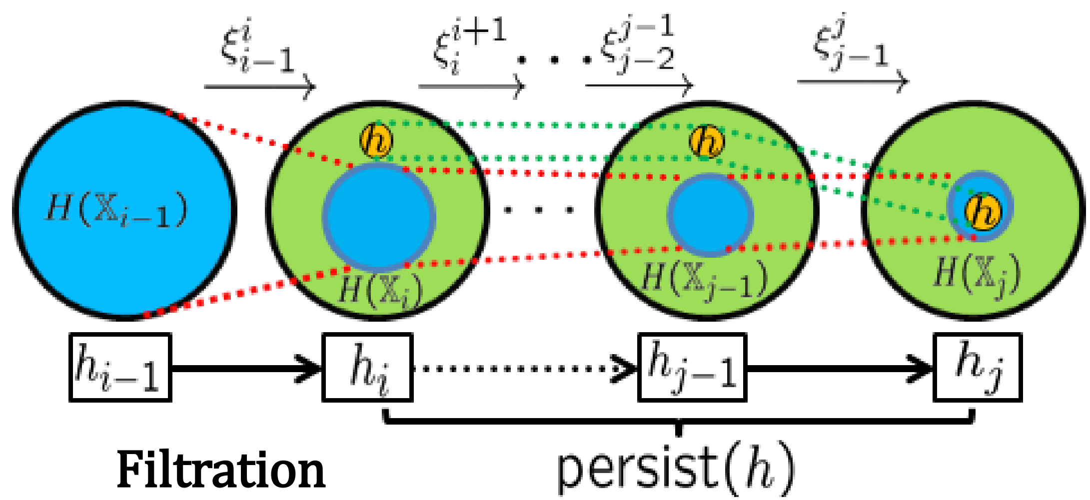


Cycle group: $Z_k = \mathrm{Ker}\ \partial_k$

Boundary group: $B_k = \mathrm{Im}\ \partial_{k+1}$

Homology group: $H_k = \dfrac{Z_k}{B_k}$

Betti number: $\beta_k = \mathrm{Rank}(H_k)$



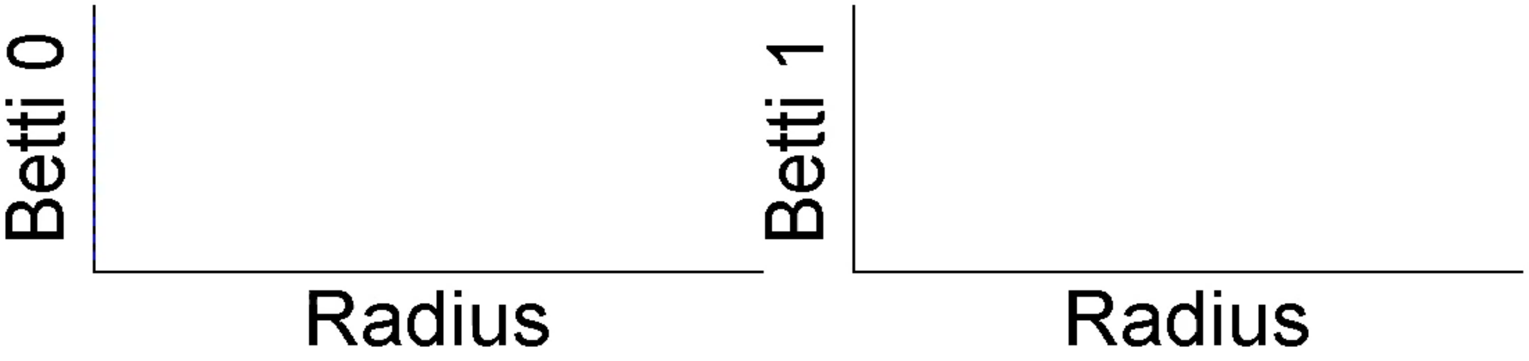
Filtration

Xia, Wei, IJNMBE, 2014;
Xia, Feng, Tong, Wei, JCC, 2015

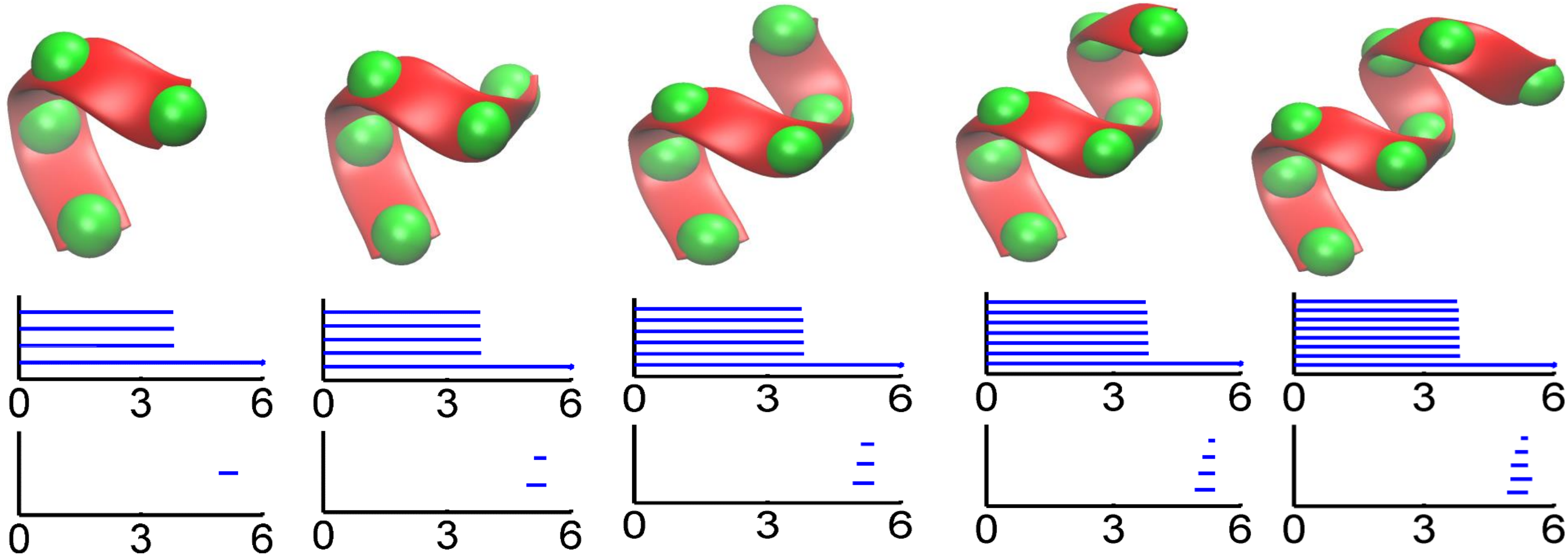
Algebraic Topology

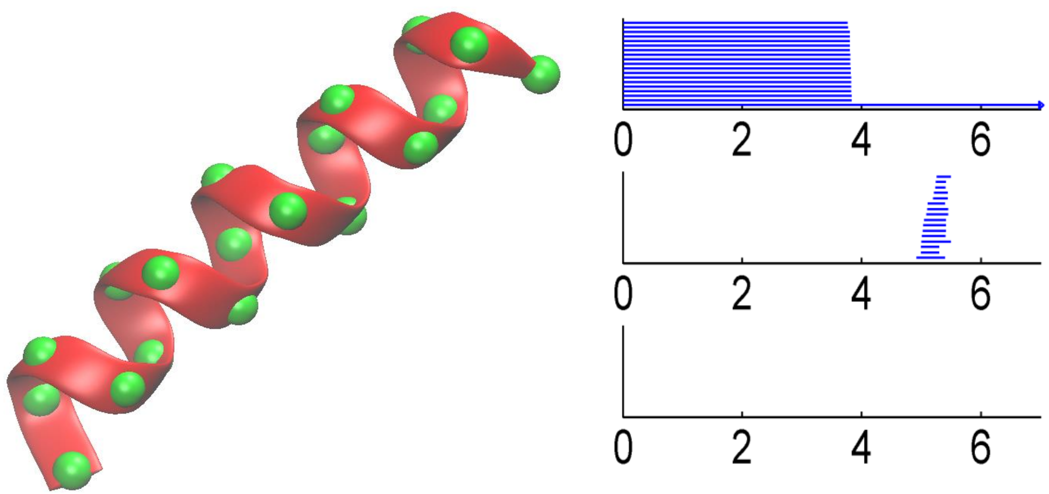
Vietoris-Rips complexes, persistent homology and topological fingerprint

(Xia, Wei, 2014)



Topological fingerprints of an alpha helix



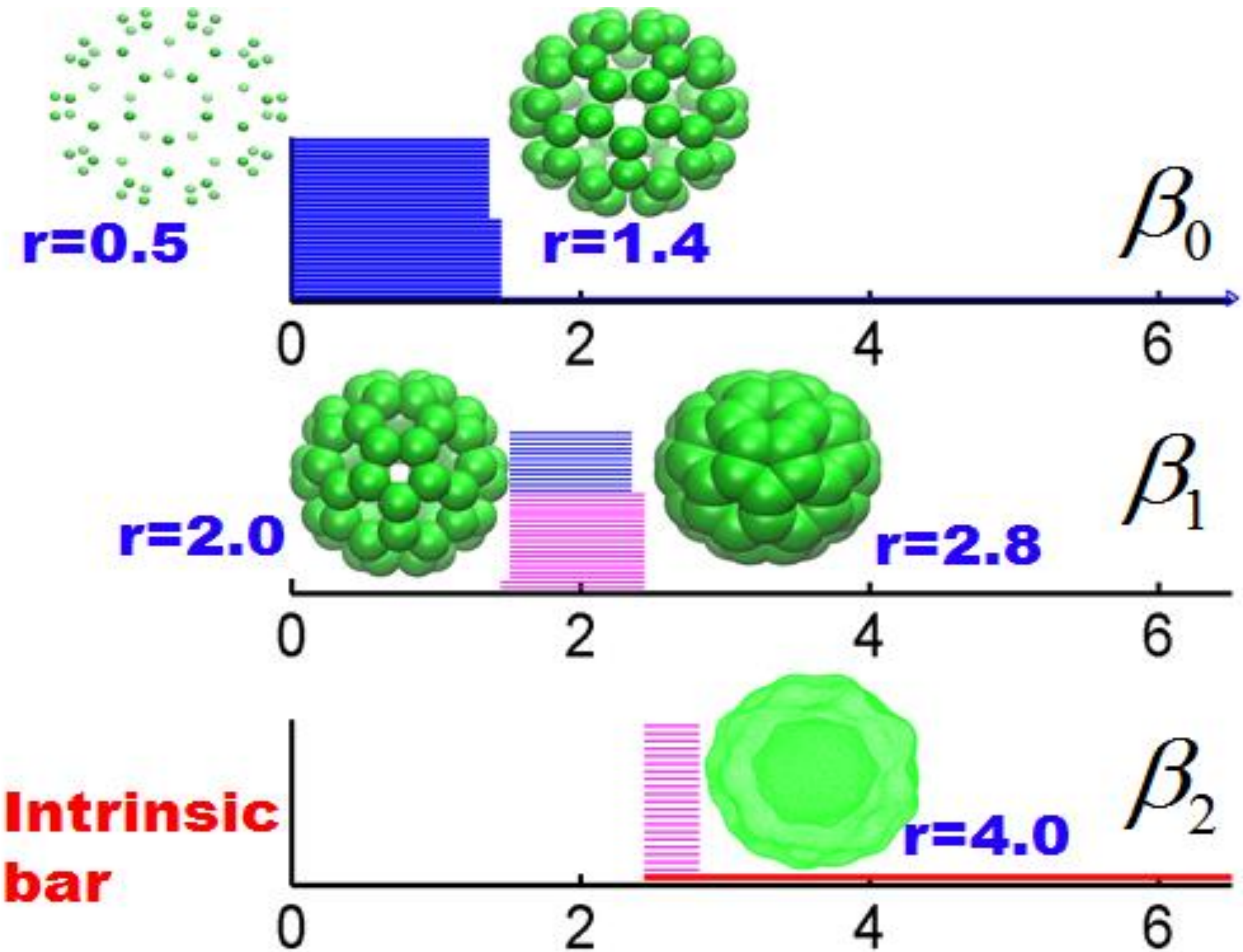


Short bars are NOT noise!

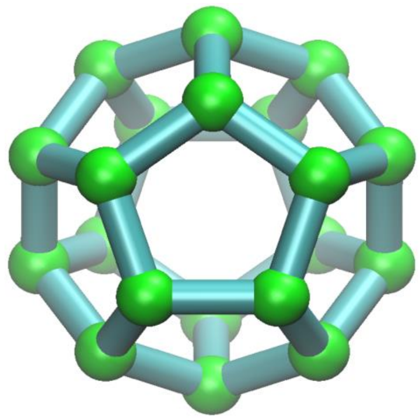
(Yao et al. ,JCP, 2012; Chang et al PLOS One , 2013) (Xia & Wei, IJNMBE, 2014)

Radius filtration of C_{60}

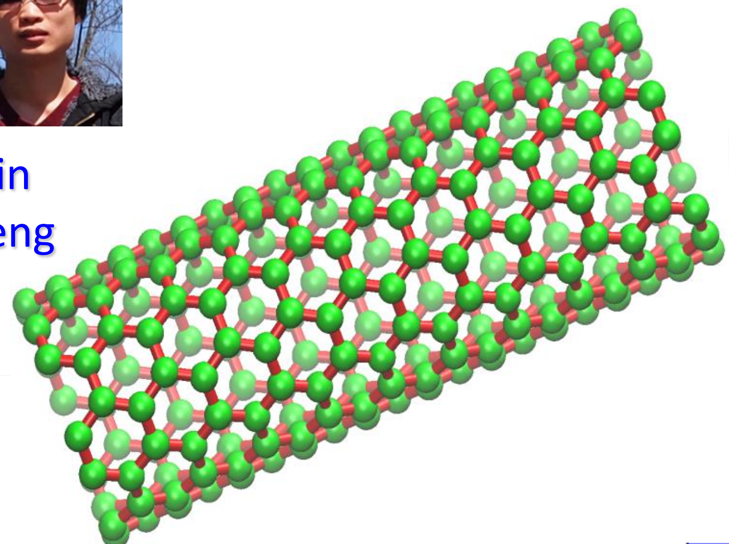
(Xia, Feng, Tong & Wei, JCC, 2015)

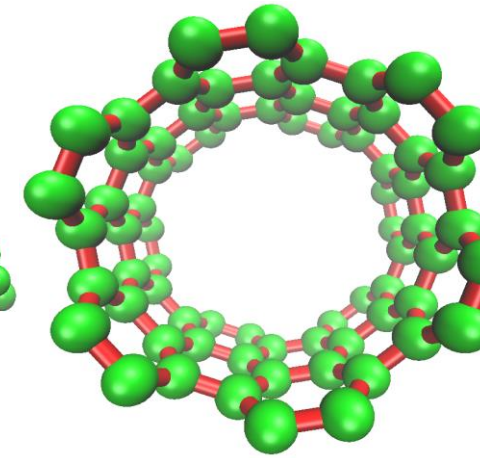


Topological fingerprints of nano molecules

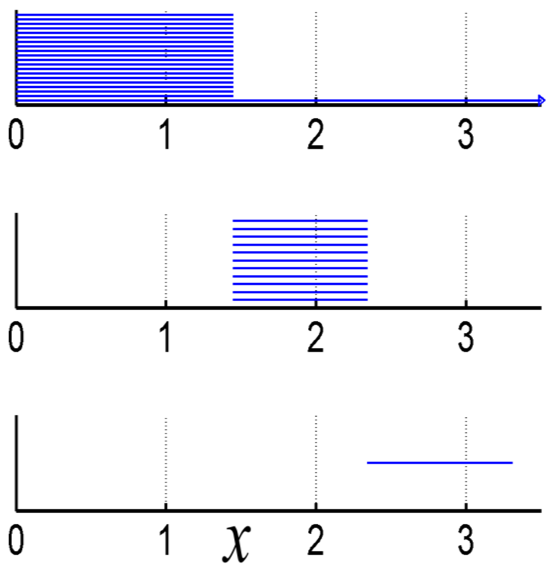


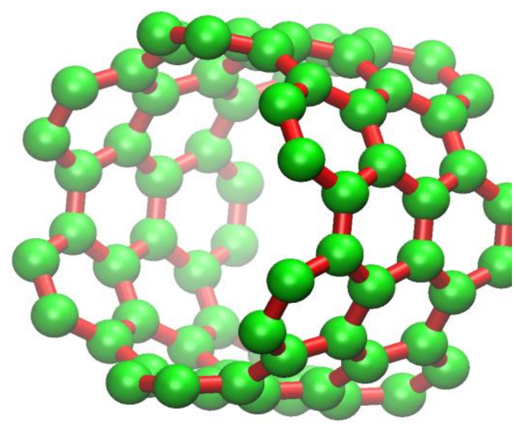
Xin feng



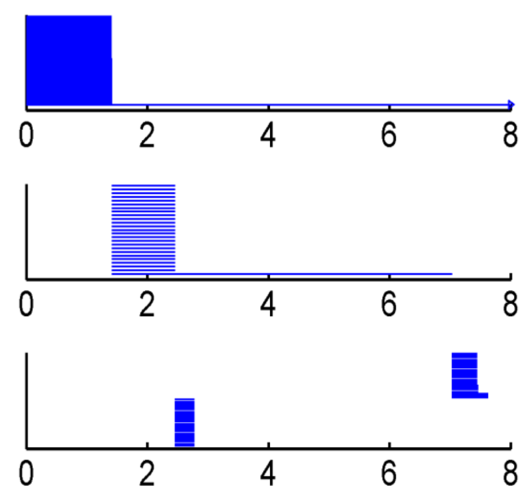


Fullerene C_{20}





nanotube



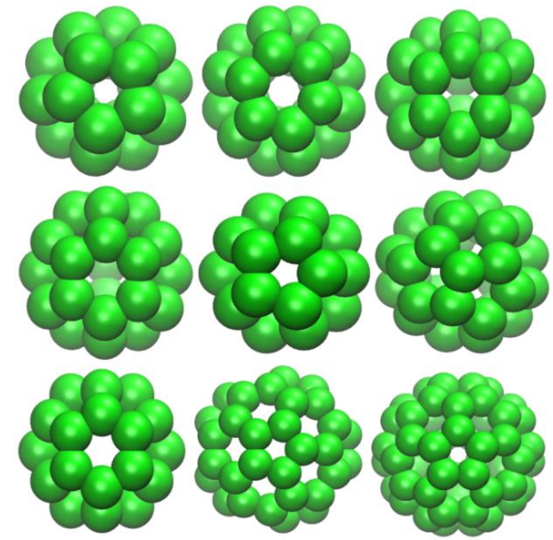
(Xia, Feng, Tong & Wei, JCC, 2015)

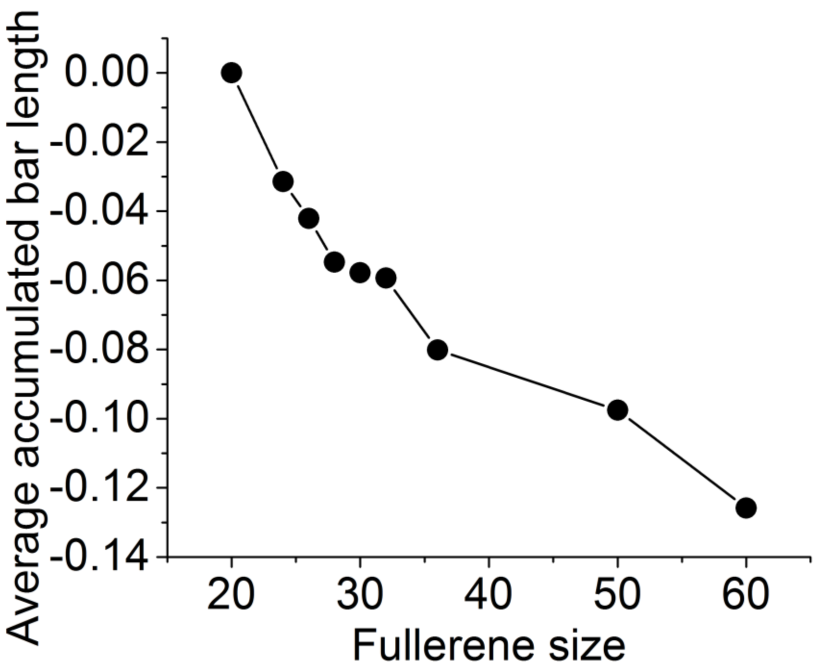
Topology-function relationship --- fullerene heat formation energies

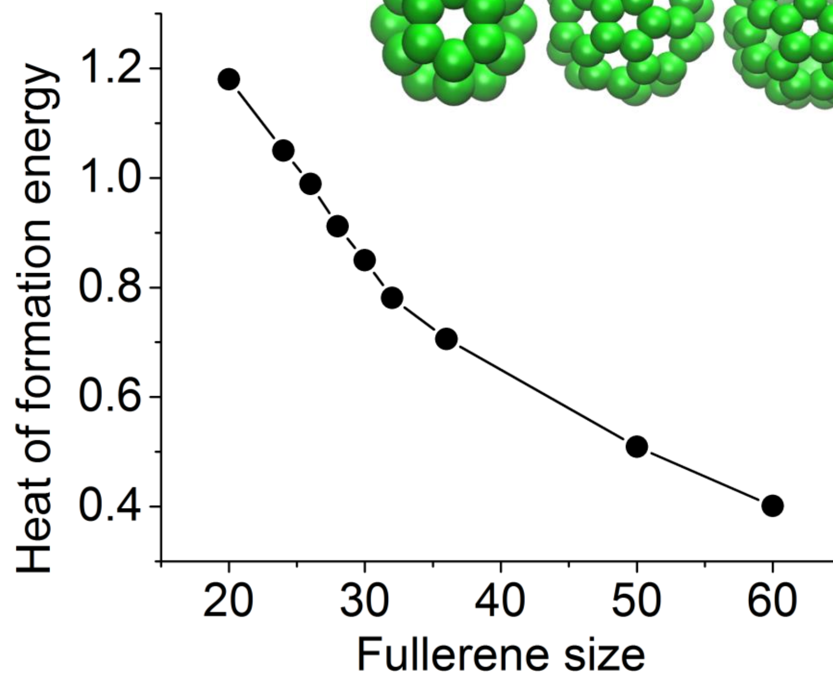
(Xia, Feng, Tong & Wei, JCC, 2015)

$$E \approx \frac{1}{B_2}$$

$$B_2 = \frac{1}{N} {\sum_j}' l_j(\beta_2)$$

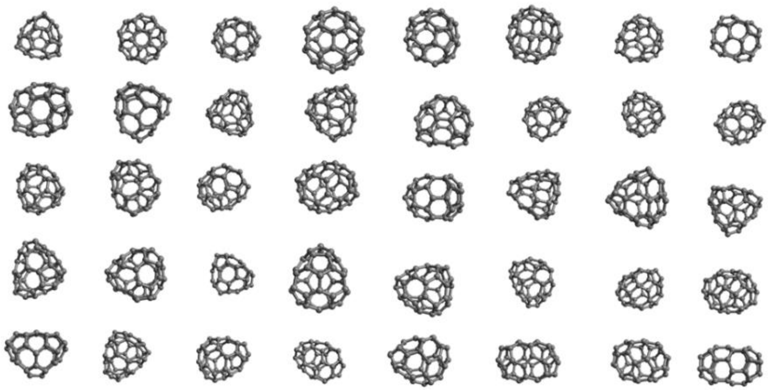




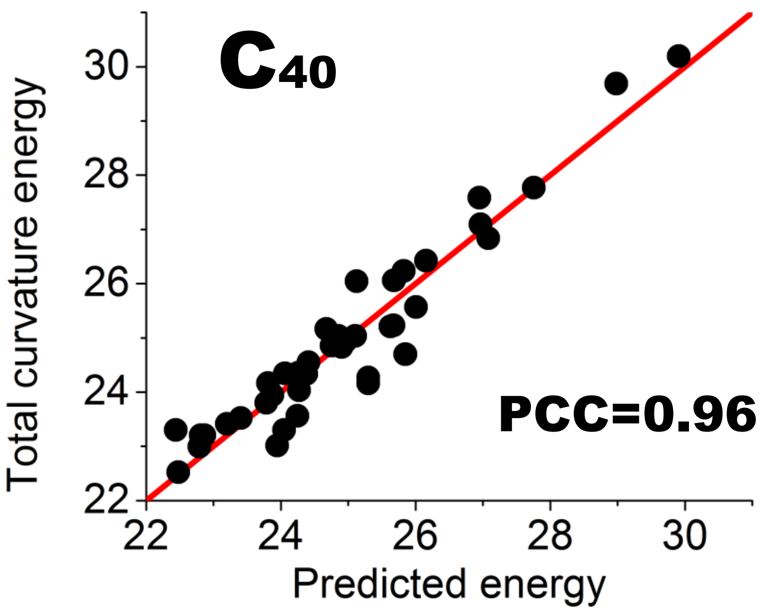


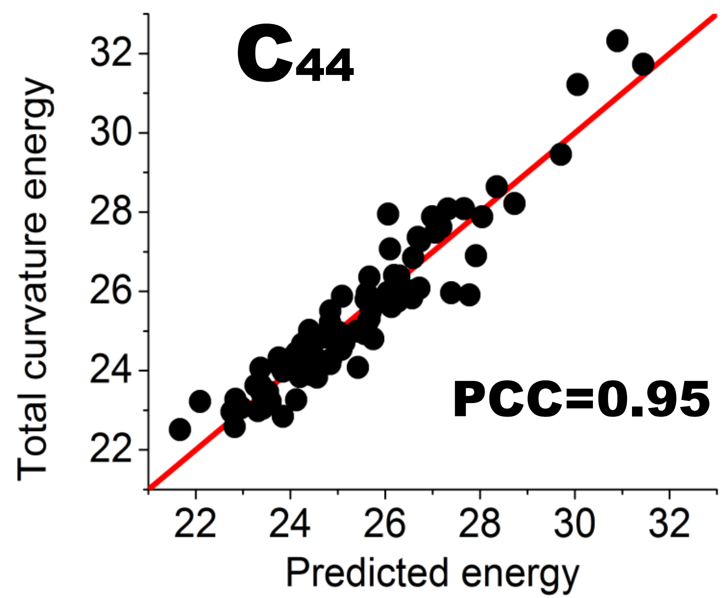
Persistent homology prediction of fullerene isomer total strain energies

(Wang & Wei, JCP, 2016;
Xia, Feng, Tong & Wei, JCC, 2015)

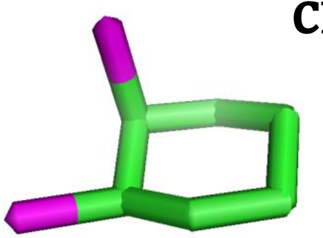


$$E \approx \frac{1}{l_j(\beta_2)}$$

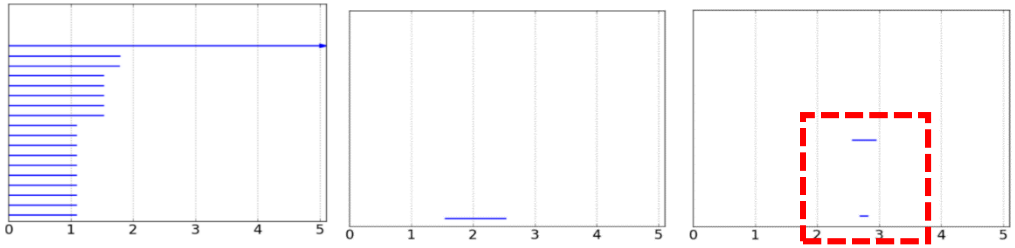


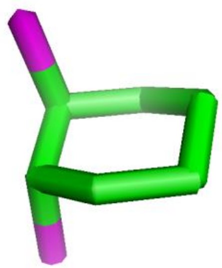


Topological fingerprints of stereoisomers

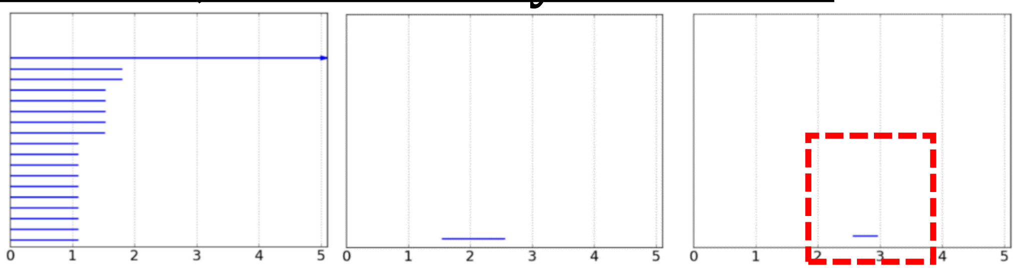


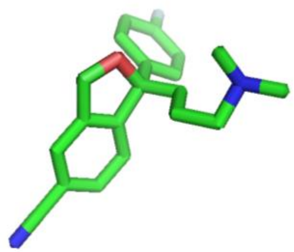
cis-1,2-Dichlorocyclohexane



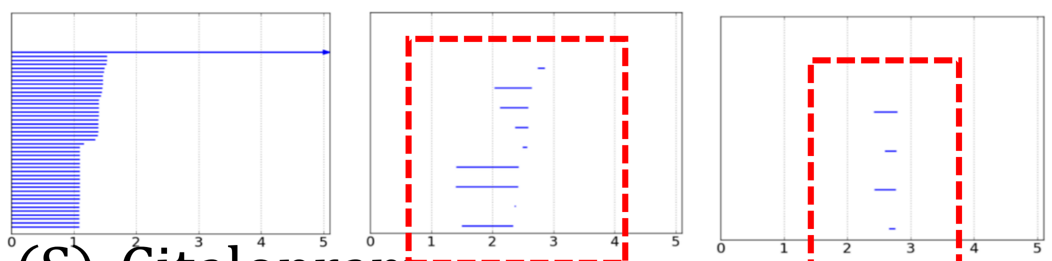


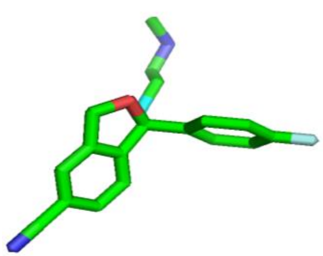
trans-1,2-Dichlorocyclohexane



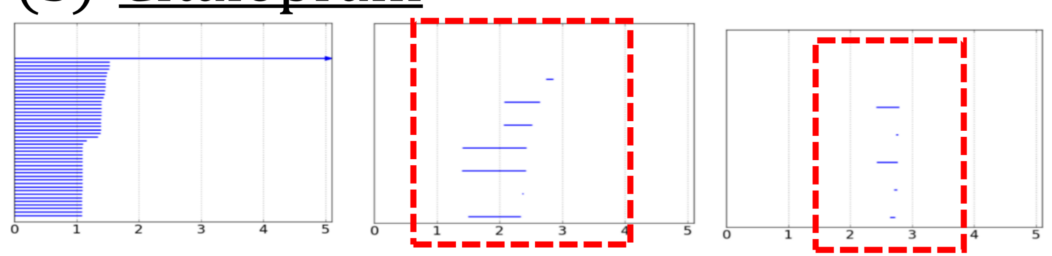


(R)-citalopram



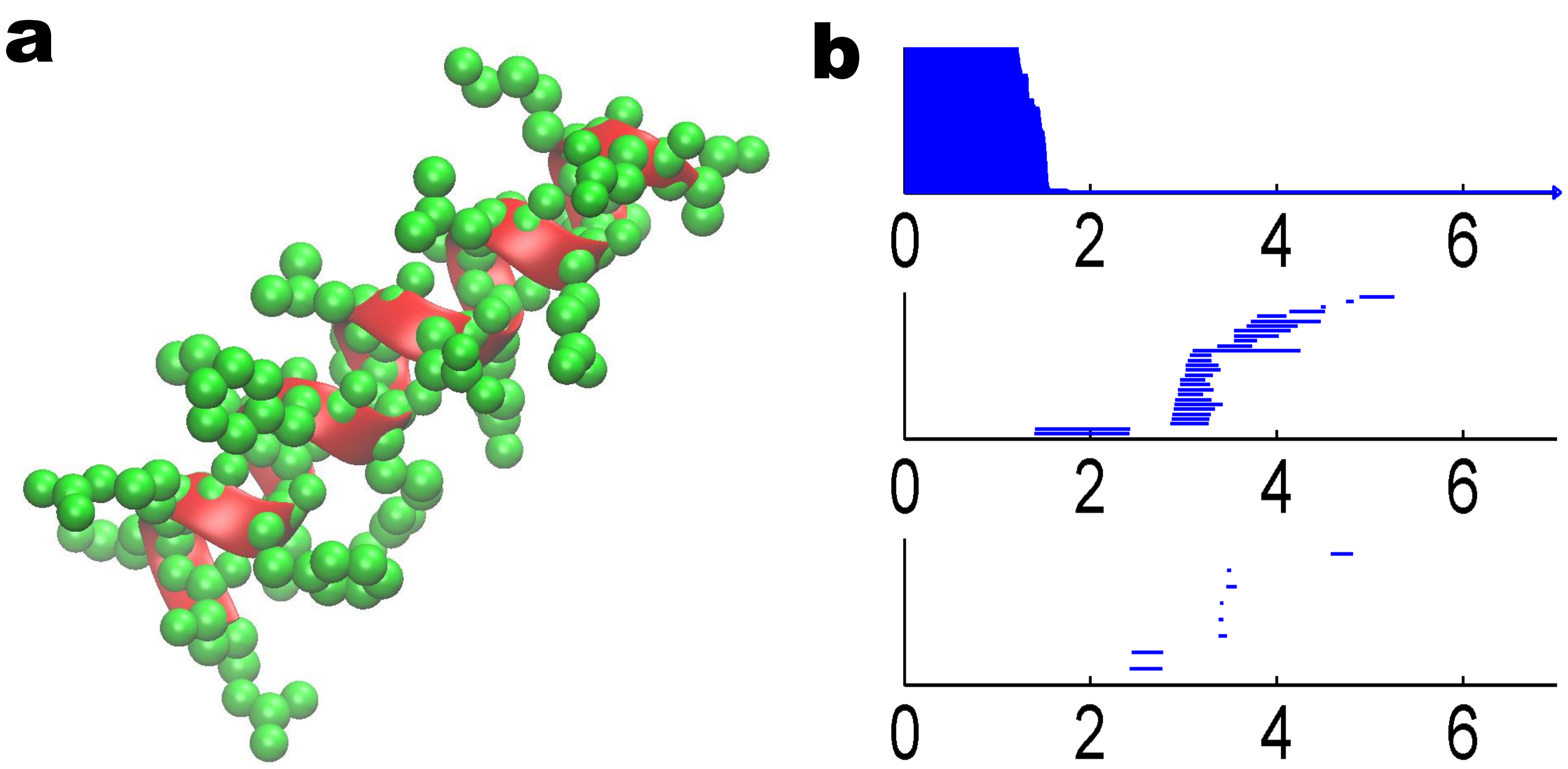


(S)-Citalopram



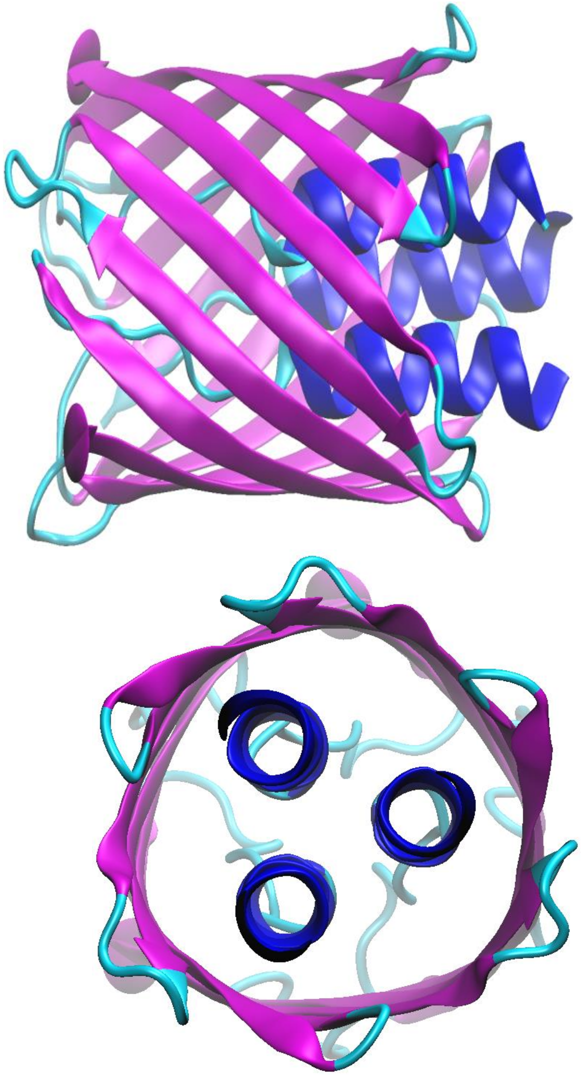
(Wu, Zhao, Wang & Wei, JCC, 2018)

Topological fingerprints of an alpha helix

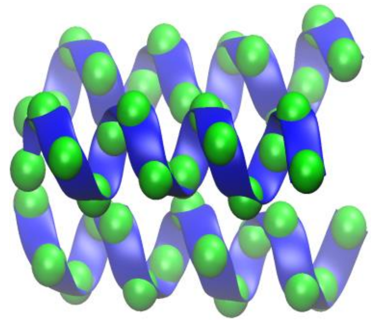


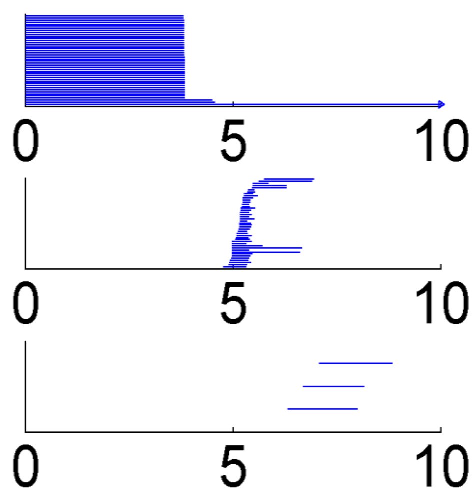
Topological fingerprints of a beta barrel

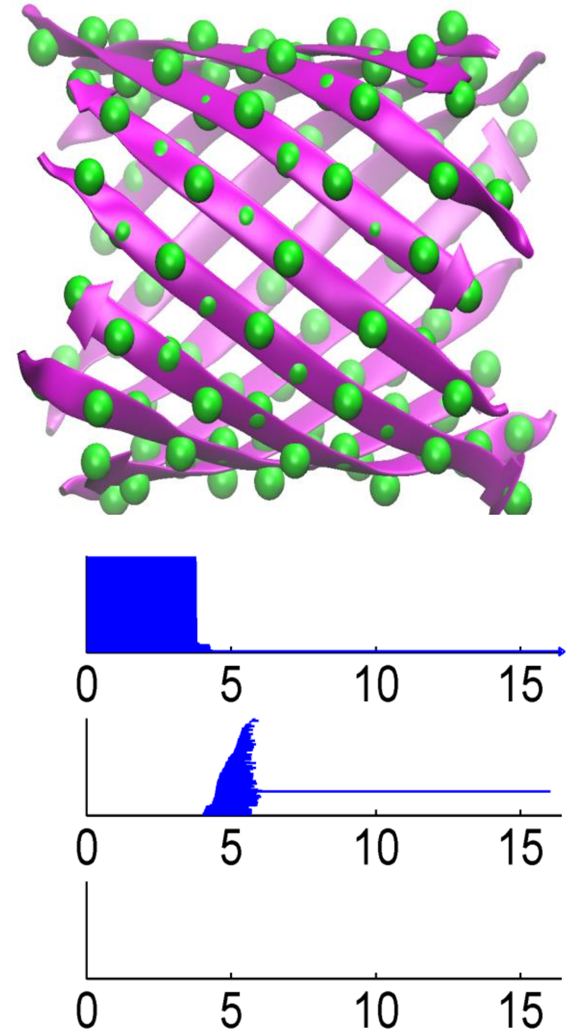
(Xia & Wei, IJNMBE, 2014)



Protein:2GR8

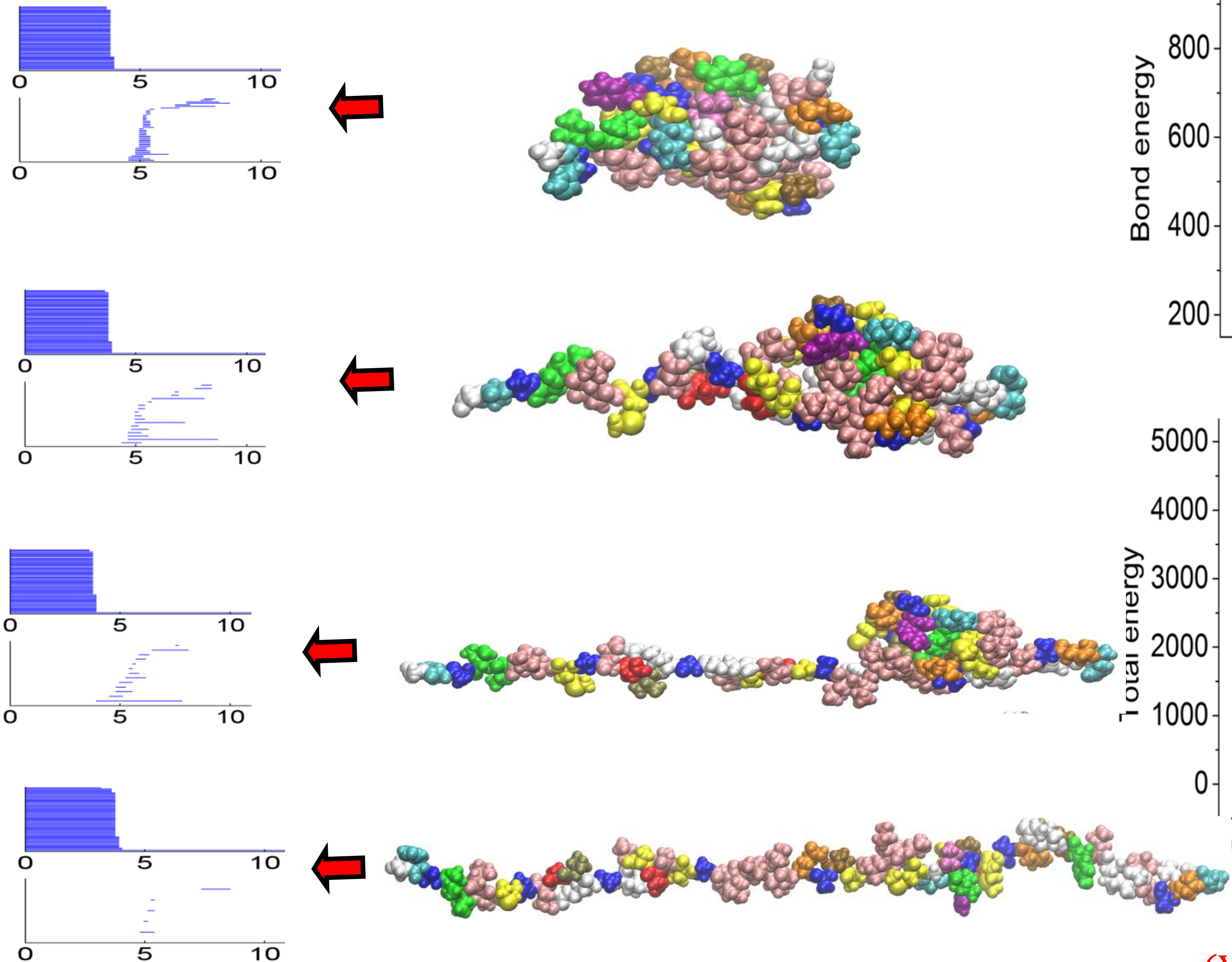


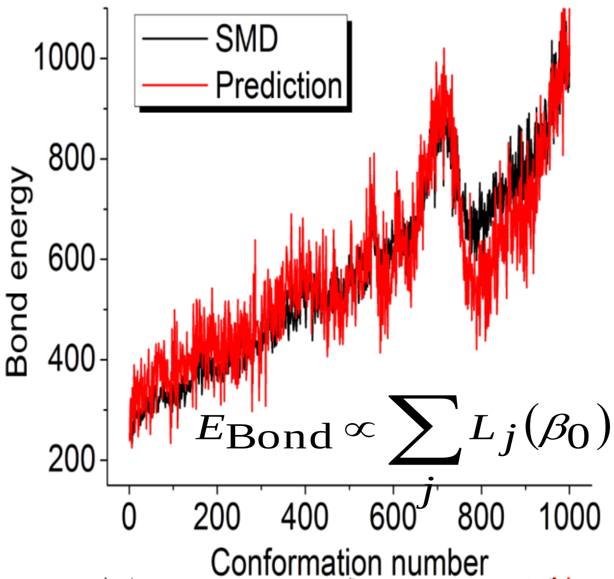


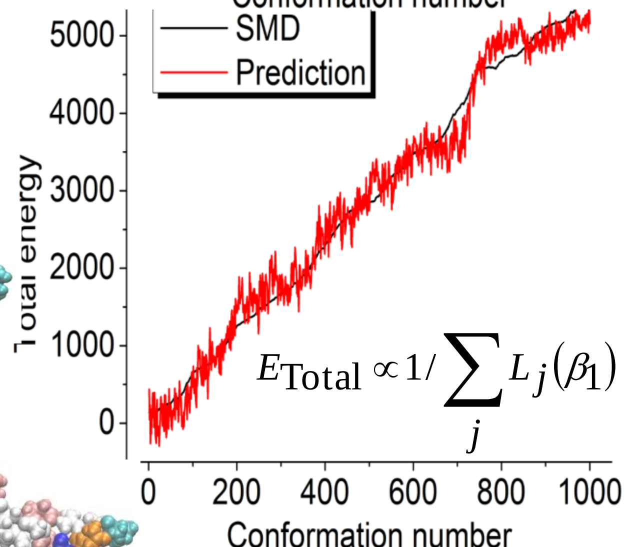


Quantitative topological analysis of protein folding

ID: 1I2T



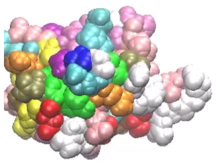




(Xia, Wei, IJNMBE, 2014)

Topological representation of protein folding

2D persistent homology of protein 1UBQ unfolding

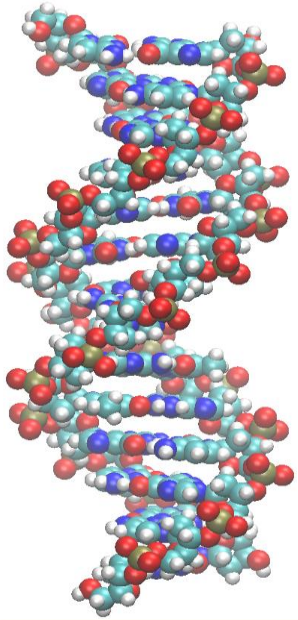


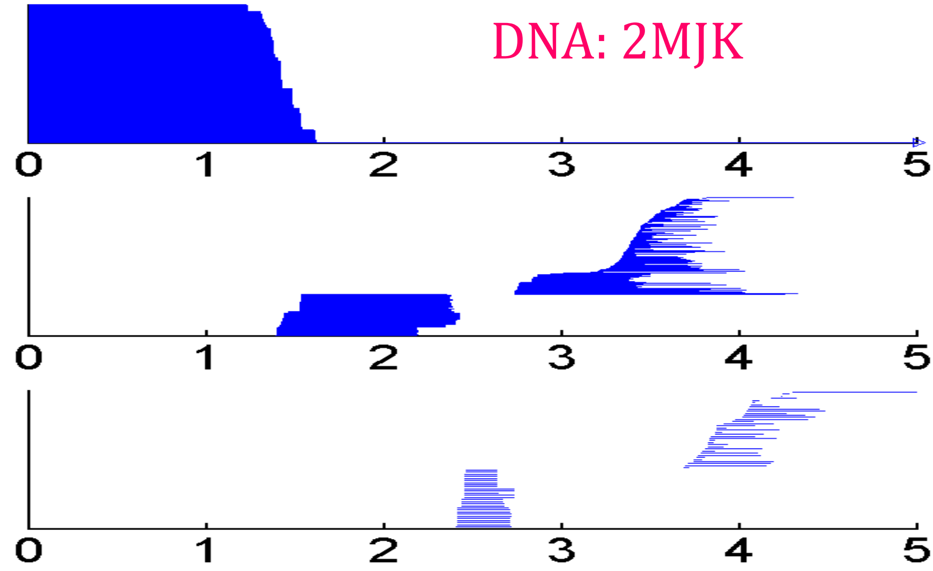
(Xia & Wei, JCC, 2015)

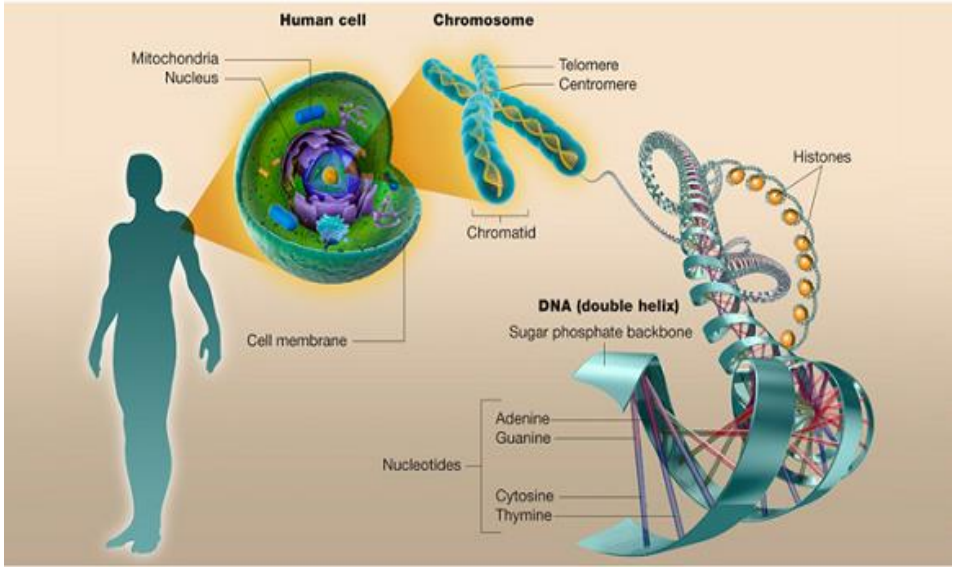
Kelin Xia

DNA topological fingerprints

Personalized topological genome library







Multiscale persistence homology

Scale=1cm

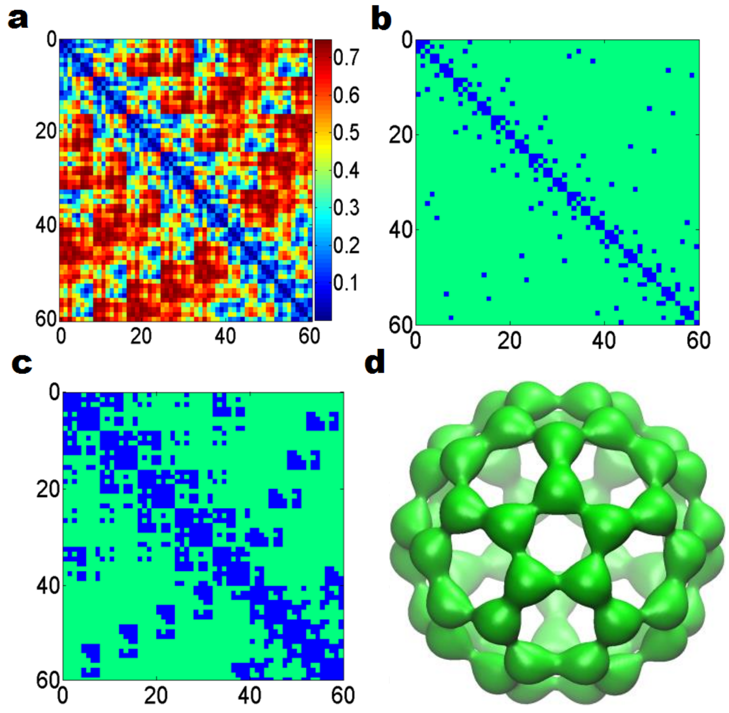
Scale=10cm

Scale=20cm

Matrix filtration:

$$M_{\mathrm{ij}} = 1 - \exp\left(-\frac{(\boldsymbol{r}_i - \boldsymbol{r}_j)^2}{\eta^2}\right)$$

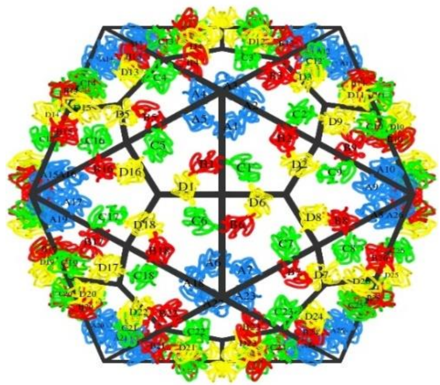
$$\{M_j \geq c_k\}_{k=1}^{N}$$

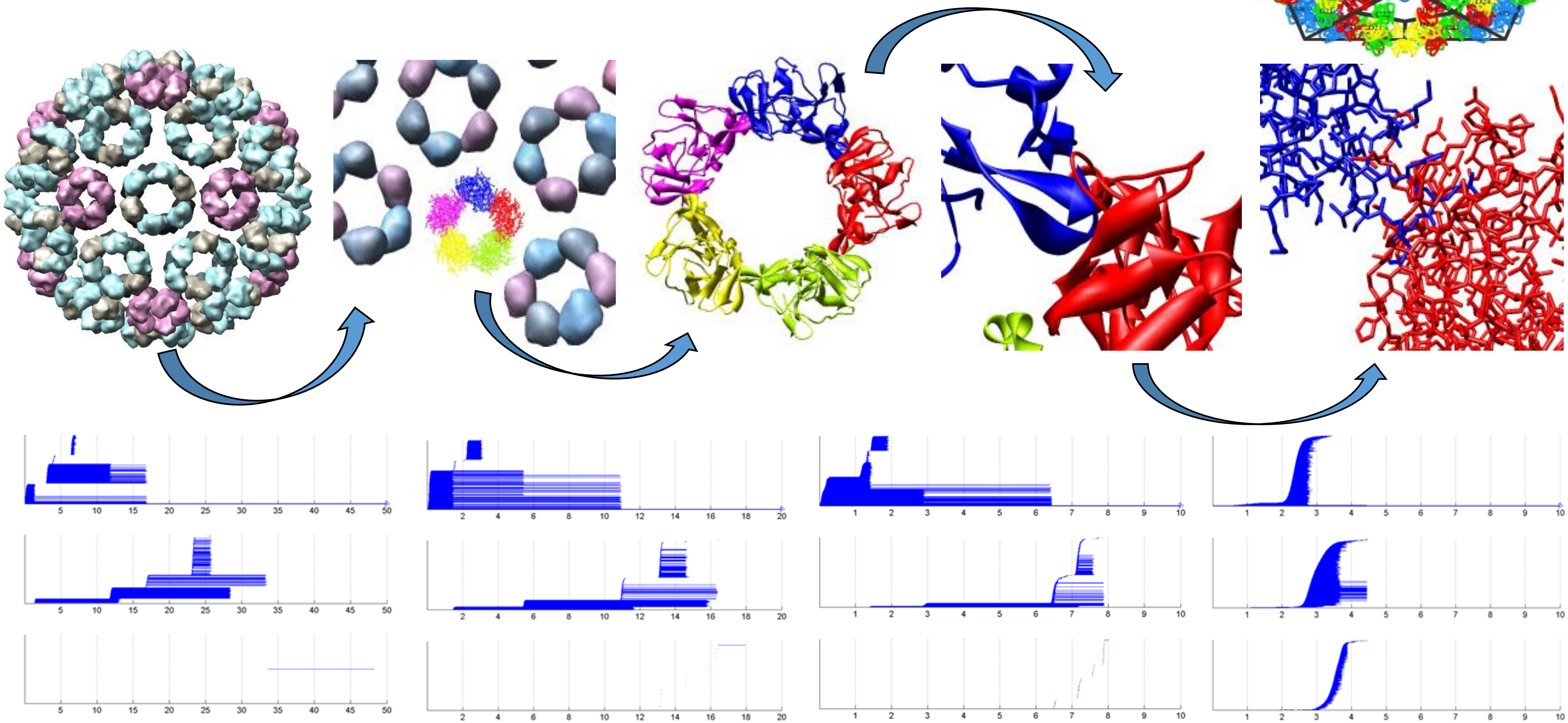


(Xia & Wei, JCC, 2014)

Multiscale topological persistence of a virus

Virus 1DYL has 12 pentagons and 30 hexagons with icosahedral symmetry. The diameter is about 700 Angstrom.





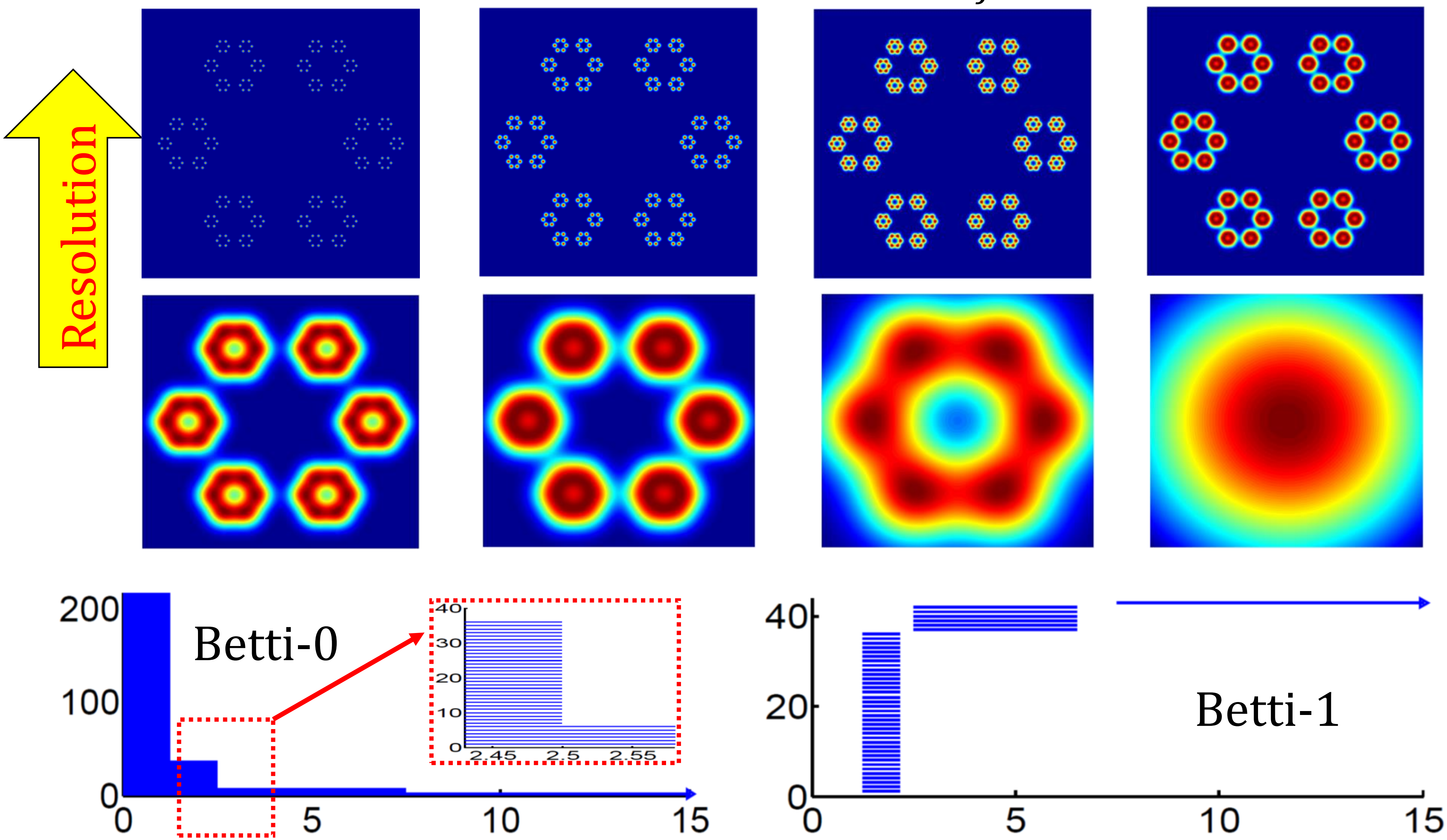
Scale=12A Scale=8A Scale=4A Scale=2A

Multiresolution induced multiscale of a fractal

Introducing the resolution:

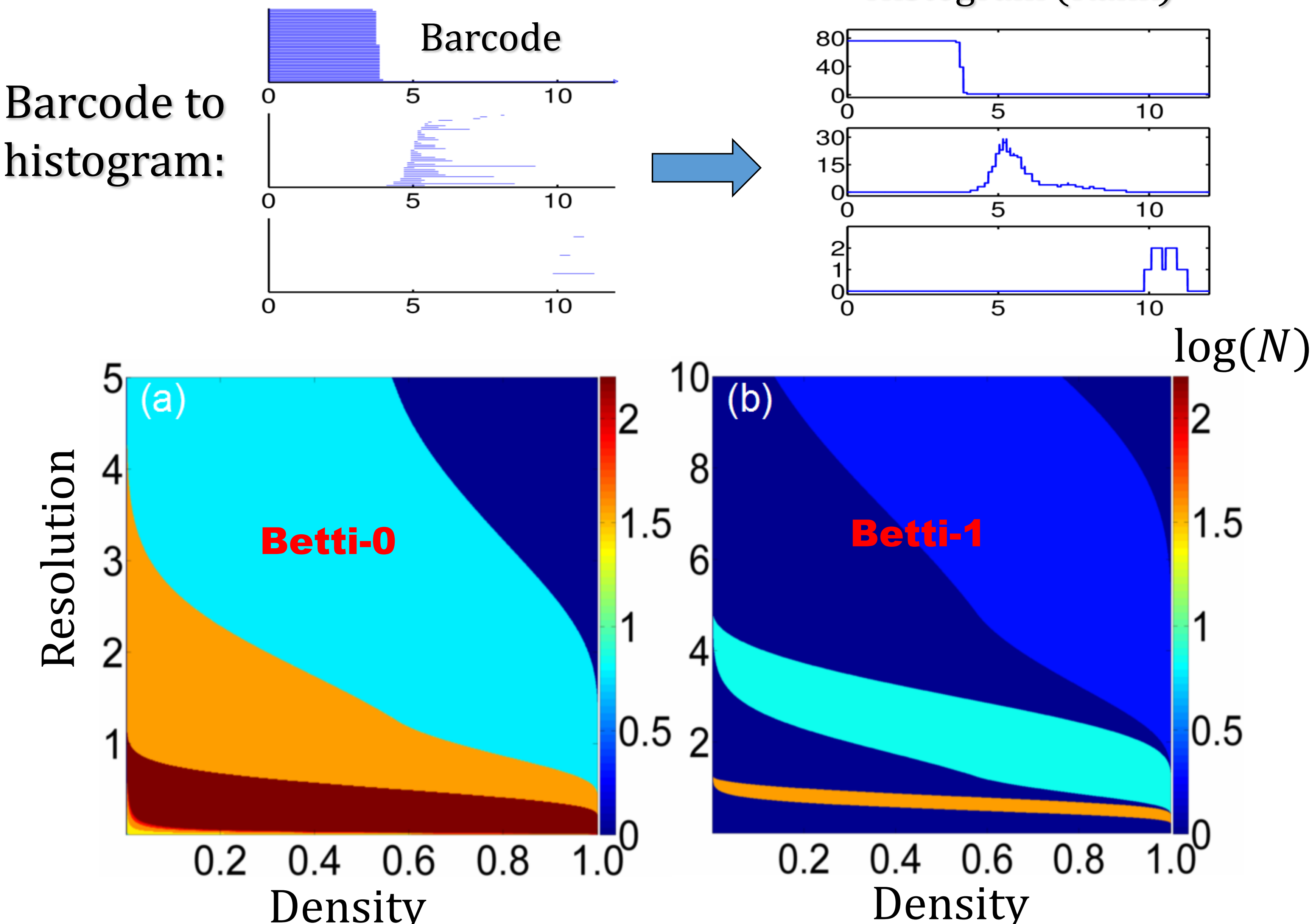
(Xia, Zhao & Wei, JCB, 2015)

$$\rho(\boldsymbol{r},\eta) = \sum_j \exp\left(-\frac{\left(\boldsymbol{r}-\boldsymbol{r}_j\right)^2}{\eta^2}\right)$$



Multiresolution induced multidimensional topology of the fractal

(Xia, Zhao & Wei, JCB, 2015)



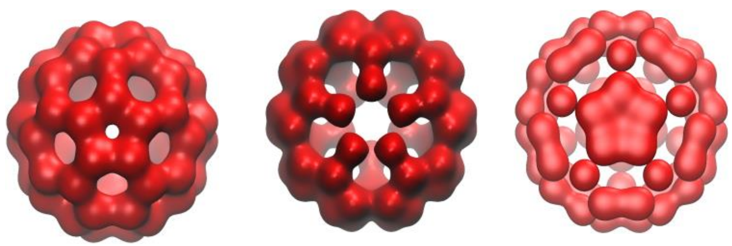
3D persistence

(Xia & Wei, JCC, 2015)

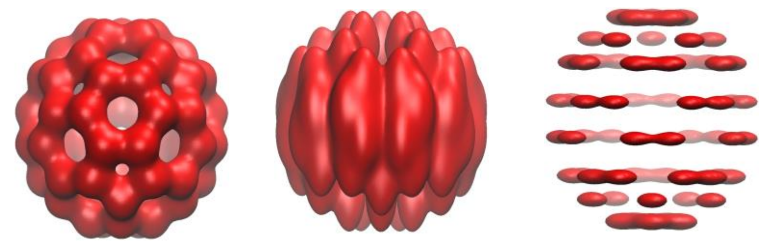
Anisotropic resolution (3D):

$$\rho(\boldsymbol{r}, \eta_x, \eta_y) = \sum_j \exp\left(-\left(\frac{(x-x_j)^2}{\eta_x^2} + \frac{(y-y_j)^2}{\eta_x^2} + \frac{(z-z_j)^2}{\eta_z^2}\right)\right)$$

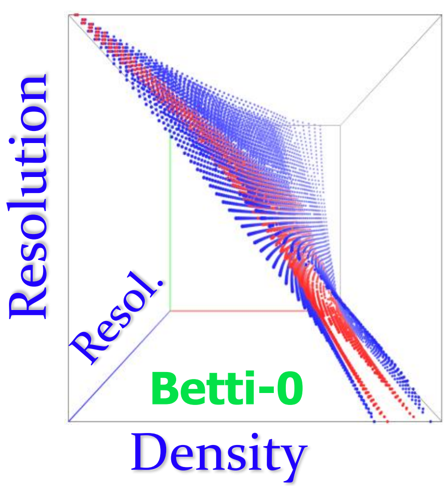
$$\{\rho(\boldsymbol{r}, \eta_x, \eta_z) \geq c_k\}_{k=1}^{N}$$

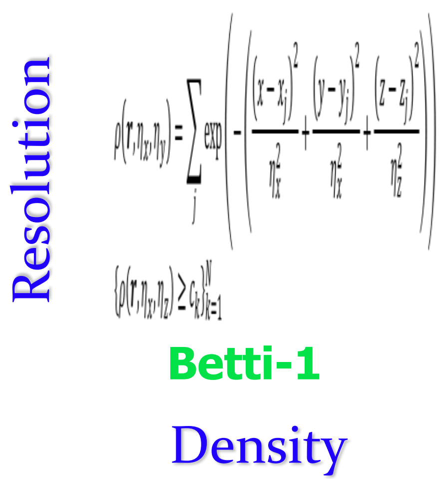


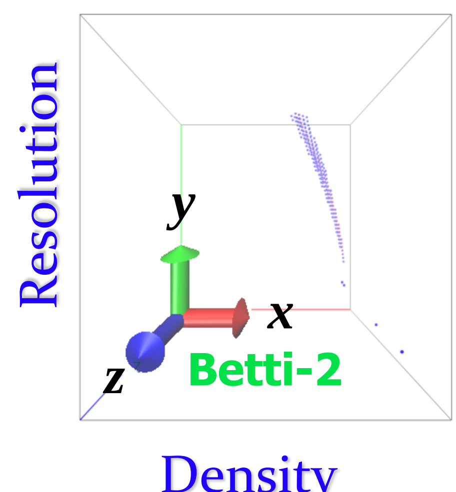
$\eta_x = \eta_y = 0.2\dot{A}, \eta_z = 0.5\dot{A}$



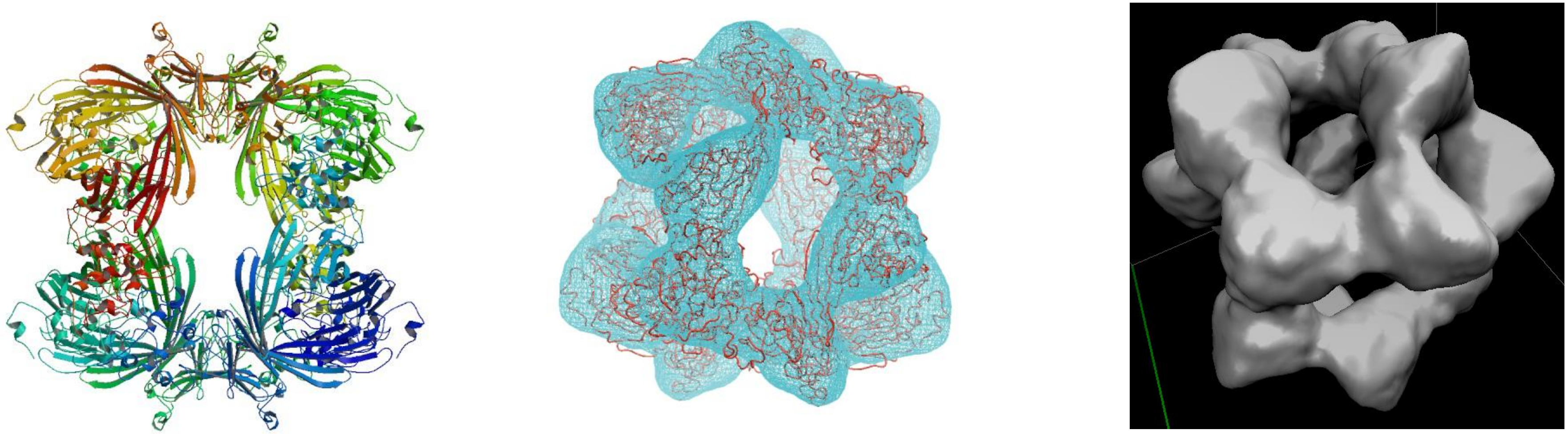
$\eta_x = \eta_y = 0.5\dot{A}, \eta_z = 0.2\dot{A}$

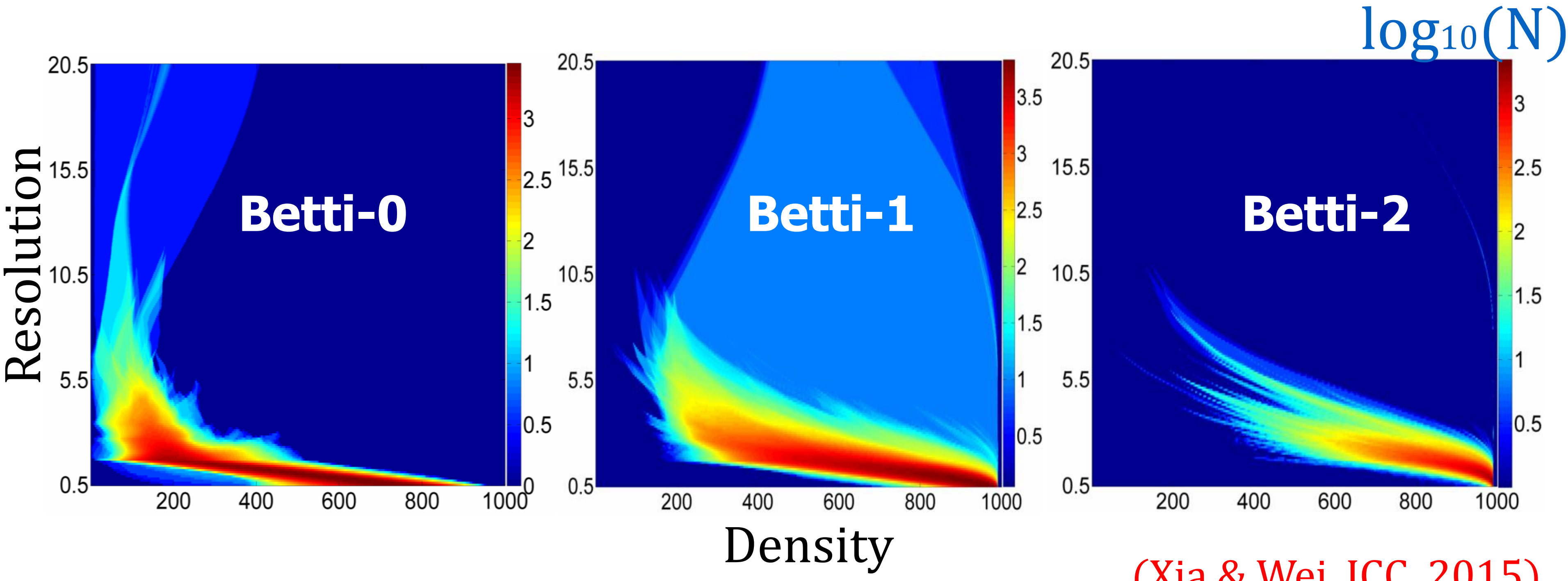






Multiresolution 2D persistence in protein complex 2YGD





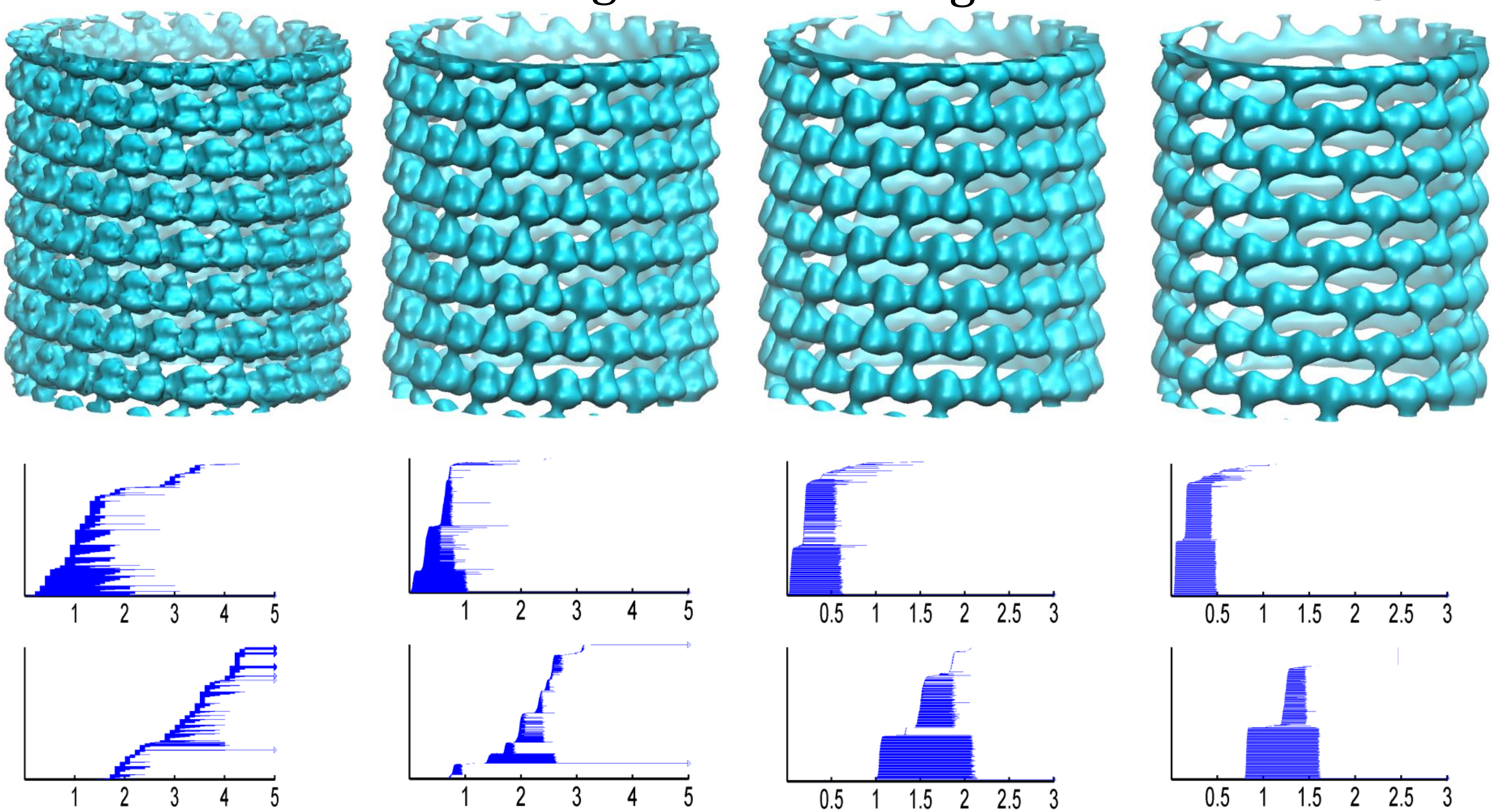
(Xia & Wei, JCC, 2015)

Topological noise reduction via geometric PDE

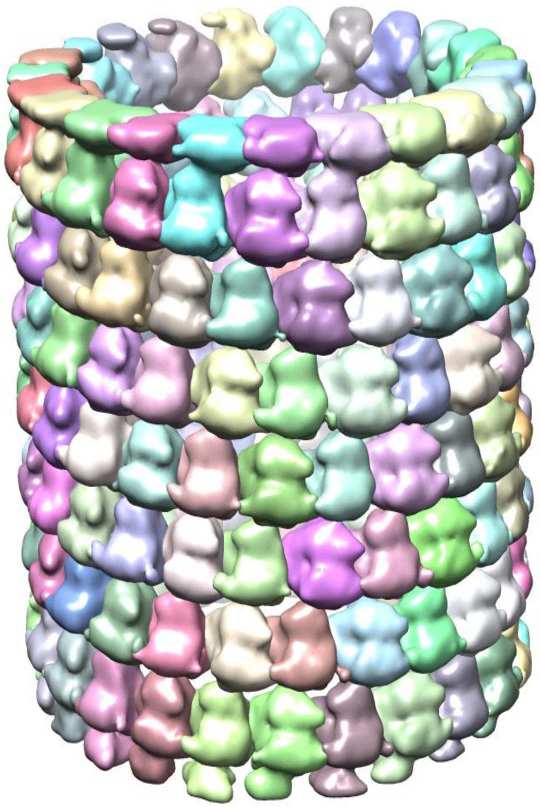
(Xia & Wei, IJNMBE 2015)

$$\frac{\partial S}{\partial t} = |\nabla S| \, \nabla \cdot \left(\frac{\gamma \nabla S}{|\nabla S|} \right)$$

Original data | Ten-iteration denoising | Twenty-iteration denoising | Forty-iteration denoising

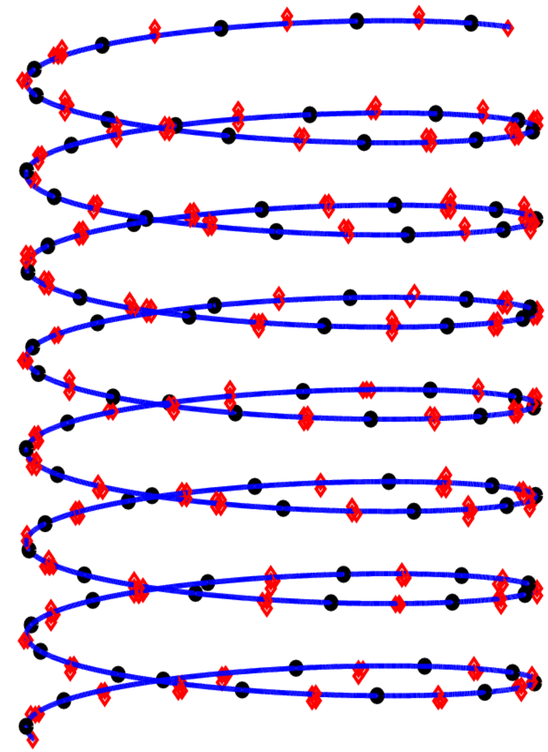


Microtubule analysis with cryo-EM data



EMD_1129

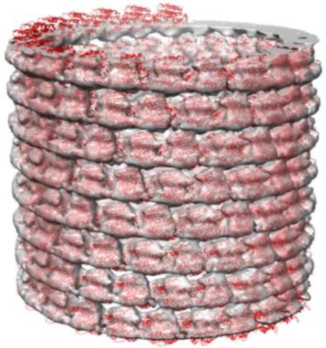
Molecular structure
fitted with tubulin 1JFF



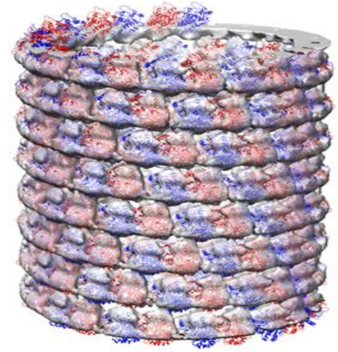
Helix backbone

(Xia & Wei, JCC, 2014)

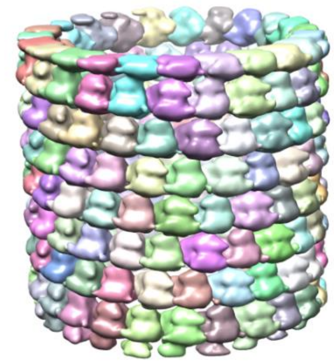
Persistent homology for ill-posed inverse problems



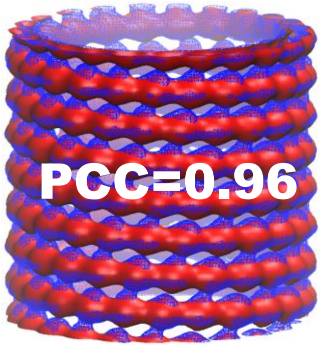
Fitted with one-type of tubulins

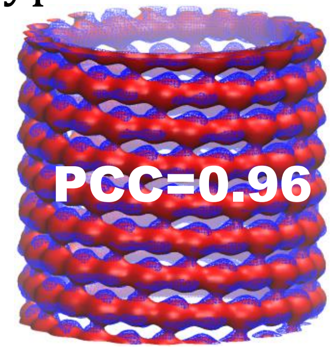


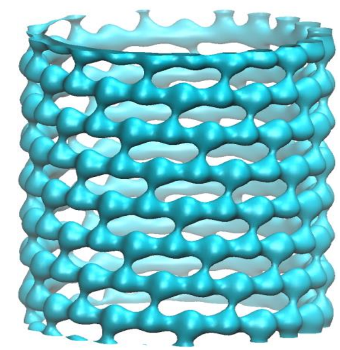
Fitted with two-types of tubulins



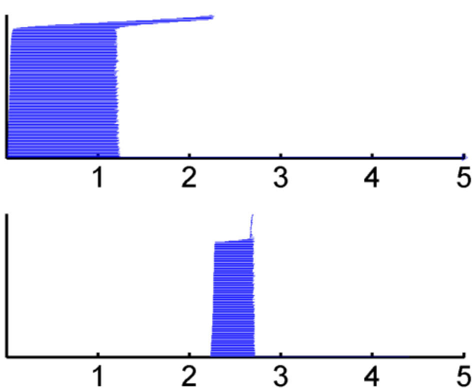
Original data: microtubule

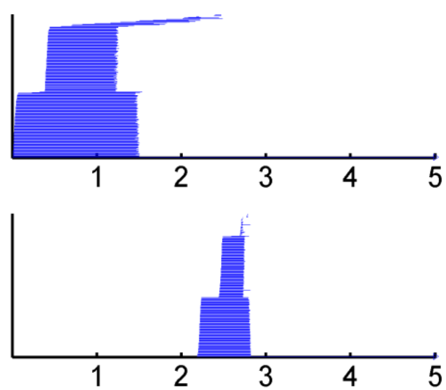


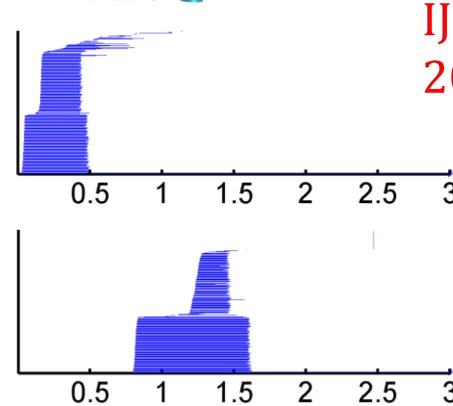




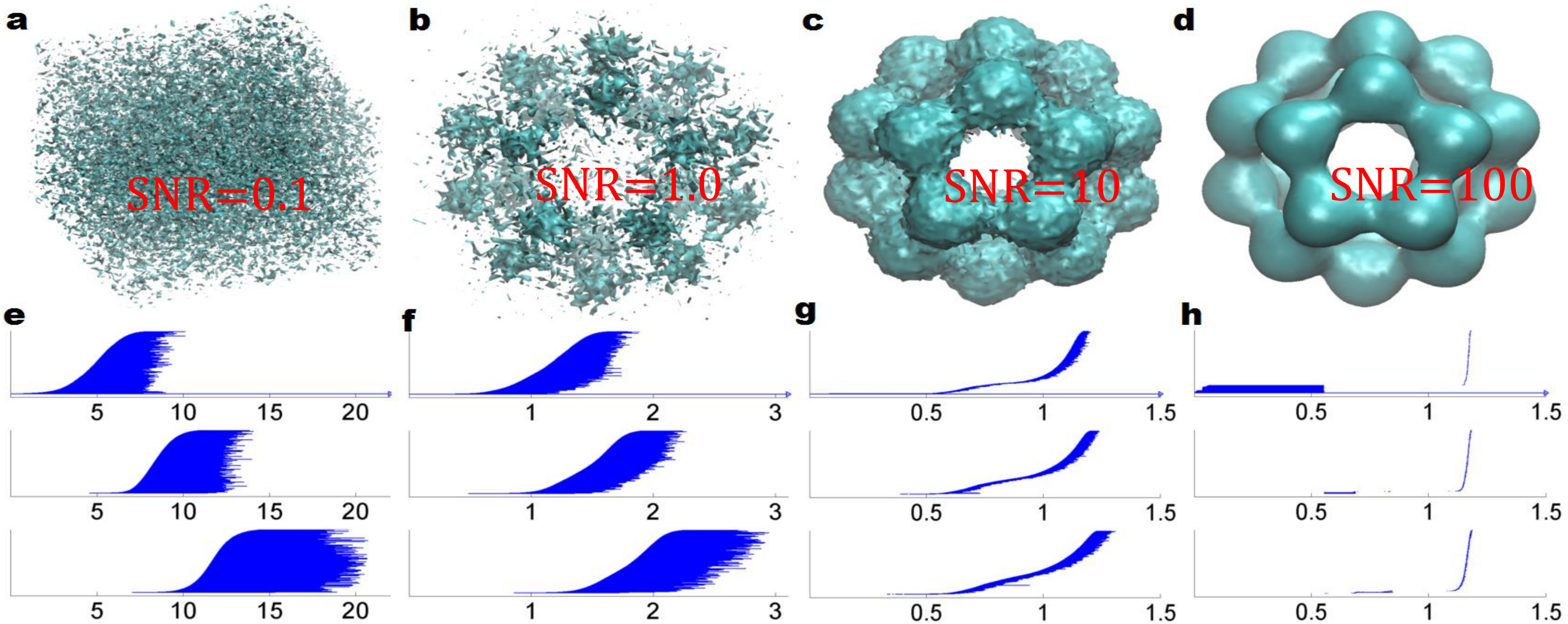
(Xia, Wei, IJNMBE, 2015)



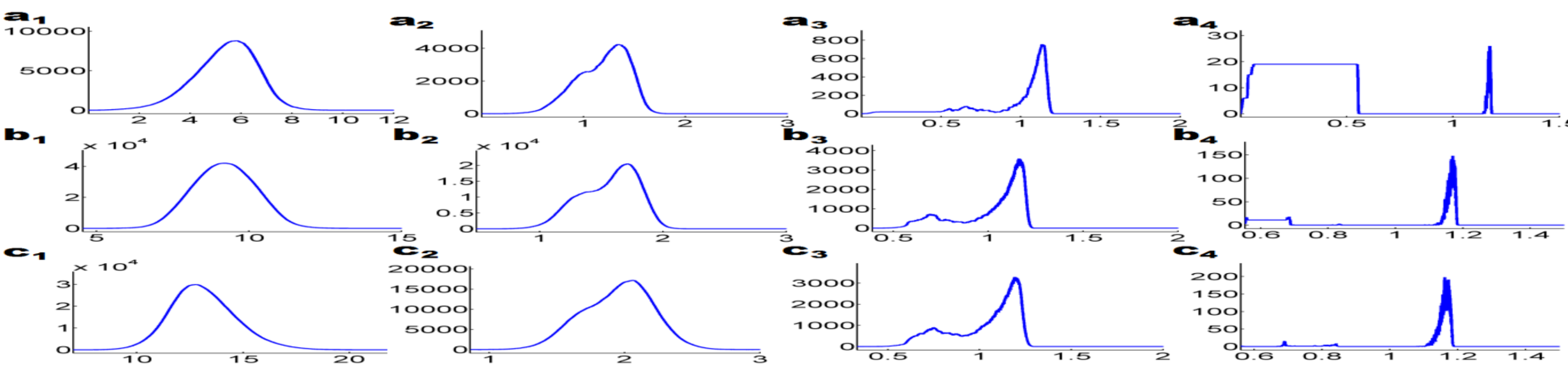




Topological analysis of Gaussian noise



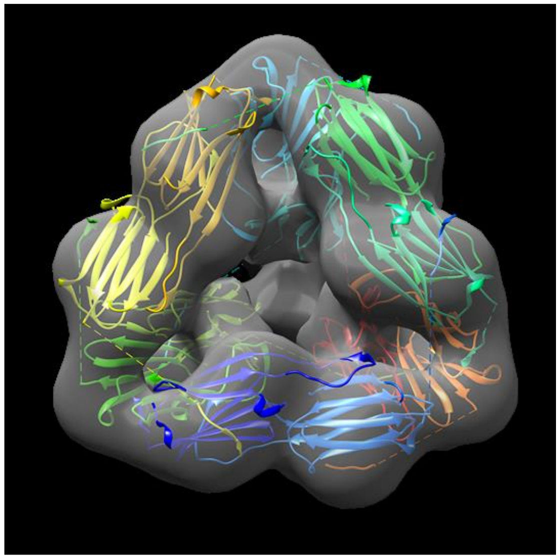
Barcode histograms show the Gaussian distribution of the noise



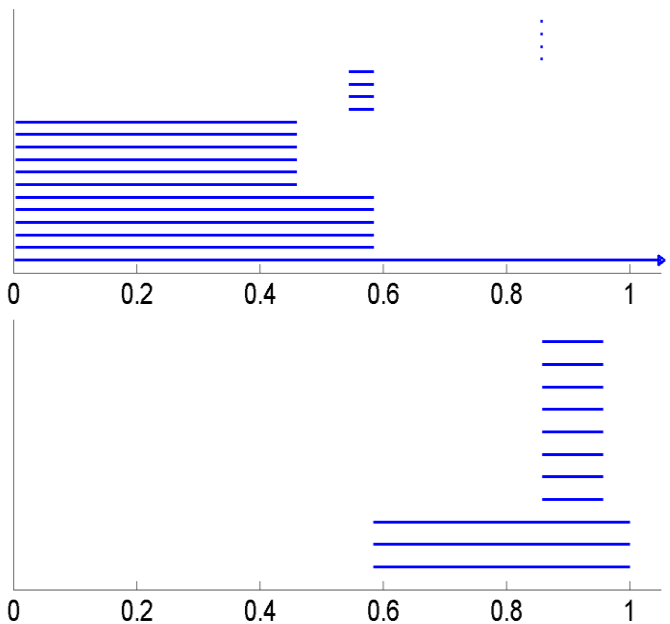
(Xia & Wei, JCC, 2014)

Detection of cryo-EM structure reconstruction errors

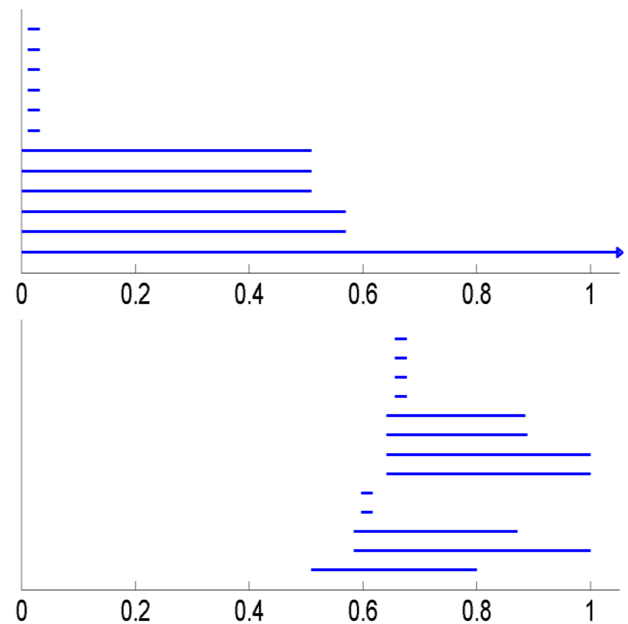
Structure map



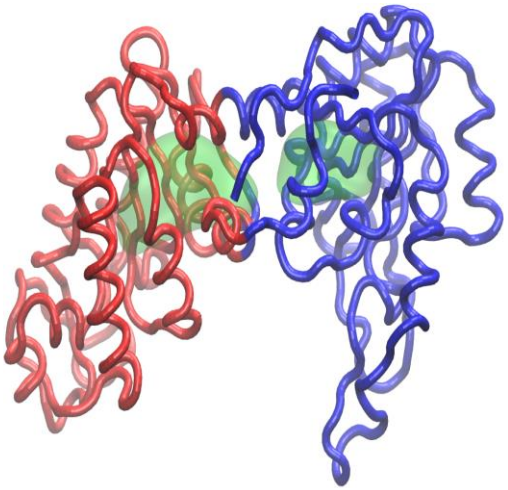
Structure barcodes

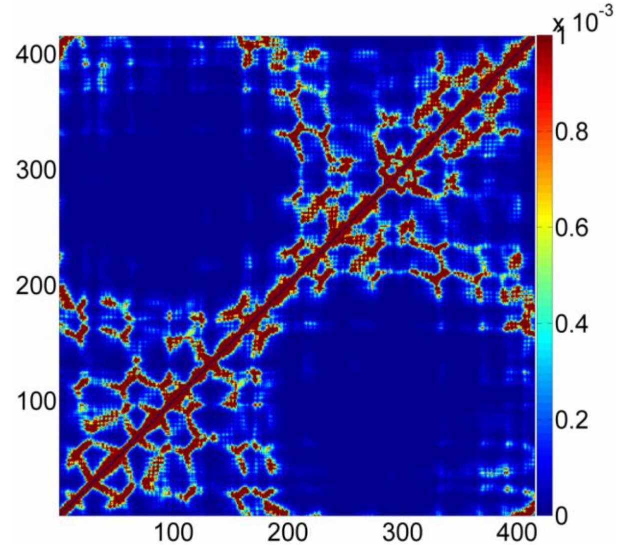


Map barcodes



Topology based protein domain classification

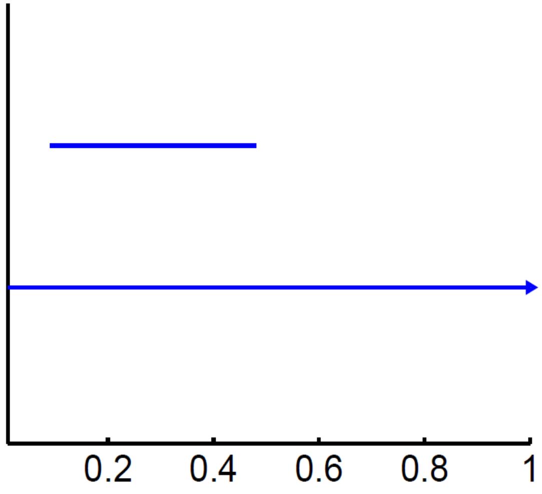


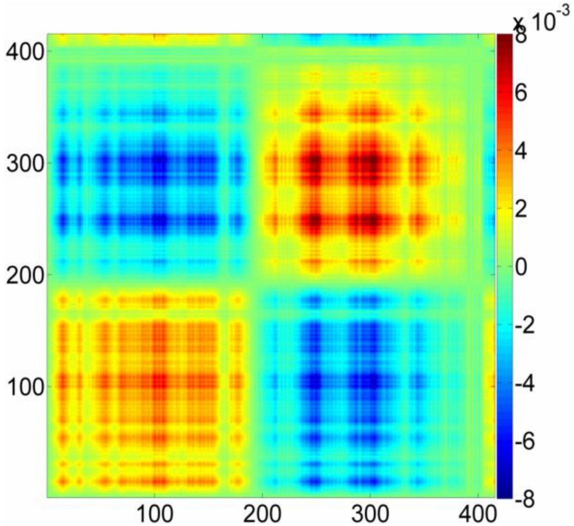


Modularity (graph)

Flexibility-rigidity index

(Xia, Zhao & Wei, JCB, 2015)



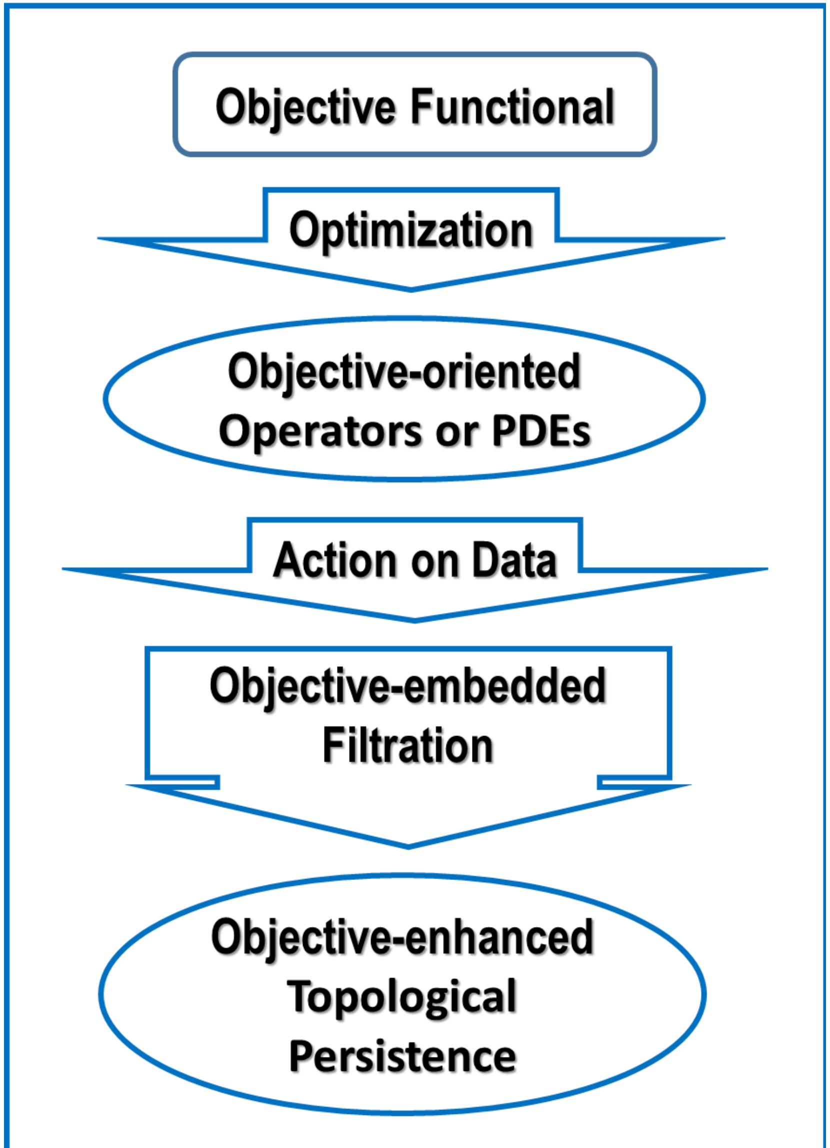


Topology

Graph Laplacian

Objective oriented persistent homology

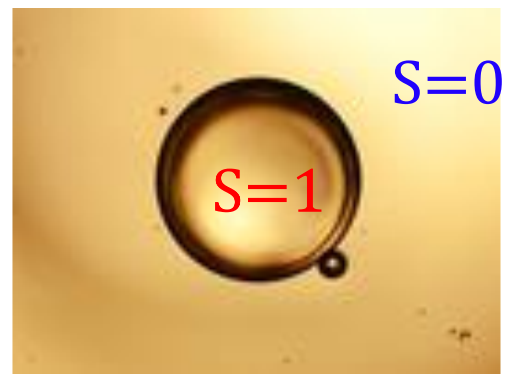
(Wang & Wei, JCP, 2016)



Objective: Minimal surface energy

$$G = \int \gamma[\text{area}]d\boldsymbol{r}, \quad [\text{area}] = |\nabla S|$$

where gamma (γ) is the surface tension, and S is a surface characteristic function:



Generalized Laplace-Beltrami flow

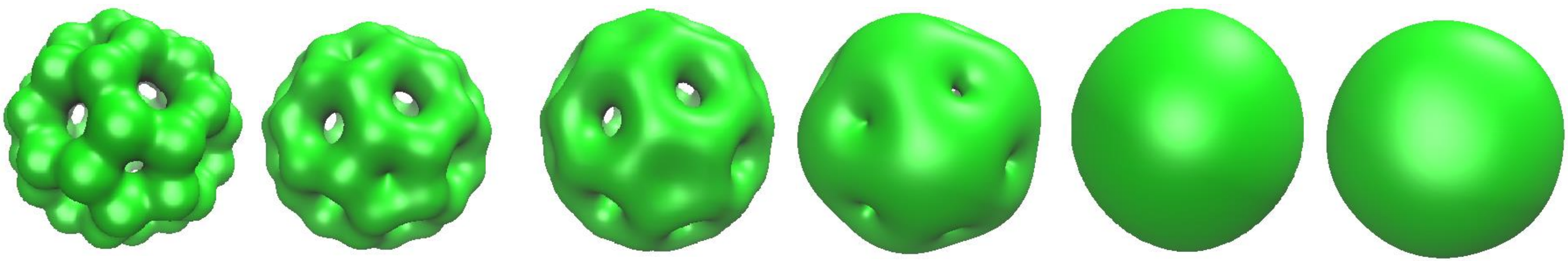
$$\frac{\partial S}{\partial t} = |\nabla S| \, \nabla \cdot \left(\frac{\gamma \nabla S}{|\nabla S|} \right)$$

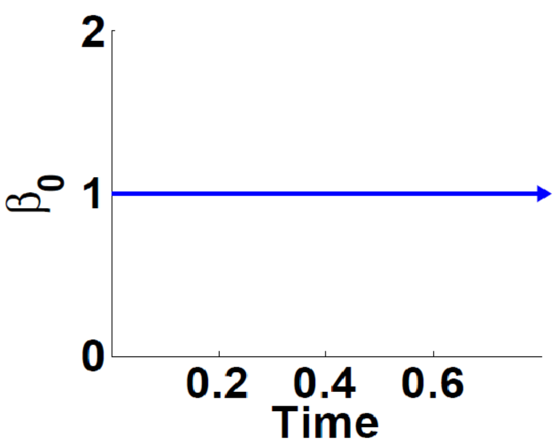
Objective oriented persistent homology

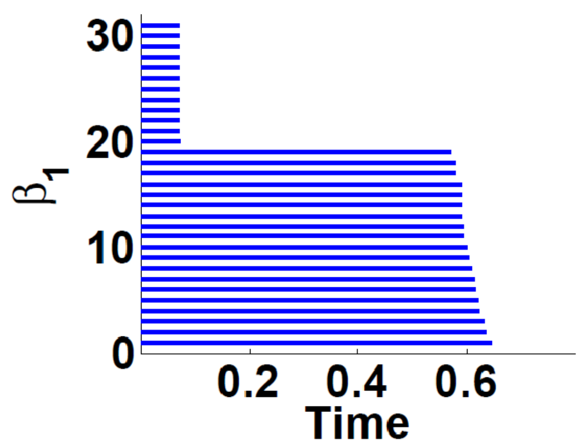
Level sets generated from Laplace-Beltrami flow

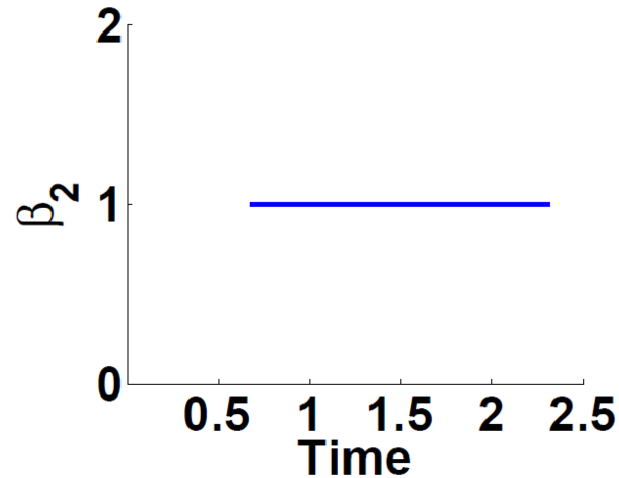
$$\frac{\partial S}{\partial t} = |\nabla S|\,\nabla \cdot \left(\frac{\gamma \nabla S}{|\nabla S|}\right)$$

(Wang & Wei, JCP, 2016)



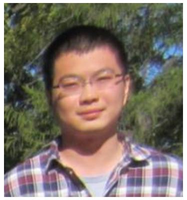






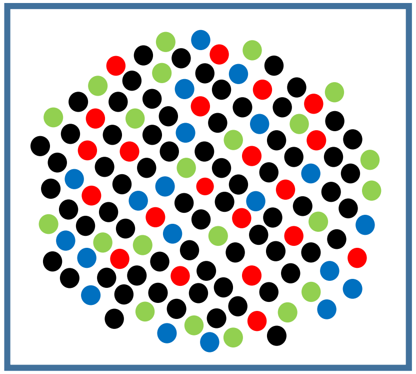
Barcodes are generated by cubical complex and cubical homology. Geometric PDE induced persistent homology

Element-specific persistent homology

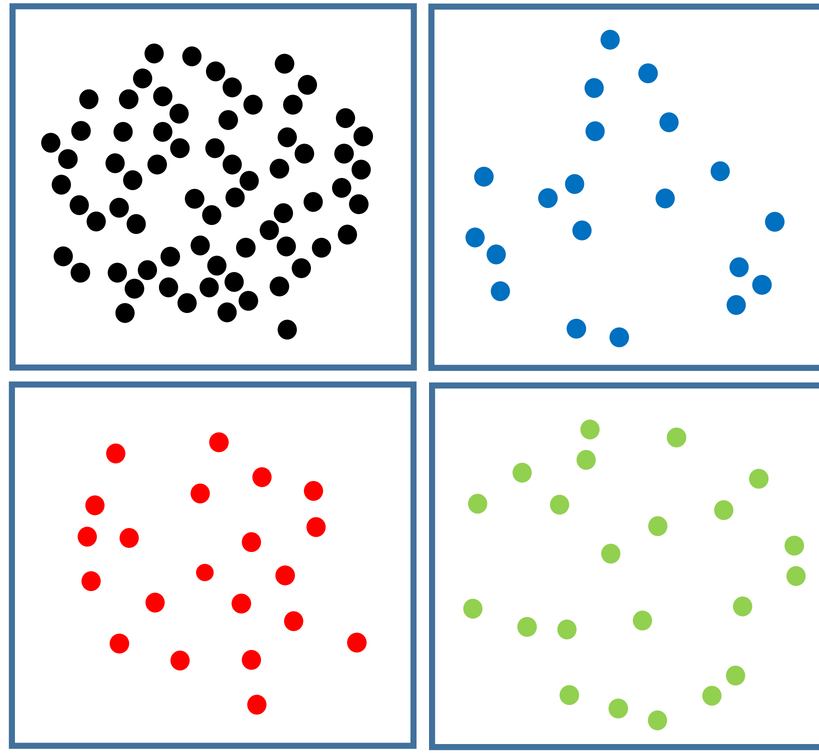


Zixuan Cang
(Cang and Wei, Bioinform. 2017)

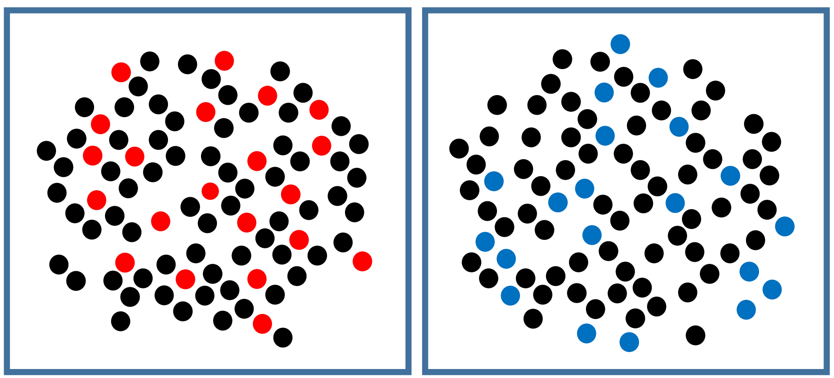
Original data

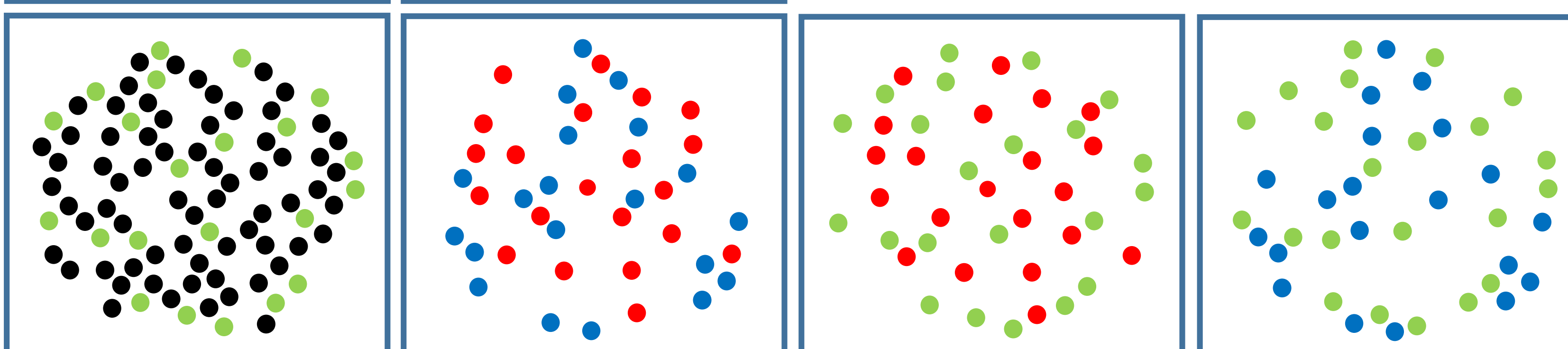


Single element selections



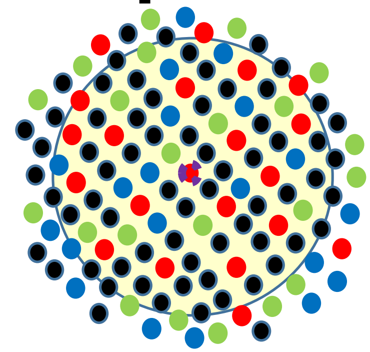
Two element selections





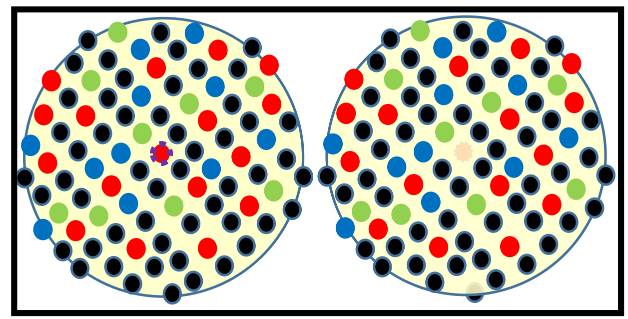
Atom-specific and element-specific persistent homology

Original data. The atom of interest is at the center.

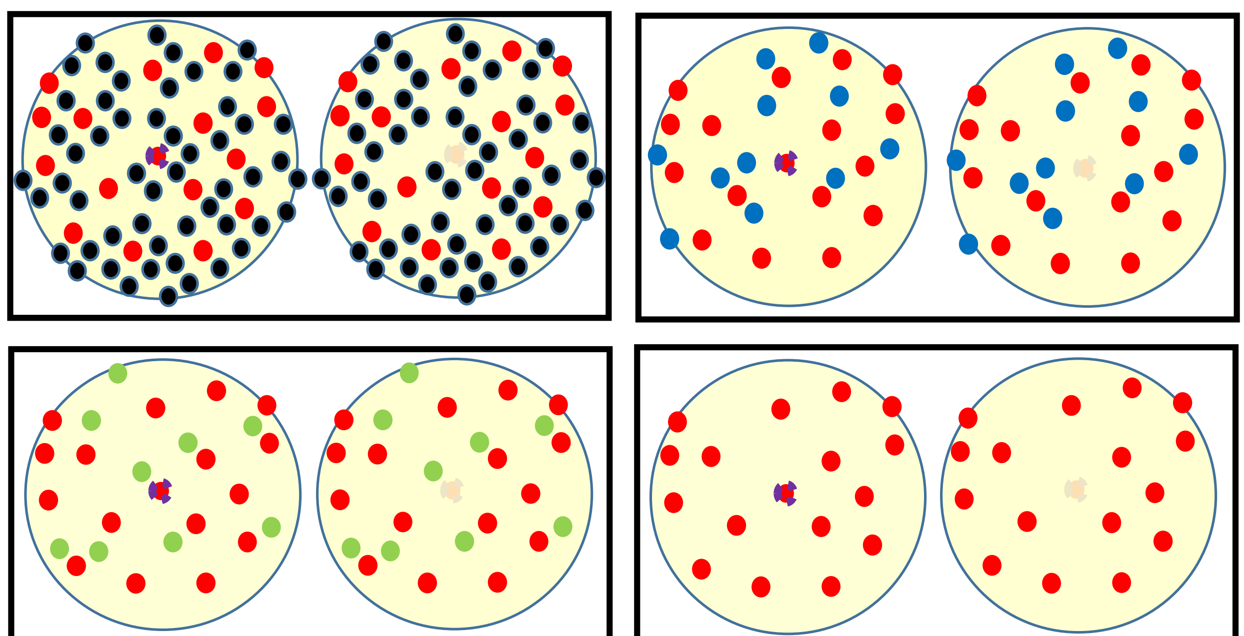


Dave Bramer
Bramer, Wei, 2018

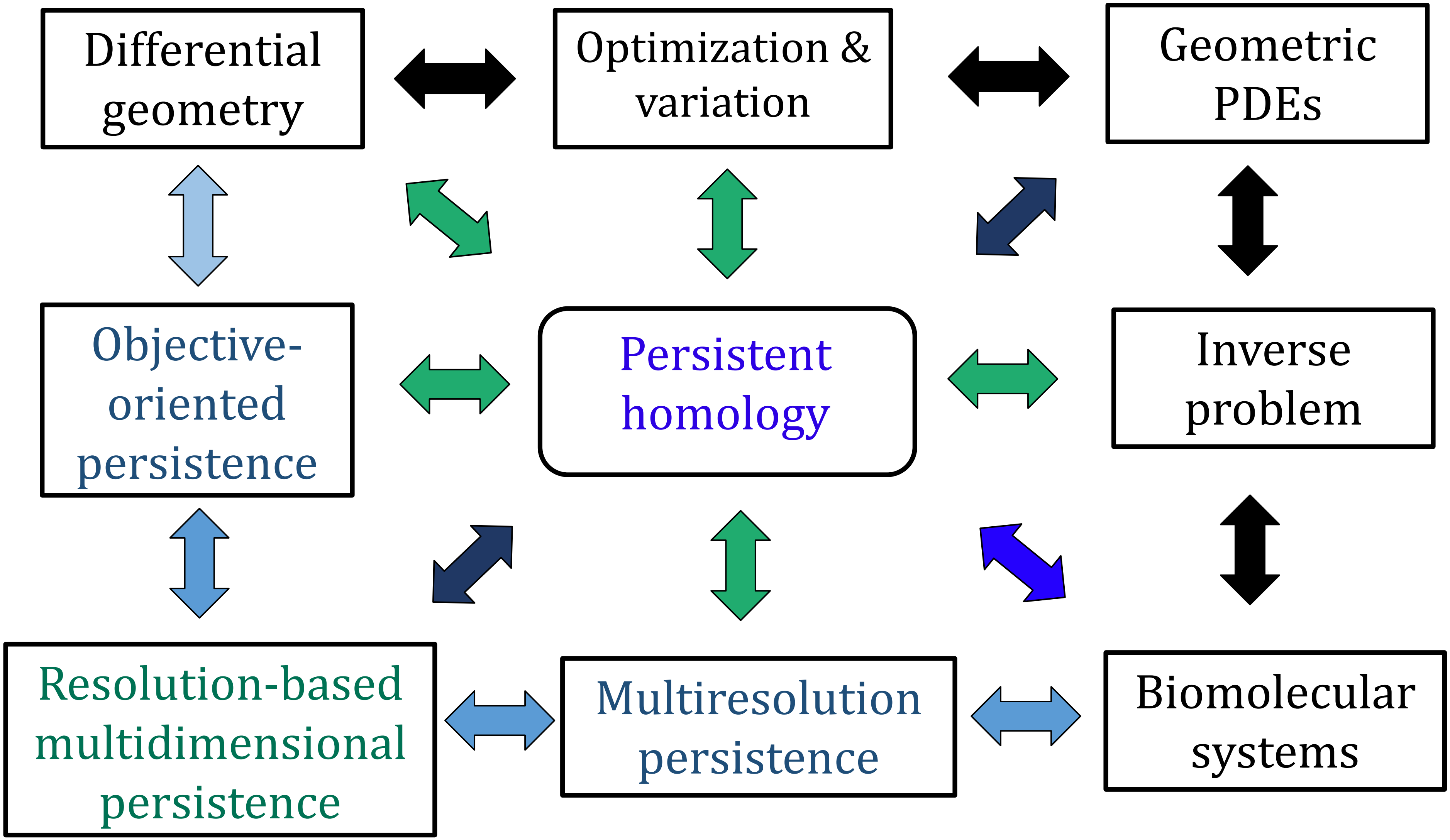
A pair of conjugated atom-specific datasets



Four pairs of conjugated atom-specific and element-specific datasets



Conclusion (connectivity)



Further topics and future directions

- Knot theory based modeling and analysis of DNA/RNA.
- Knot theory analysis of Hi-C data.
- Mathematical analysis of virus capsid assembly.
- Morse theory analysis of biomolecules.
- Reeb graph analysis of biomolecules.
- Conley index analysis of biomolecules.
- Evolutionary homology analysis of biomolecules.
- Atom specific topological modeling of atomic interactions and enzymatic processes.
- Topological representation of protein-protein interactions.
- Topological simplification of protein complexes, molecular machines, and subcellular organelles.
- Multicomponent persistent homology, multi-level persistent homology, and electrostatic persistence of biomolecules.
- Spectral sequence based modeling of biomolecules.
- Biology inspired new topological models for better understanding biology.


thank you