Lecture 2: Differential Geometry Based Biomolecular Surface Modeling

Guowei Wei

Mathematics

Michigan State University

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

NSF-CBMS Conference on Mathematical Molecular Bioscience and Biophysics

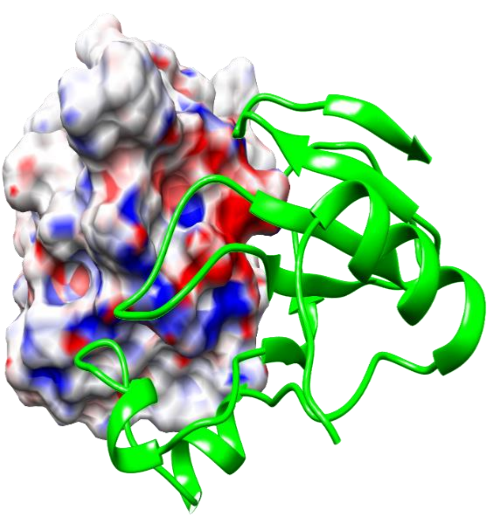
University of Alabama

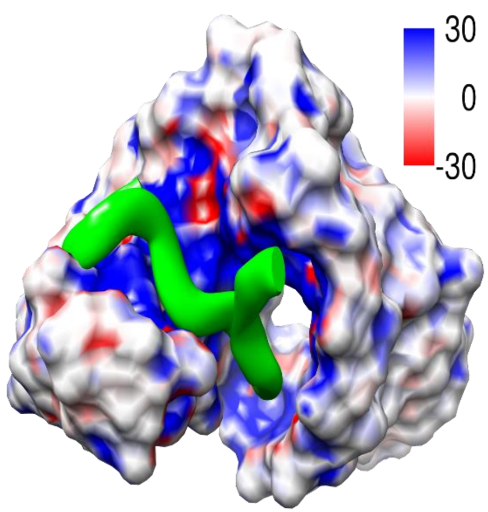
Tuscaloosa, May, 13-17, 2019

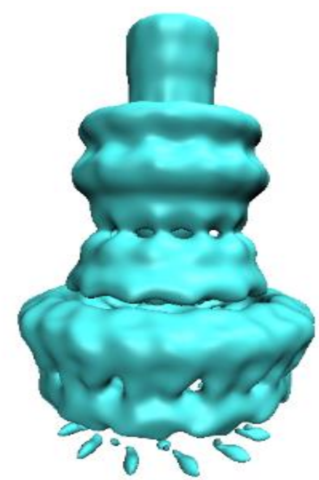
Grant support: NSF, NIH, MSU, BMS, and Pfizer

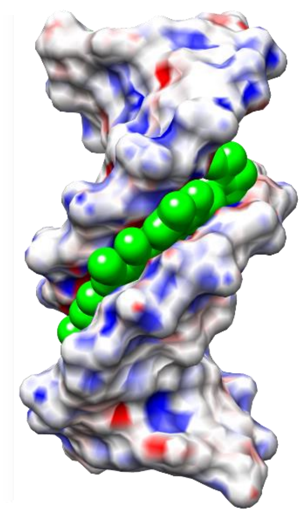
Motivations

How to define a biomolecular surface?
How to characterize the shape of proteins?
How to compute the surface area of protein?
How to minimize free energies on protein surface?
How to compute electrostatic forces on a biomolecular surface?
How to describe the shape matching and electrostatic matching of protein-protein, protein-ligand, and protein-nucleic acid interactions?
How to detect protein binding pockets and hot spots?

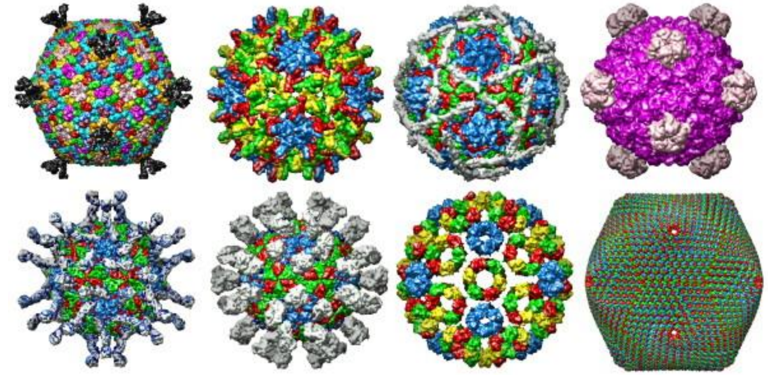




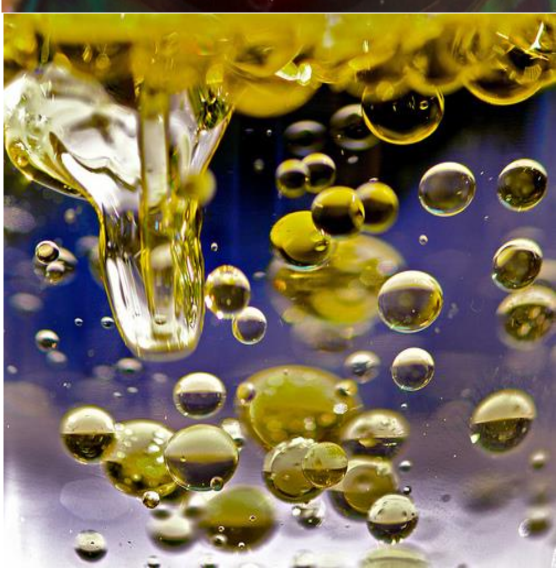


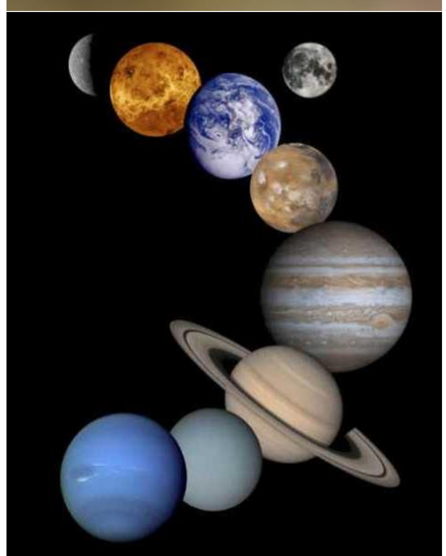


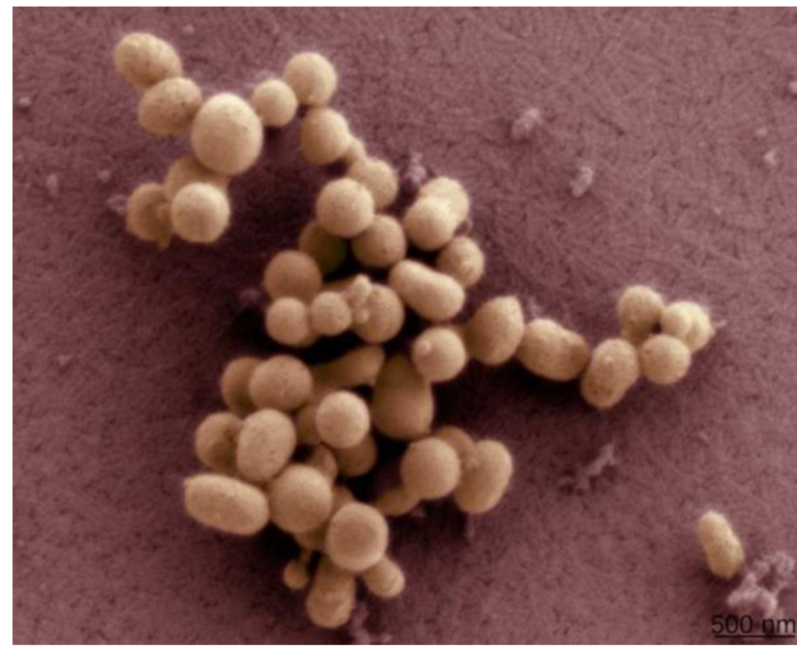
Minimal Surfaces
A way to minimize energy and maximize stability



Viral morphology







The first man-made life, Bacterium, M. mycoides, based on computer information

Brief History of Differential Geometry

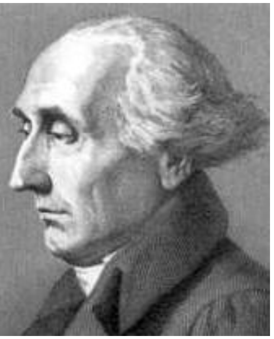
Leonhard Paul Euler (Swiss Mathematician, April 15, 1707 - Sept 18 1783) Developed the general idea of natural equations for obtaining curves from local curvature in 1736.

Gaspard Monge, Comte de Péluse (French mathematician, May 9, 1746-July 28, 1818) Established the general theory of curves of curvature of a surface between 1770-1790.

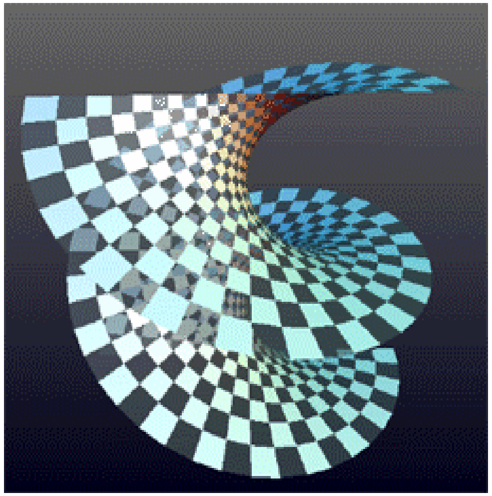
Jean Baptiste Meusnier (French, June 19, 1754-June 13, 1793) Studied the curvature of surfaces and discovered helicoid.

Johann Carl Friedrich Gauss (German mathematician, April 30, 1777 - February 23, 1855) Defined Gaussian curvature and its enbedding.

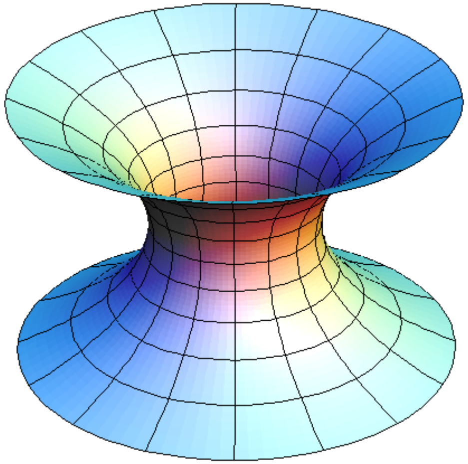
Brief History of Differential Geometry



Joseph-Louis Lagrange (Italian Mathematician, January 25 1736 – April 10, 1813) Euler-Lagrange variation for surface minimization.



Helicoid



Catenoid

Differential geometry of surfaces

An immersion $\boldsymbol{f}: U \to \mathbb{R}^{n+1}$ with $U \subset \mathbb{R}^n$ an open set and $\boldsymbol{f} \in C^2$.

Point vector on a hypersurface:

$$\boldsymbol{f}(\boldsymbol{u}) = \big(f_1(\boldsymbol{u}), f_2(\boldsymbol{u}), \dots, f_{n+1}(\boldsymbol{u})\big) = \boldsymbol{X}$$

$$\boldsymbol{u} = (u_1, u_2, \dots, u_n) \in U$$

Tangent vectors: $\boldsymbol{X}_i := \frac{\partial f}{\partial u_i} \in \mathrm{T}_{\boldsymbol{f}}\, \mathbb{R}^{n+1}$, the tangent space of $\mathbb{R}^{n+1}$ at $\boldsymbol{X}$.

Tangent plane $T_{\boldsymbol{u}}\boldsymbol{f}$ and normal space $\perp_{\boldsymbol{u}} \boldsymbol{f}$: $\perp_{\boldsymbol{u}} \boldsymbol{f} \oplus T_{\boldsymbol{u}}\boldsymbol{f} = \mathrm{T}_{\boldsymbol{f}(\boldsymbol{u})}\, \mathbb{R}^{n+1}$

Jacobian matrix: $D\boldsymbol{f} = (\boldsymbol{X}_1, \boldsymbol{X}_2, \dots, \boldsymbol{X}_n)$

Euclidean inner product in $\mathbb{R}^{n+1}$:

$$I\big(\boldsymbol{X}_i, \boldsymbol{X}_j\big) \coloneqq \big\langle \boldsymbol{X}_i, \boldsymbol{X}_j \big\rangle \quad \forall \boldsymbol{X}_i, \boldsymbol{X}_j \in \mathrm{T}_{\boldsymbol{u}}\, \boldsymbol{f}$$

The first fundamental form: $(g_{ij}) = \Big(I\big(\boldsymbol{X}_i, \boldsymbol{X}_j\big)\Big) = (D\boldsymbol{f})^T \circ (D\boldsymbol{f})$

Gram determinant: $g = \mathrm{Det}(g_{ij})$

Gauss map: $\boldsymbol{N}(\boldsymbol{u})$: $U \to \mathbb{R}^{n+1}$, an n-linear vector product to give a unit normal: $\boldsymbol{N}(u_1, u_2, \dots, u_n) =: \frac{\boldsymbol{X}_1 \times \boldsymbol{X}_2 \times \cdots \times \boldsymbol{X}_n}{\|\boldsymbol{X}_1 \times \boldsymbol{X}_2 \times \cdots \times \boldsymbol{X}_n\|} \in \perp_{\boldsymbol{u}} \boldsymbol{f}$

Differential geometry of surfaces

The second fundamental form (Hessian matrix):

$$II(\boldsymbol{X}_i, \boldsymbol{X}_j) = (h_{ij})_{i,j=1,2,\ldots,n} = \left(\left\langle -\frac{\partial \boldsymbol{N}}{\partial u_i}, \boldsymbol{X}_j \right\rangle\right)_{ij} = \left(\left\langle \boldsymbol{N}, \frac{\partial^2 \boldsymbol{f}}{\partial u_i \partial u_j} \right\rangle\right)_{ij}$$

The Weingarten map: $L = -D\boldsymbol{N} \circ (D\boldsymbol{f})^{-1}$

Mean curvature: $H = \frac{1}{n}\mathrm{Tr}(L) = \frac{1}{n} h_{ij} g^{ji} = \frac{1}{n}(\kappa_1 + \kappa_2 + \cdots + \kappa_n)$

with $g^{ji} = \left((g_{ij})^{-1}\right)_{ij}$ and

κ_i are egenvalues of L (i.e., principal curvatures)

Gauss curvature: $K = \mathrm{Det}(L) = \kappa_1 \cdot \kappa_2$ (2D surface)

For biomolecular surfaces $n = 2$:

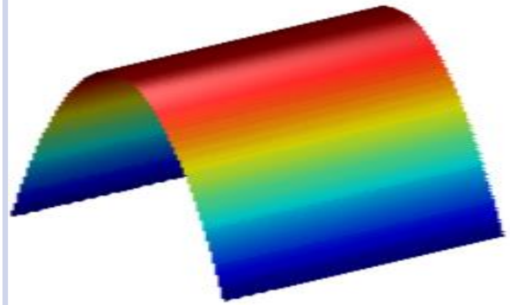
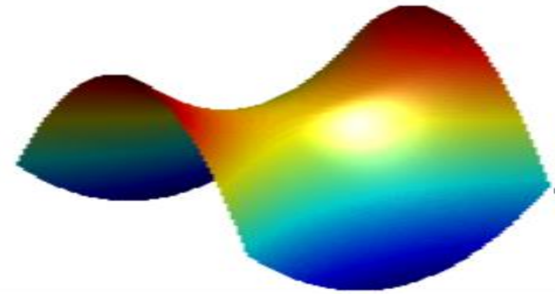
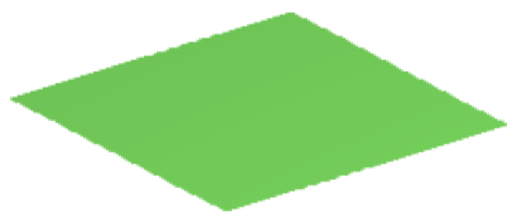
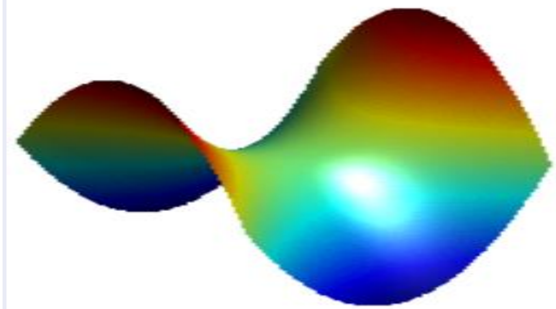
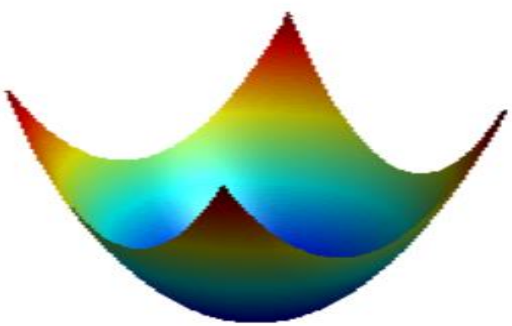
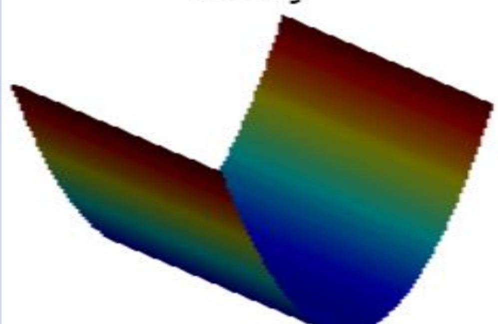
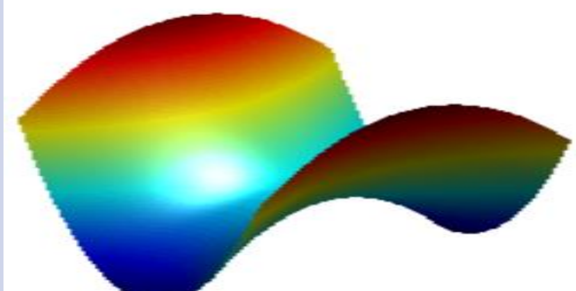
$K > 0$: Elliptic

$K < 0$: Hyperbolic

$K = 0$ and $H \neq 0$: Parabolic

$\kappa_1 = \kappa_2$: Umbilic

Surface parches characterized by the Gauss and mean

	K>0	*K=0*	*K<0*
H>0	Peak 	Ridge 	Saddle Ridge 
H=0	None	Flat 	Minimal Surface 
H<0	Pit 	Valley 	Saddle Valley 

Differential geometry of surfaces

For a special construction, consider a hypersurface function $S(x, y, z)$

$$\boldsymbol{f}(\boldsymbol{u}) = \big(x, y, z, S(x, y, z)\big)$$

$$(g_{ij}) = \begin{pmatrix} 1+S_x^2 & S_xS_y & S_xS_z \\ S_xS_y & 1+S_y^2 & S_yS_z \\ S_xS_z & S_yS_z & 1+S_z^2 \end{pmatrix}$$

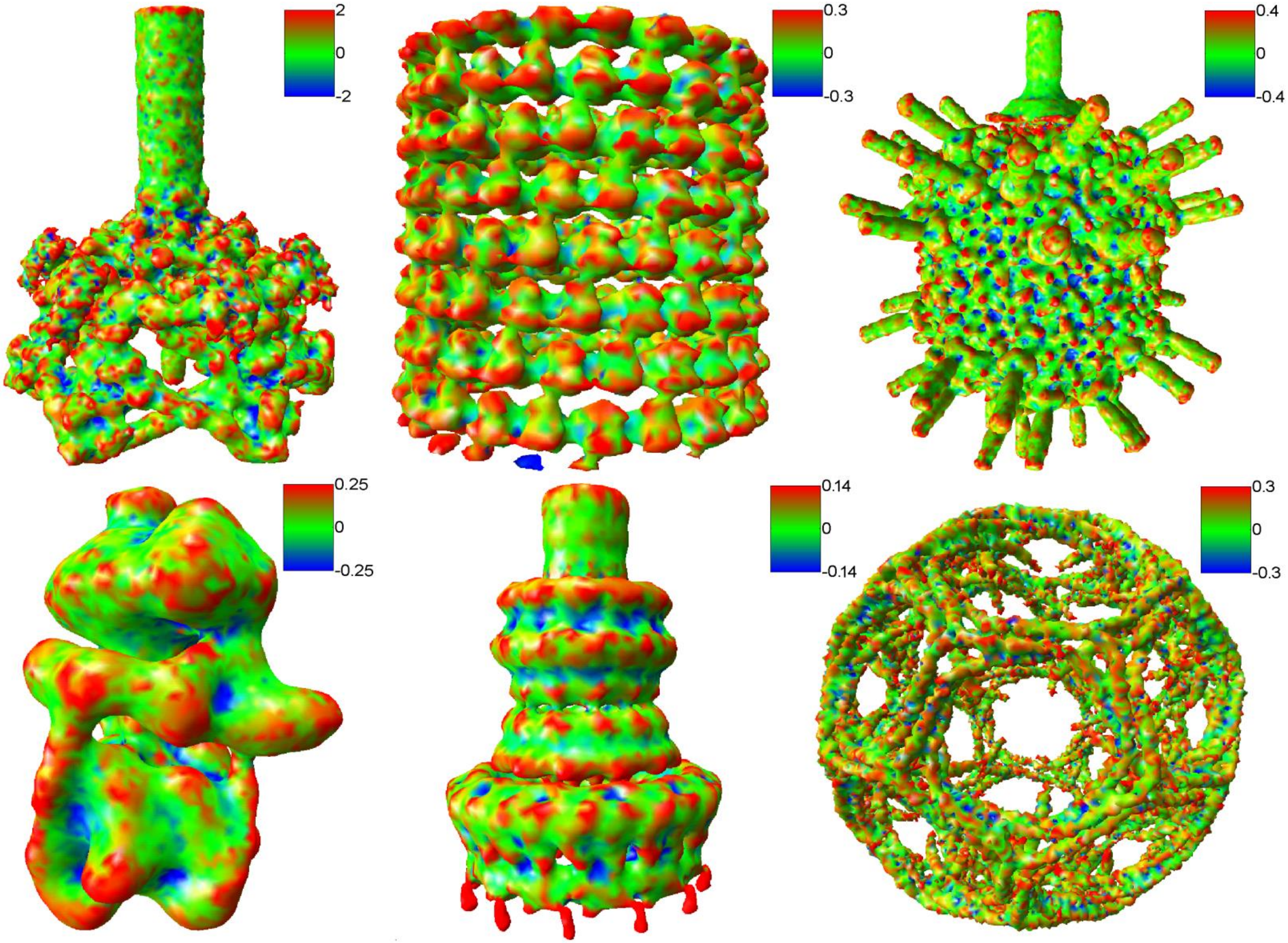
The Gram determinant: $g = \mathrm{Det}(g_{ij}) = 1 + S_x^2 + S_y^2 + S_z^2$

Gauss map: $\boldsymbol{N} = \frac{(-S_x, -S_y, -S_z, 1)}{\sqrt{g}}$

The second fundamental form: $(h_{ij}) = \left(\frac{1}{\sqrt{g}} S_{x_i} S_{x_j}\right)$

Mean curvature: $H = \frac{1}{3} \nabla \cdot \left(\frac{\nabla S}{\sqrt{g}}\right)$

Gaussian curvature of subcellular organelles

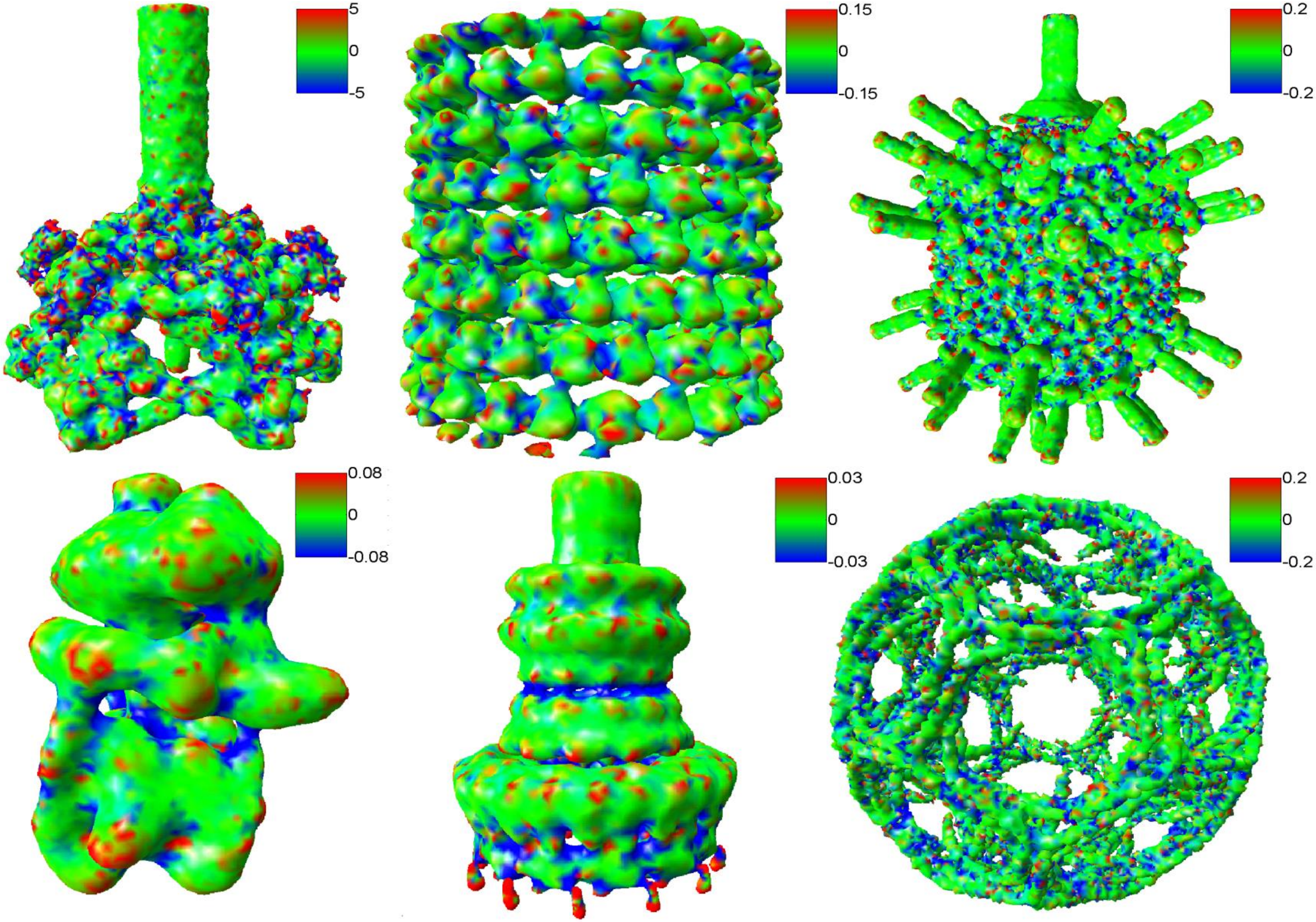


Kelin Xia

Xin Feng

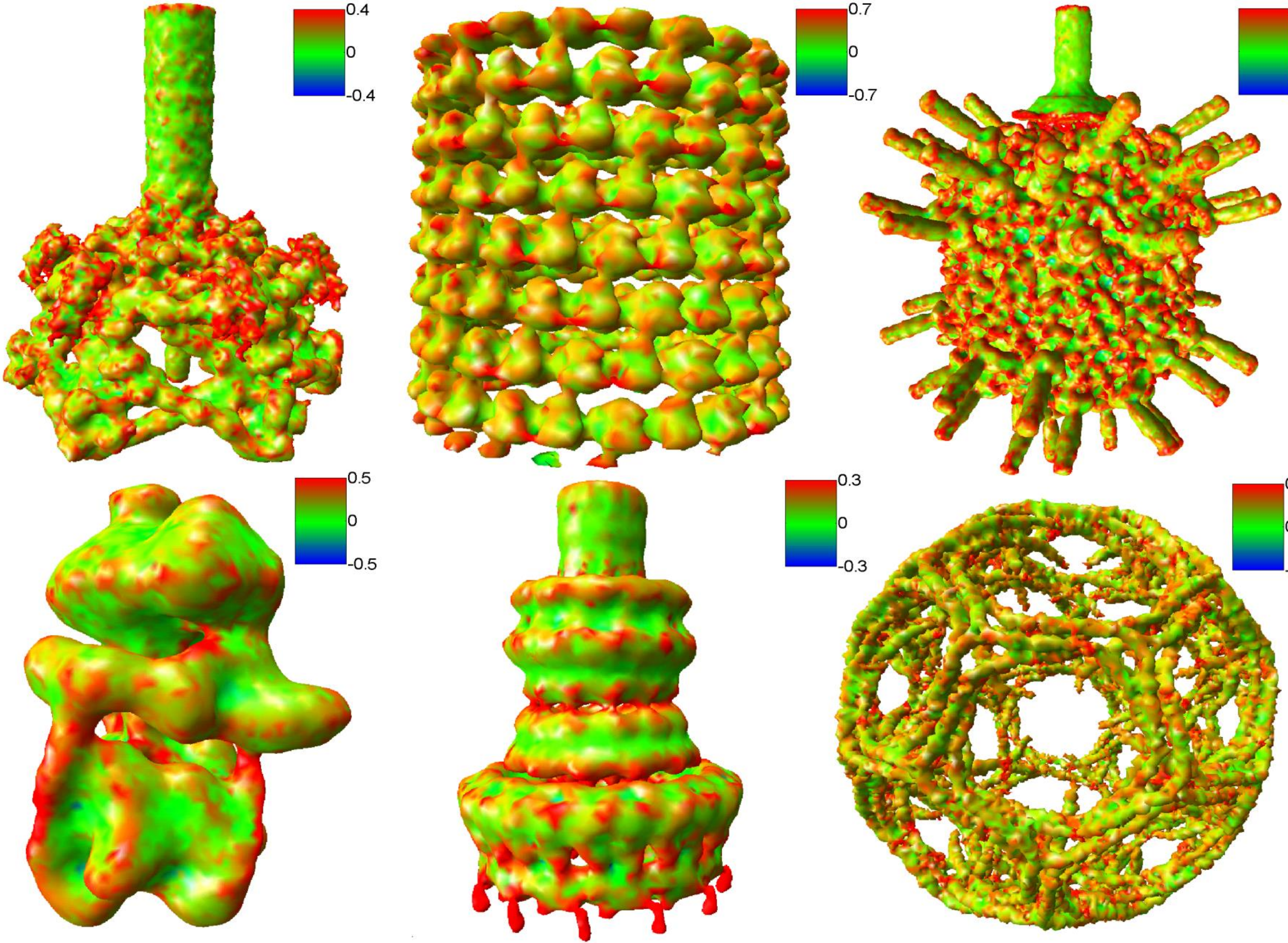
From upper left to lower right: EMD1048, T4 baseplate; EMD1129, GDP-tubulin; EMD1265, bacteriophage 29; EMD1590 ; EMD1617; EMD5119, clathrin lattice.

Mean curvature of subcellular organelles

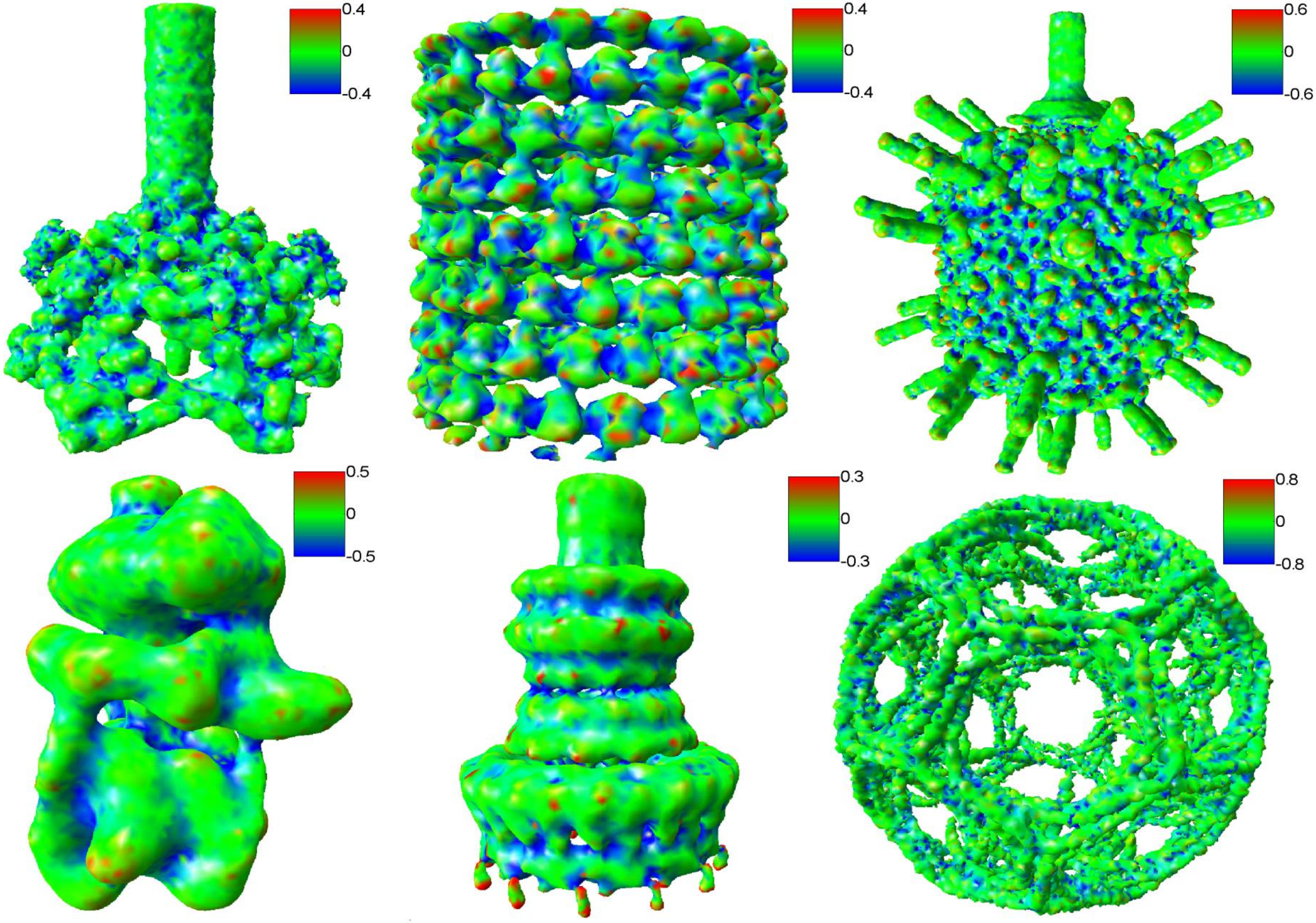


(Feng, Xia, Tong and Wei, IJNMBI,2012)

Maximum curvature (k1)

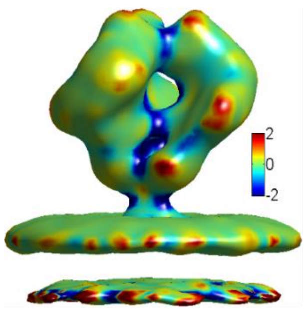


Minimum curvature (k2)

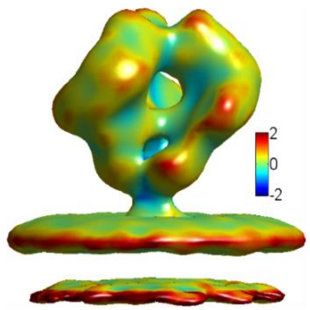


Surface characterization of an HIV viral receptor

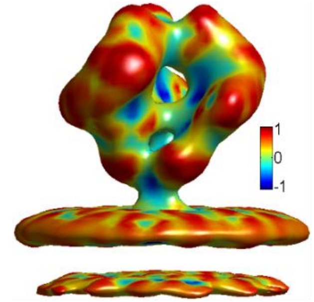
Gauss



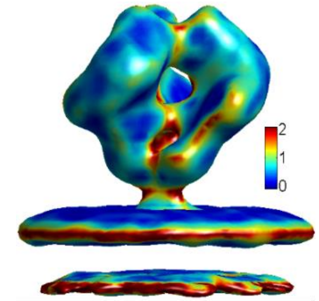
Mean



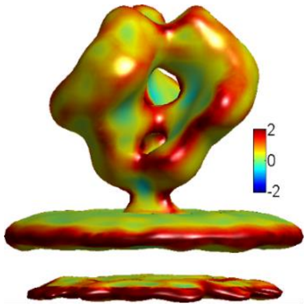
Shape index



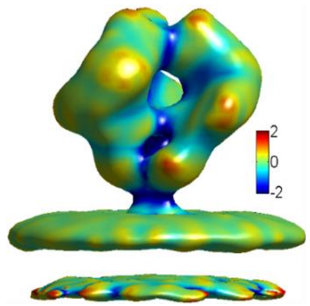
Curvedness



Minimum



Maximum



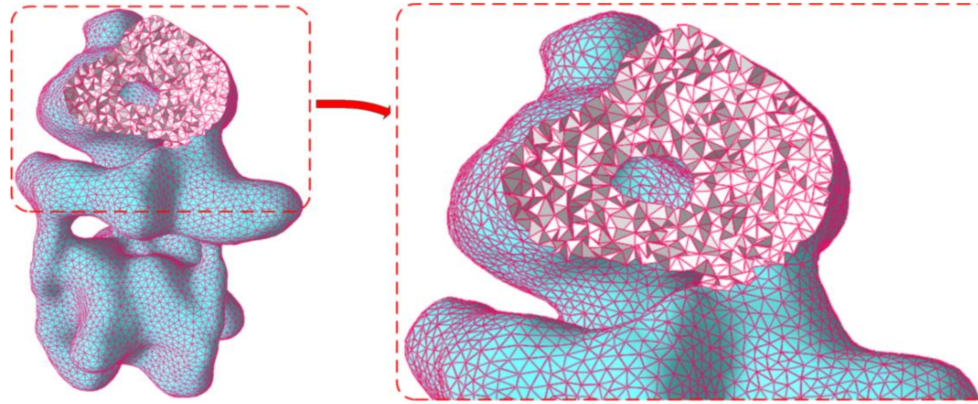
Kelin Xia

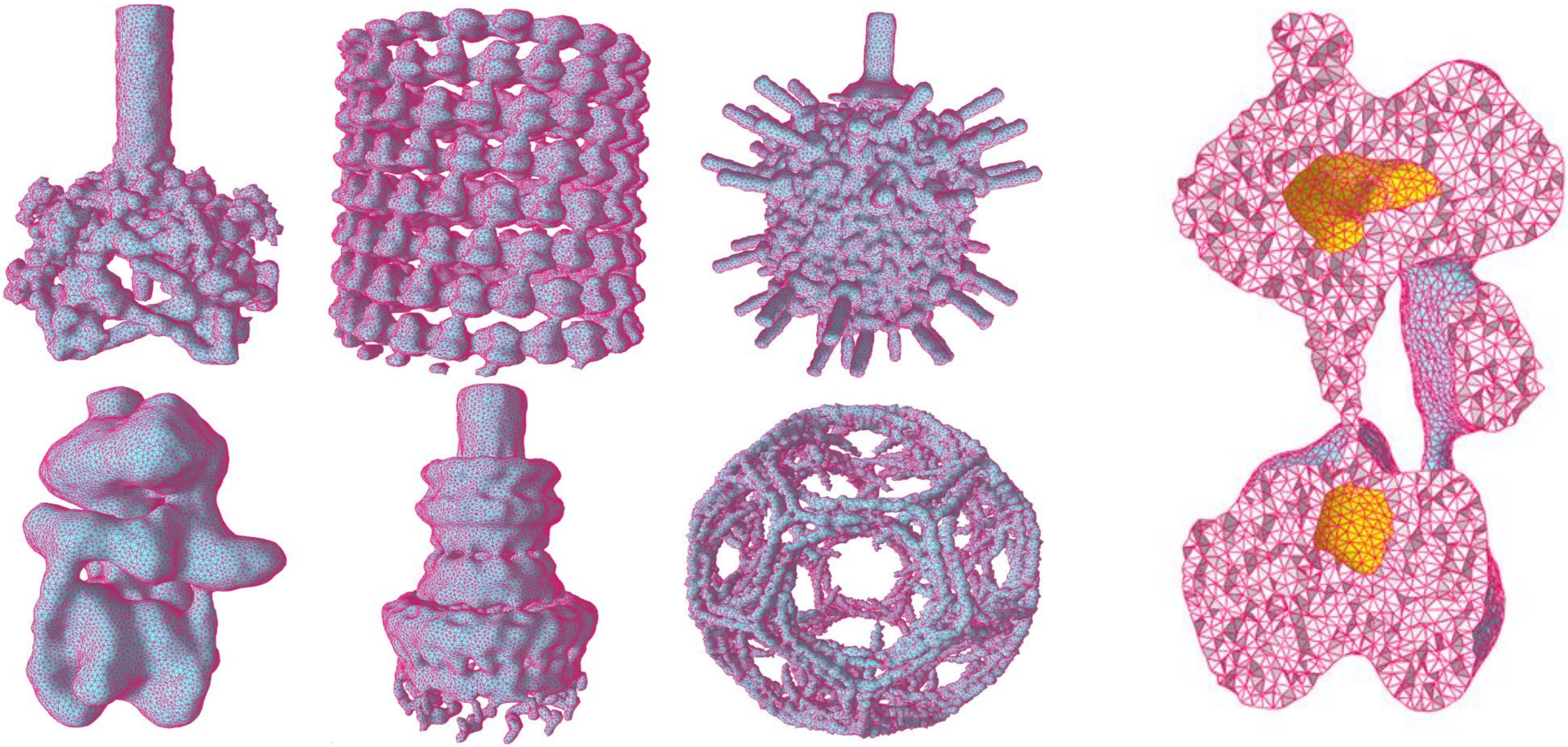
(Xia, Feng, Tong and Wei, JCP, IJNMBI,2012)

Geometric modeling

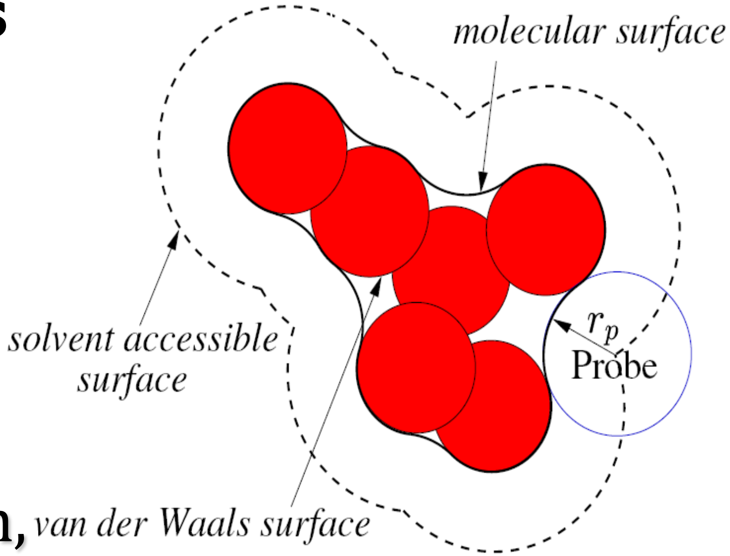
Curvature based meshing for subcellular structures and organelles

(Feng, Xia, Tong and Wei, IJNMBI,2012)

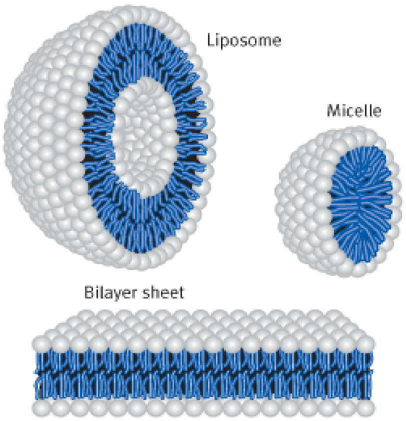




Biomolecular Surface Models



- Concepts (Richards and Lee 1971, Richards, 1977)
- Analytical methods (Connolly, 1983) (msroll)
- Reduced representations (Sanner, Olson, and Spehner, 1996; Rocchia et al, 2002,…)
- Gaussian surface (Blinn, 1982; Grant & Pickup, 1995 Alexov et al, 2013,)
- Other methods (Greer; Bush; Duncan; Richmond; Edelsbrunner; Lu;…)



Difficulties:

- Inconsistent with free energy minimization
- Cusp and self-intersecting singularities

Phase field models for biomembrane:
Chun Liu, Q. Du, …

Higher order curvature driving flow:

$$\frac{\partial S}{\partial t} = (-1)^n |\nabla S| \, \nabla \cdot \left(\frac{\nabla^{2n} \nabla S}{|\nabla^{2n} \nabla S|} \right) + V, \qquad n = 0,1,2, \ldots$$

Models: Wei, 1999; You and Kaveh, 2000; Chan et al, 2001;...
Analysis: Bertozzi and Greer, 2004; Xu and Zhou, 2006;...)
Applications: Many

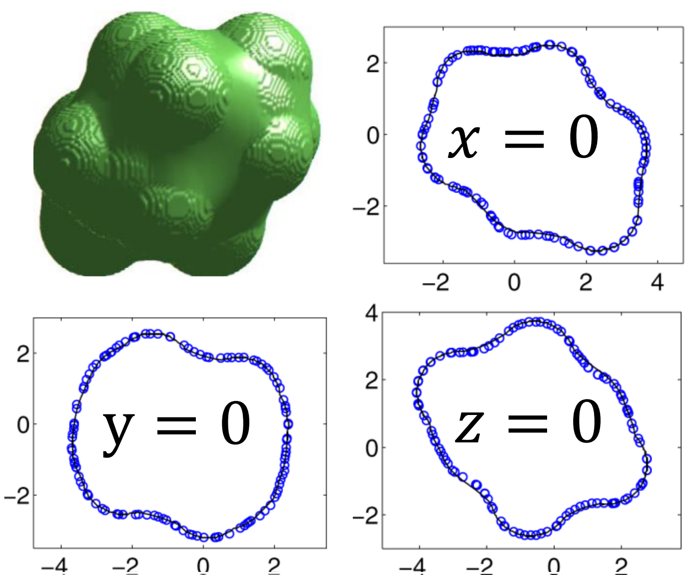
Numerical methods (Shan Zhao, 2009,2010):

Implicit Euler, CN, ADI, Euler, SOR, CG,...

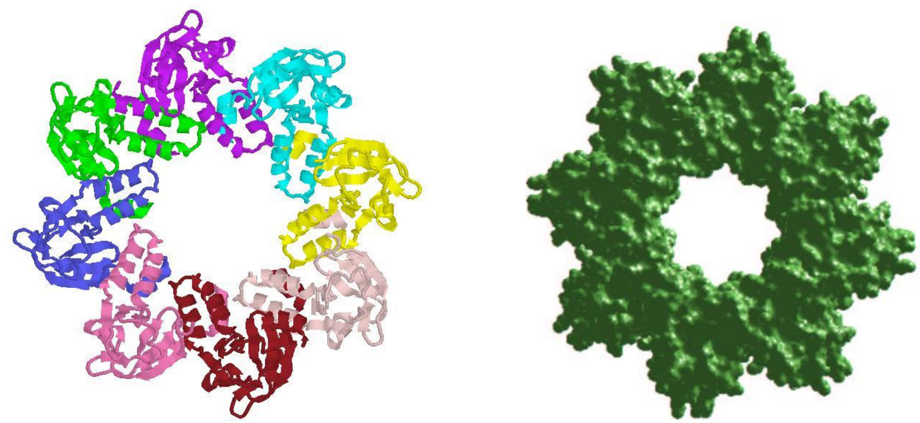
The first PDE based molecular surface modeling

$$\frac{\partial u}{\partial t} = \nabla \cdot [d(|\nabla u|)\nabla\nabla^2 u] + V$$

(Wei, IEEE Signal Proc. Lett., 1999)



cyclohexane (C_6H_{12})



The cell division protein, PDB ID: 1N0E, 9245 atoms in 1328 residues

Major feature: Starting from atomic information, instead of a given surface

(Wei, Sun, Zhou and Feig, 2005)

Yuhui Sun

Minimal molecular surface

Shan Zhao

Surface free energy functional E:

$$E = \int_U \gamma\sqrt{g}\, dxdydz = \int_U e\, dxdydz$$

where U encloses the molecule, γ is the surface tension and $g = 1 + S_x^2 + S_y^2 + S_z^2$.

The Euler-Lagrange variation:

$$\frac{\partial e}{\partial S} - \frac{\partial}{\partial x}\frac{\partial e}{\partial S_x} - \frac{\partial}{\partial y}\frac{\partial e}{\partial S_y} - \frac{\partial}{\partial z}\frac{\partial e}{\partial S_z} = 0$$

Assume a homogeneous surface tension, we have the mean curvature: $3\gamma H = \gamma \nabla \cdot \left(\frac{\nabla S}{\sqrt{g}}\right) = 0$

The Laplace-Beltrami equation achieves the minimization:

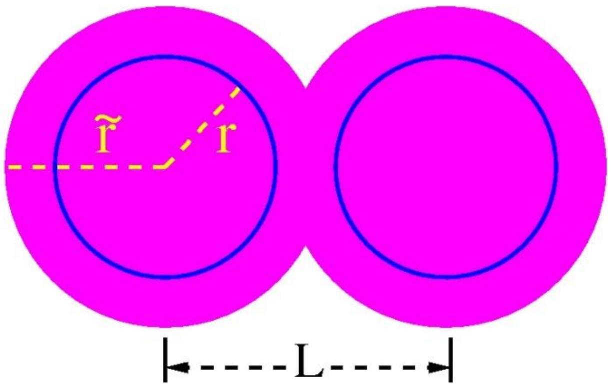
$$\frac{\partial S}{\partial t} = \sqrt{g}\nabla \cdot \left(\frac{\nabla S}{\sqrt{g}}\right) \Longrightarrow \frac{\partial S}{\partial t} = |\nabla S|\, \nabla \cdot \left(\frac{\nabla S}{|\nabla S|}\right)$$

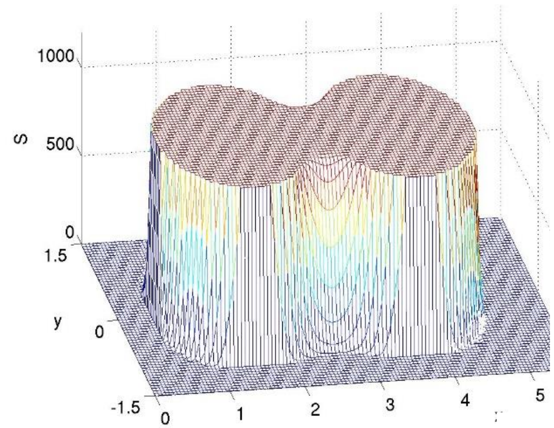
(Bates, Wei, and Zhao, 2016, JCC, 2008)

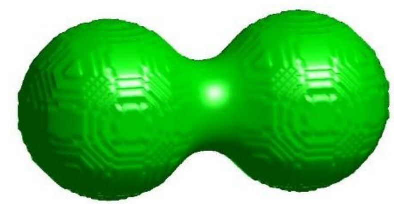
Minimal Molecular surface

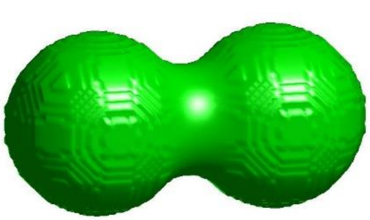
The first biomolecular surface ever constructed with the variational principle.

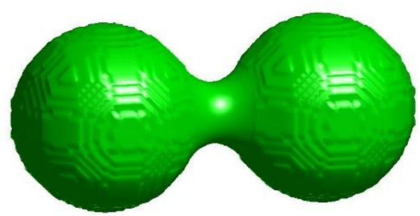
$$\frac{\partial S}{\partial t} = |\nabla S| \, \nabla \cdot \left(\frac{\nabla S}{|\nabla S|} \right)$$

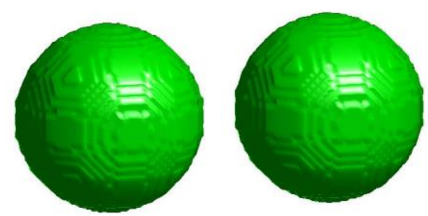












Atomic distance dependence

Shan Zhao

(Bates, Wei, and Zhao, 2016, JCC 2008)

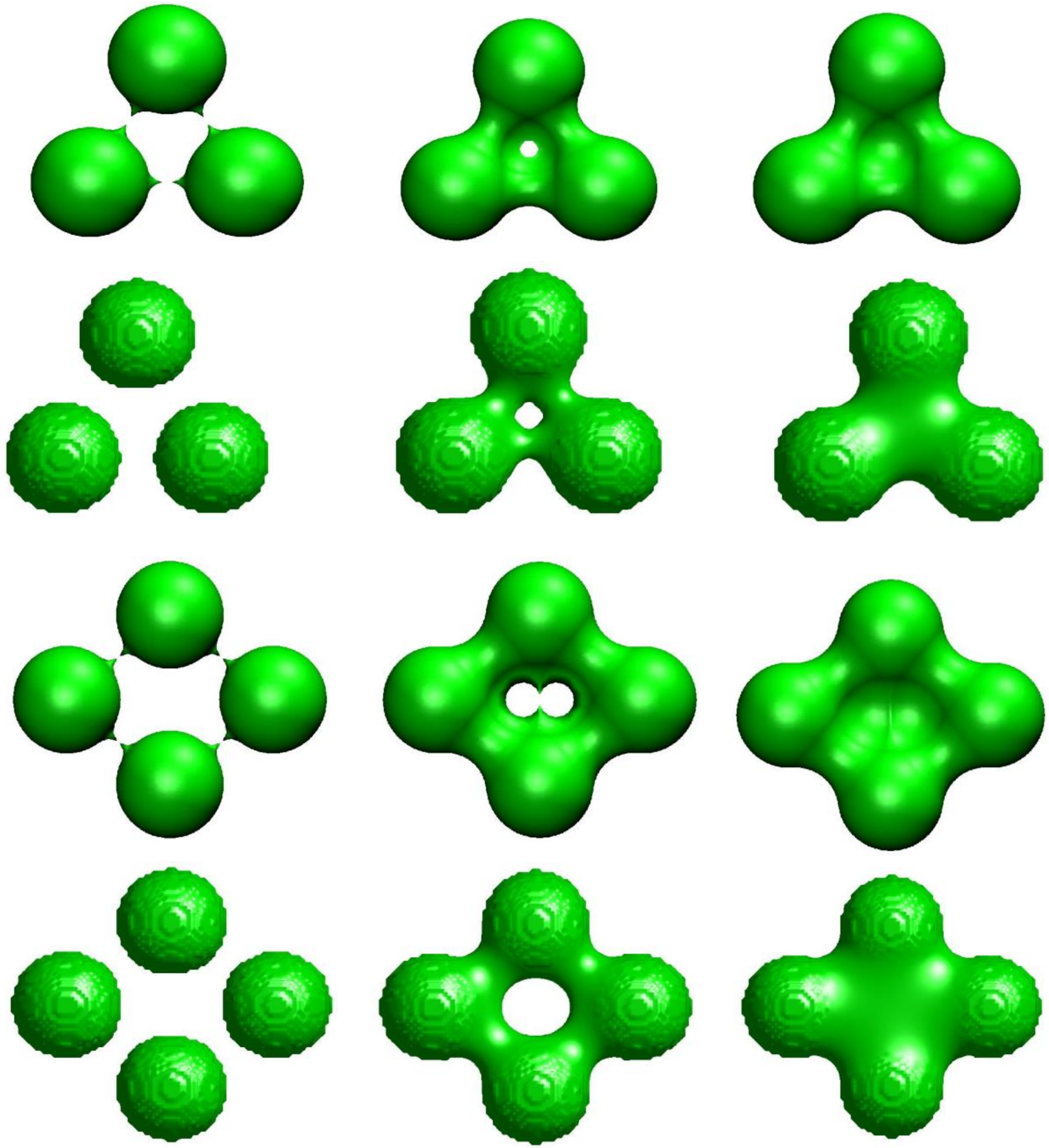
Surface singularities

MS:

MMS:

MS:

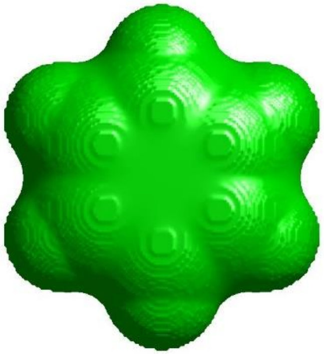
MMS:



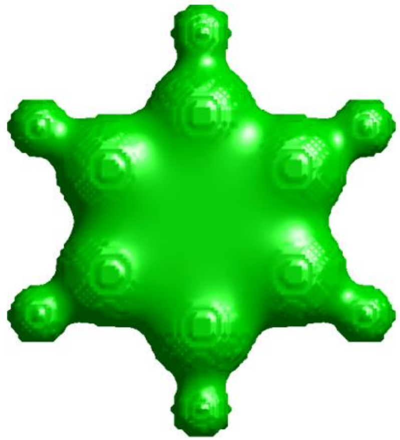
Shan Zhao

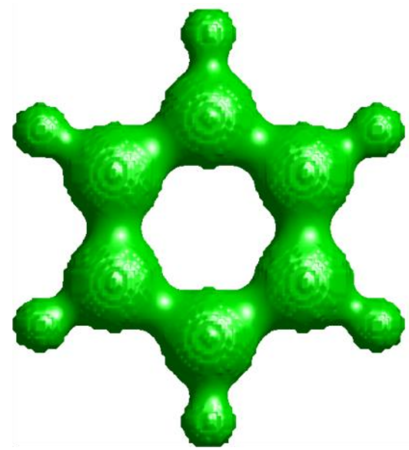
(Bates, Wei, and Zhao, 2016, JCC 2008)

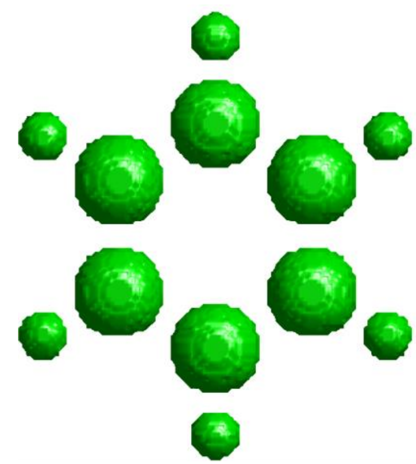
Minimal molecular surfaces of benzene at different atomic radii



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







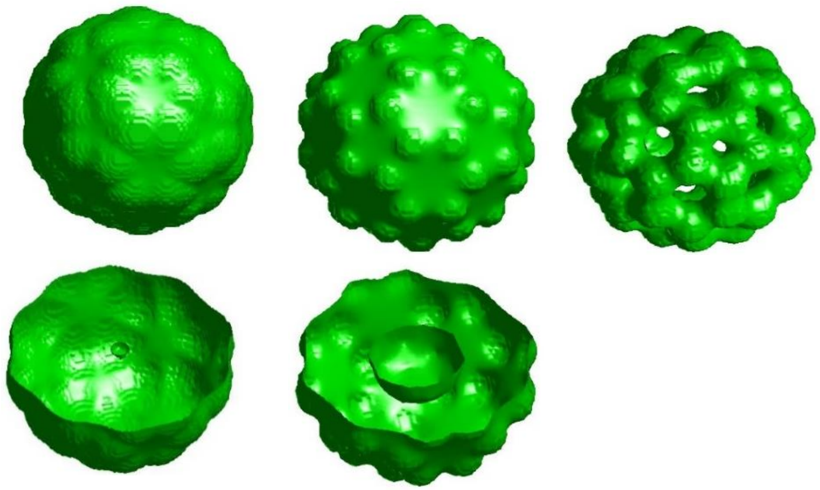
Shan Zhao

(Bates, Wei, and Zhao, 2016, JCC 2008)

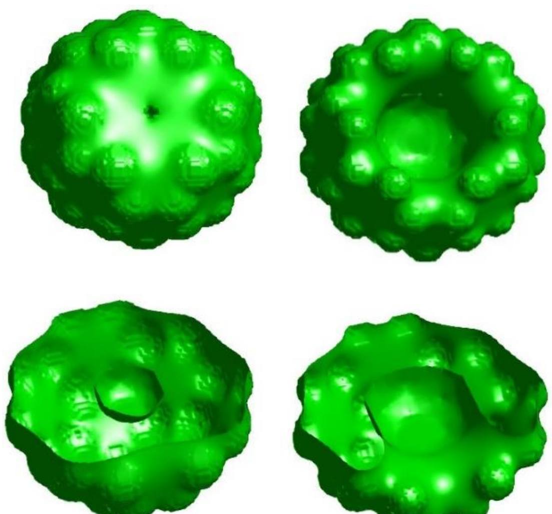
MMSs of nano-particles, DNA and protein

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

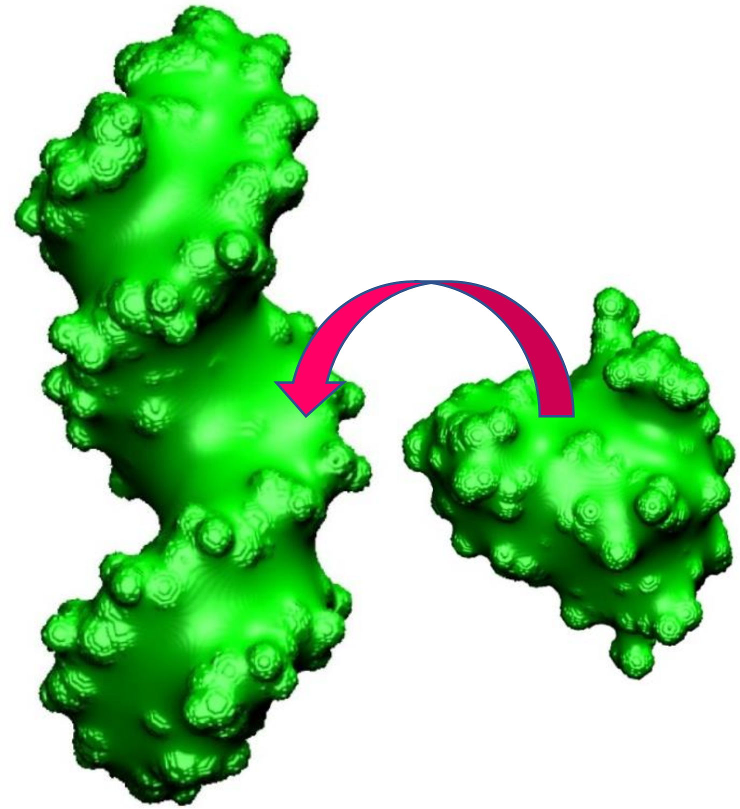
C_{60}



C_{54}



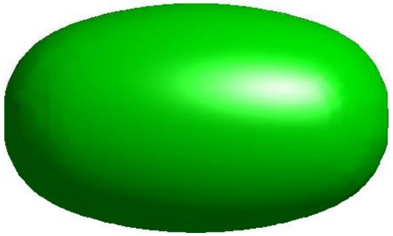
Shan Zhao



DNA-protein
molecular recognition

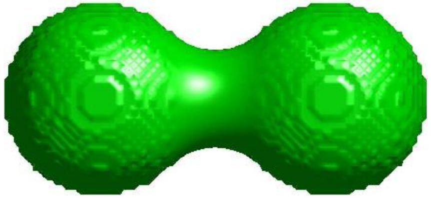
(Bates, Wei, and Zhao, 2016, JCC 2008)

Potential driving surface formation

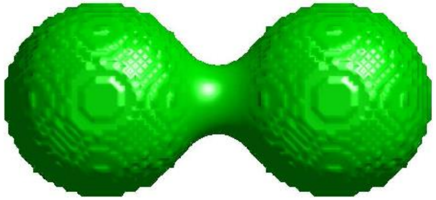


$V = 0.6$

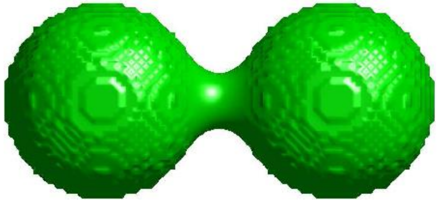
$$\frac{\partial S}{\partial t} = |\nabla S|\left[\nabla \cdot \left(\frac{\gamma \nabla S}{|\nabla S|}\right) + V\right]$$



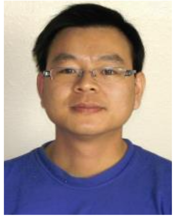
$V = 0.3$



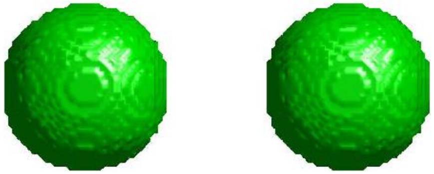
$V = 0$



$V = -0.3$



Zhan Chen



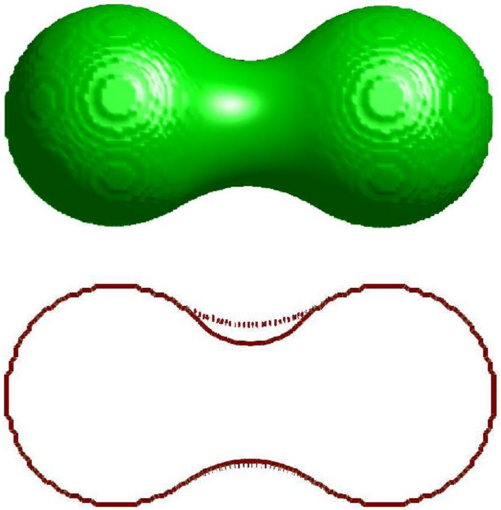
$V = -0.6$

(Bates, Chen, Sun, Wei & Zhao, J. Math. Biol. 2008)

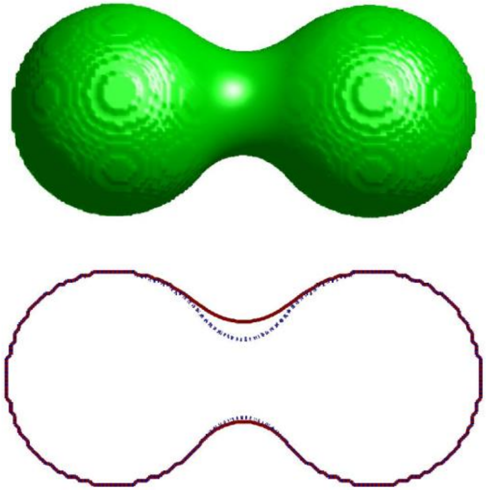
Potential driving geometric flows

Zhan Chen

Localized potentials



$$\frac{\partial S}{\partial t} = |\nabla S|\, \nabla \cdot \left(\frac{\gamma \nabla S}{|\nabla S|} \right) + V_0$$



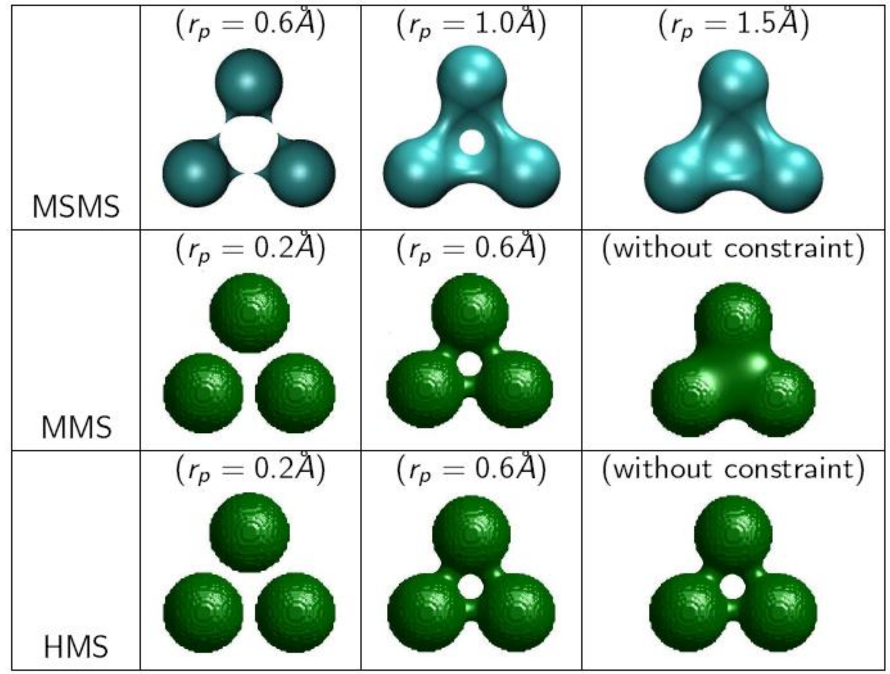
$$V_0 = -1 + 0.5 \left(\frac{4}{\sqrt{x^2 + (y-5.5)^2 + z^2}} \right)^7$$

$$V_0 = -0.5 \left(\frac{4}{\sqrt{x^2 + (y-5.5)^2 + z^2}} \right)^7$$

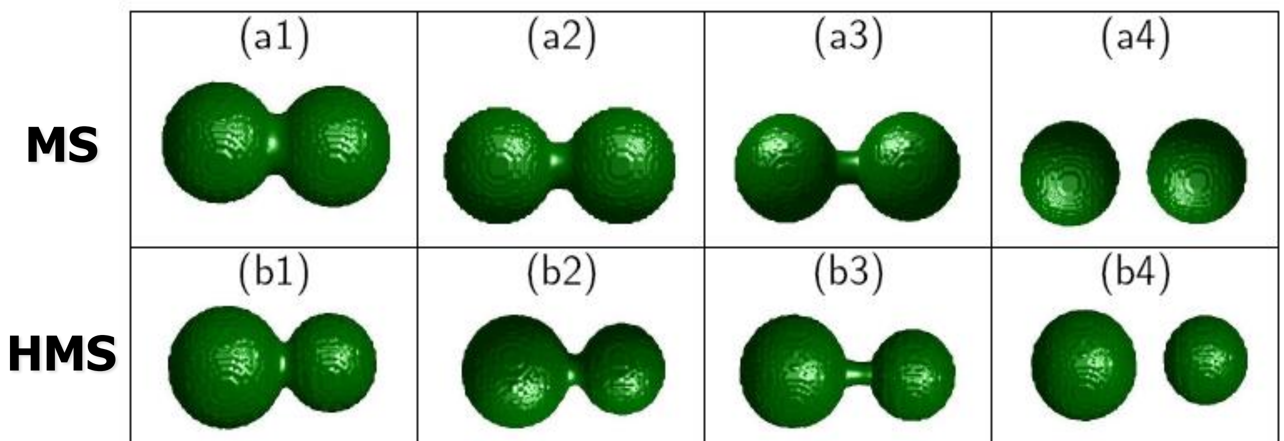
(Bates, Chen, Sun, Wei & Zhao, J. Math. Biol. 2008)

High-order molecular surface (HMS)

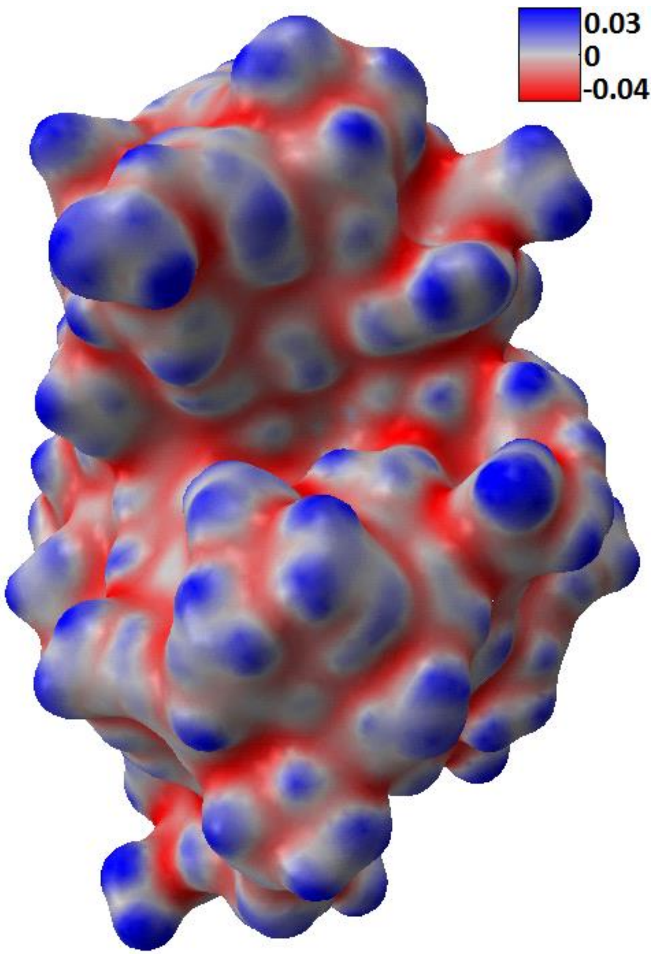
$$\frac{\partial S}{\partial t} = (-1)^n |\nabla S| \, \nabla \cdot \left(\frac{\nabla^{2n} \nabla S}{|\nabla^{2n} \nabla S|} \right) + V, \qquad n = 0,1,2,\dots$$

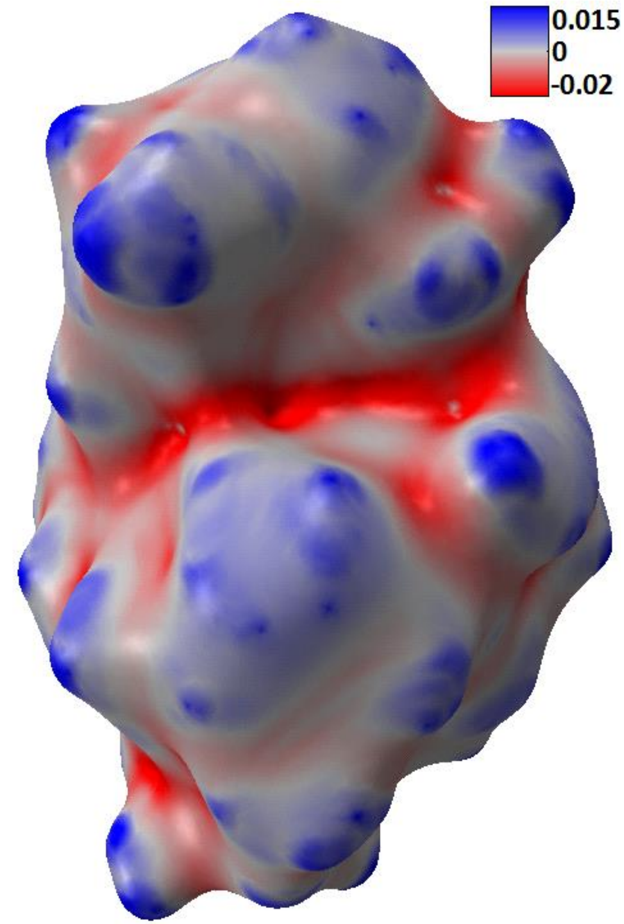


(Bates, Chen, Sun, Wei, 2008)



Multiscale minimum principle curvature on protein MMS to identify binding prockets



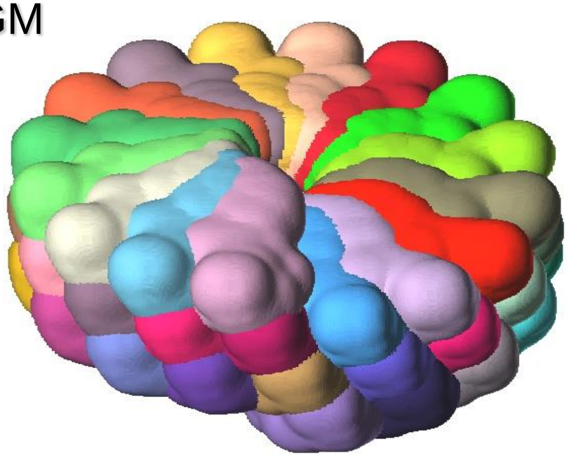


(Feng, Xia, Tong and Wei, JCC, 2013)

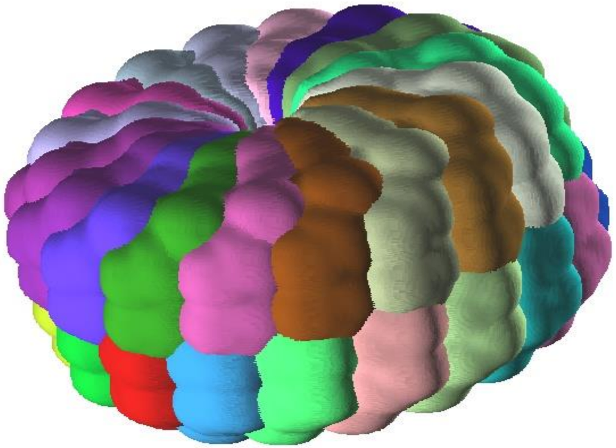
Virus surfaces

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

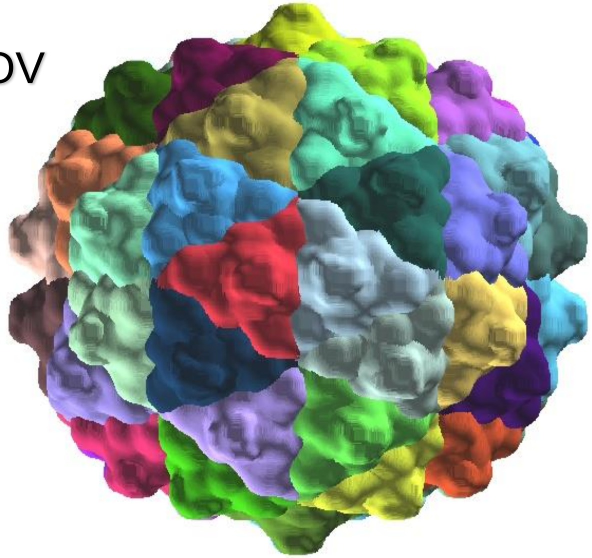
1CGM



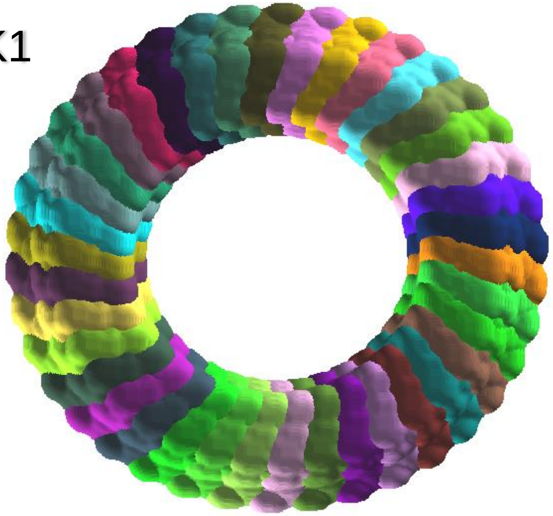
1EI7



1NOV

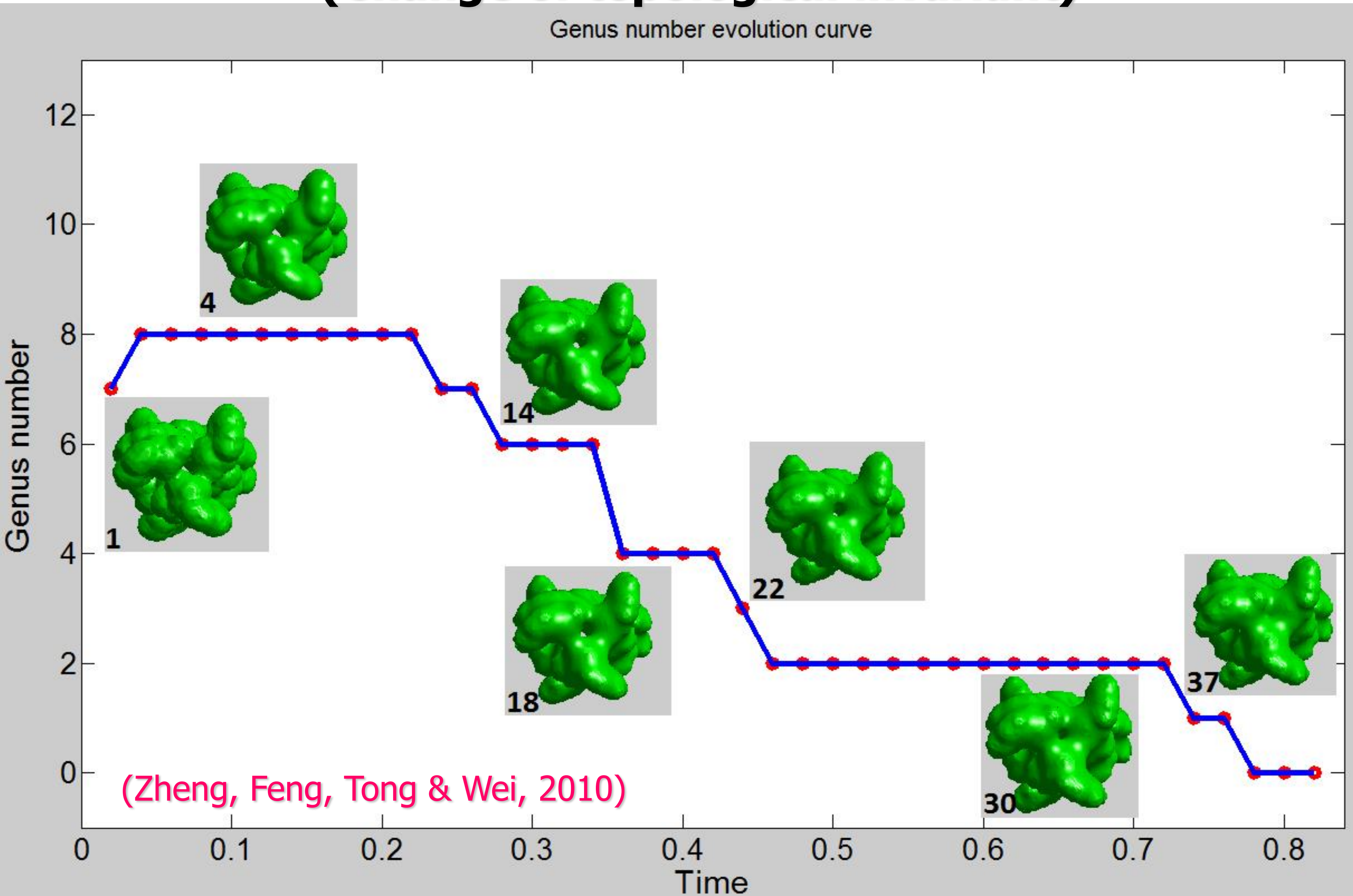


2BK1



(Chen, Saxena, Wei, IJBI, 2009)

Time evolution of genus number (Change of topological invariant)



Cryo-EM map noise reduction

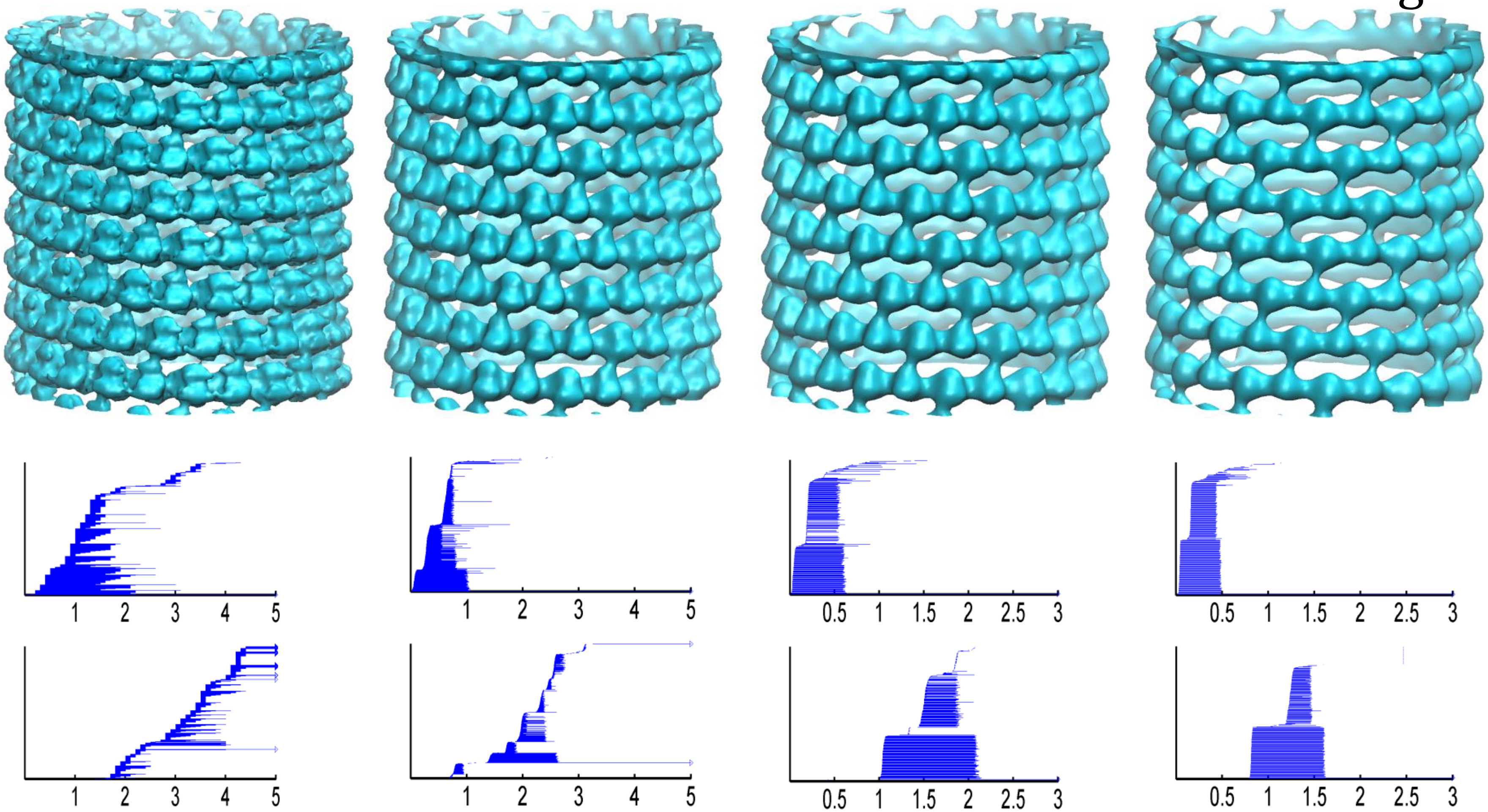
$$\frac{\partial S}{\partial t} = |\nabla S|\, \nabla \cdot \left(\frac{\gamma \nabla S}{|\nabla S|} \right)$$

Original data

Ten-iteration denoising

Twenty-iteration denoising

Forty-iteration denoising



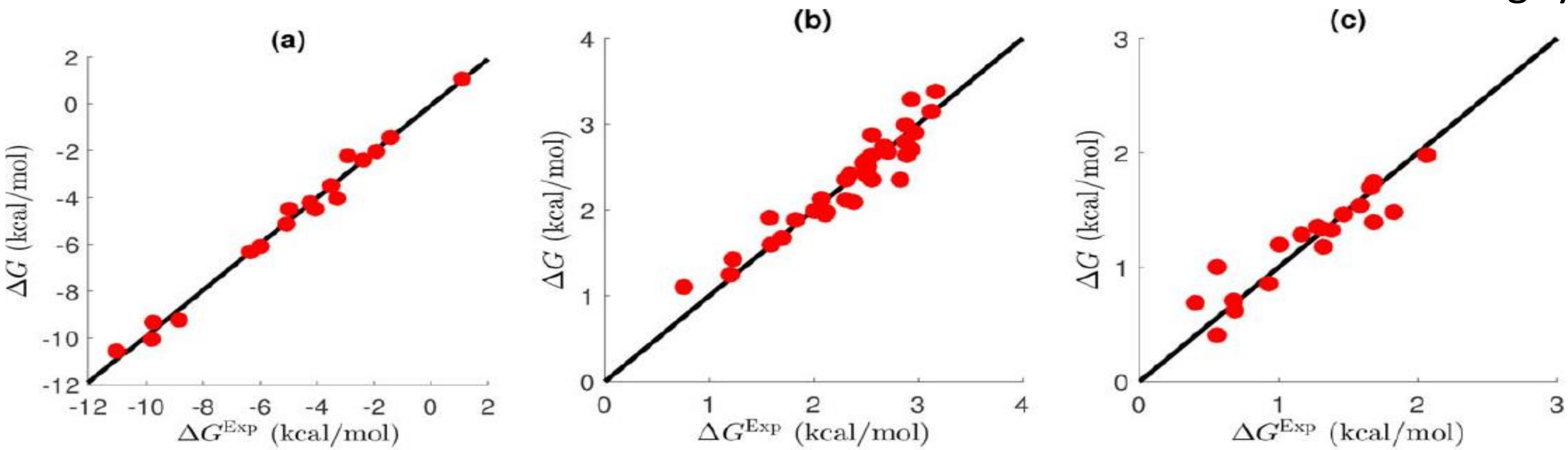
Topological fingerprints

(Xia & Wei, IJNMBE, 2014)

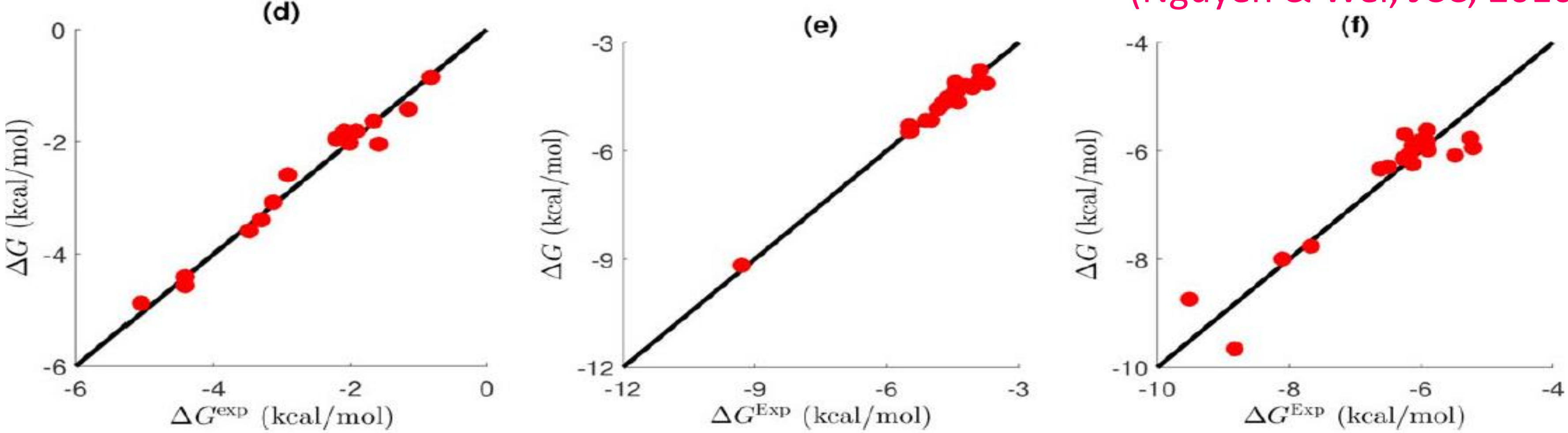
Curvature (C_j) based solvation model

$$\Delta G^{\mathrm{np}} = \gamma \mathrm{A} + p\mathrm{V} + \sum_j \lambda_j C_j + \rho_0 \int_{\Omega_s} U^{\mathrm{vdW}} d\boldsymbol{r}$$

Duc Nguyen

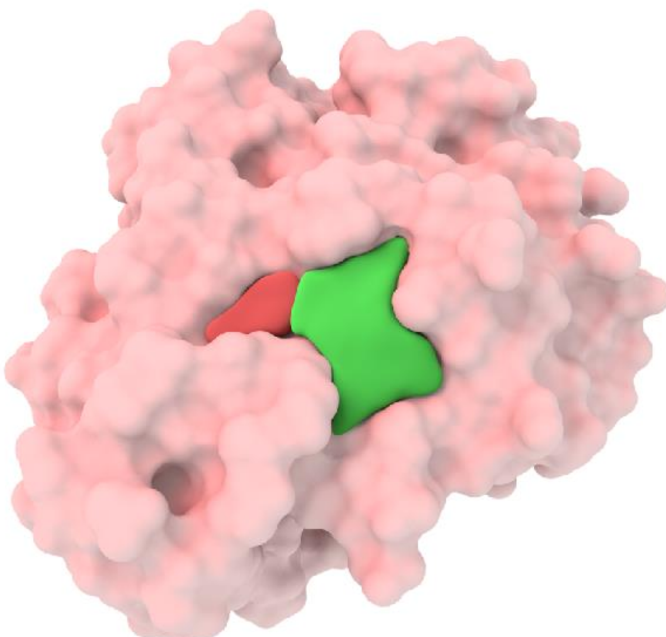


(Nguyen & Wei, JCC, 2016)

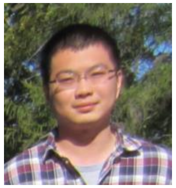


Protein Pocket Detection

Pocket of multi-ligand Interactions



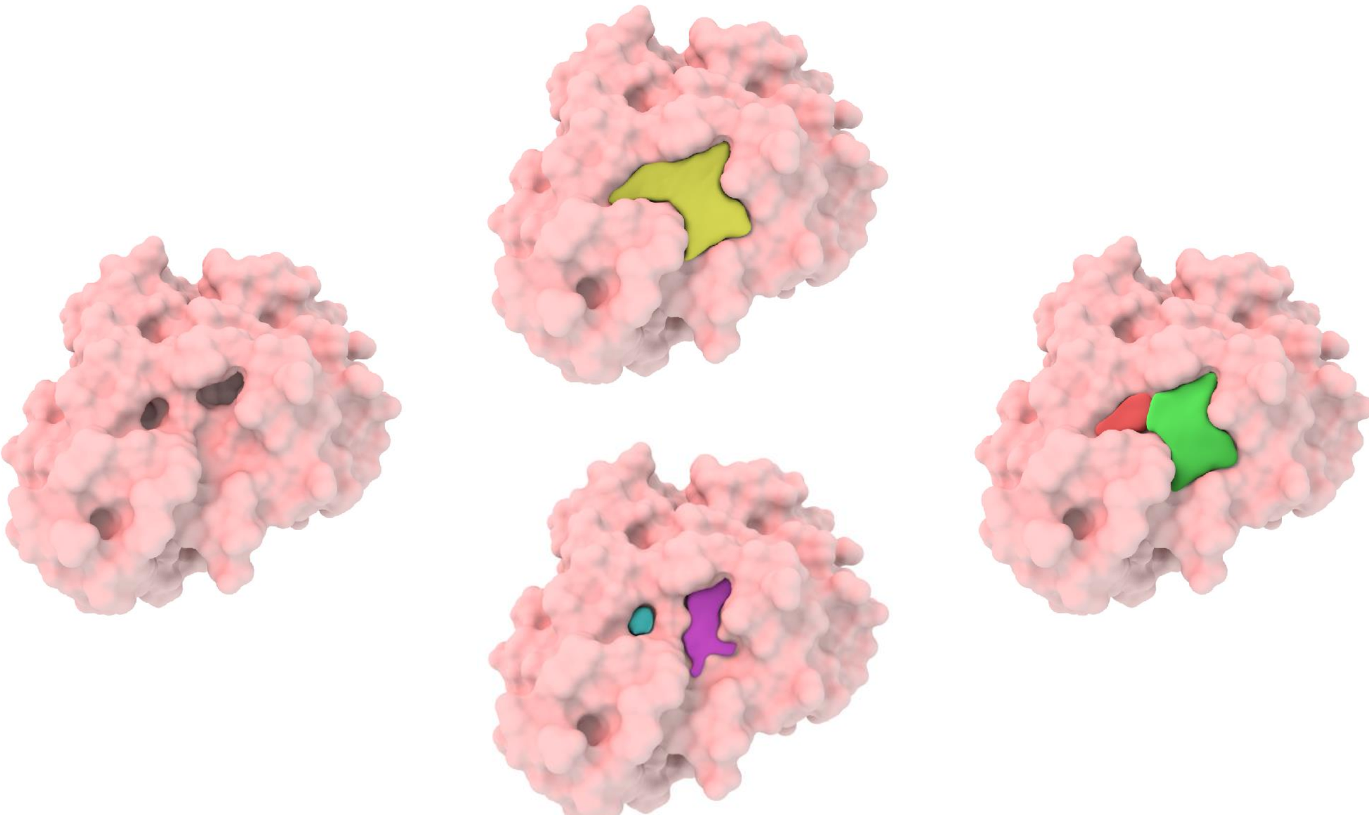
Rundong Zhao



Zixuan Cang

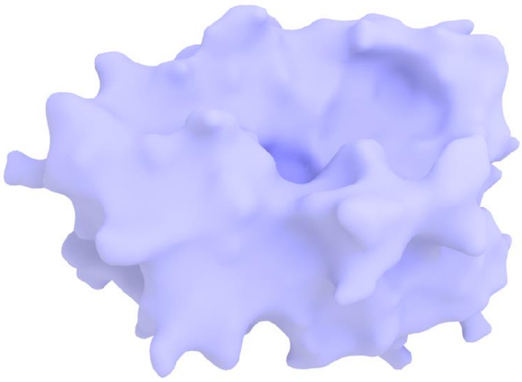
(Zhao, Cang, Tong, and Wei, Bioinformatics, 2018)

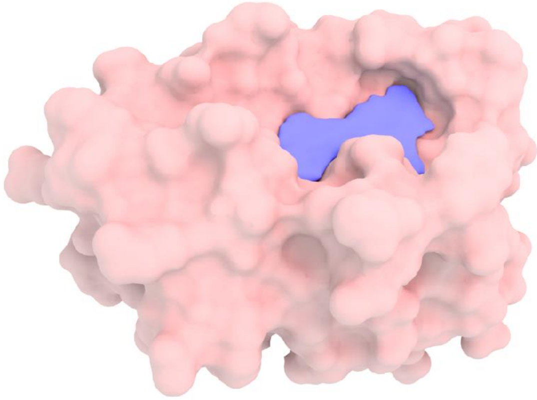
Hierarchical Pocket Structure



(Zhao, Cang, Tong, and Wei, Bioinformatics, 2018)

Algorithm Outline

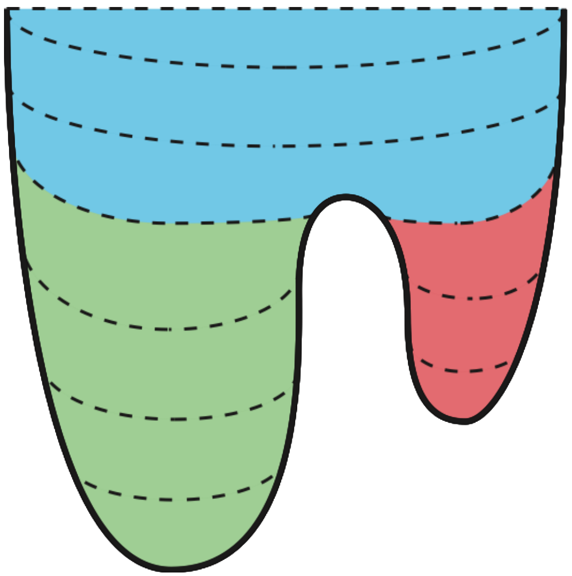




(Zhao, Cang, Tong, and Wei, Bioinformatics, 2018)

Deformation and Topology

ECCB 2018



Evolution Equation

$$\frac{\partial S}{\partial t} = |\nabla S| \nabla \cdot \frac{\nabla S}{|\nabla S|}$$

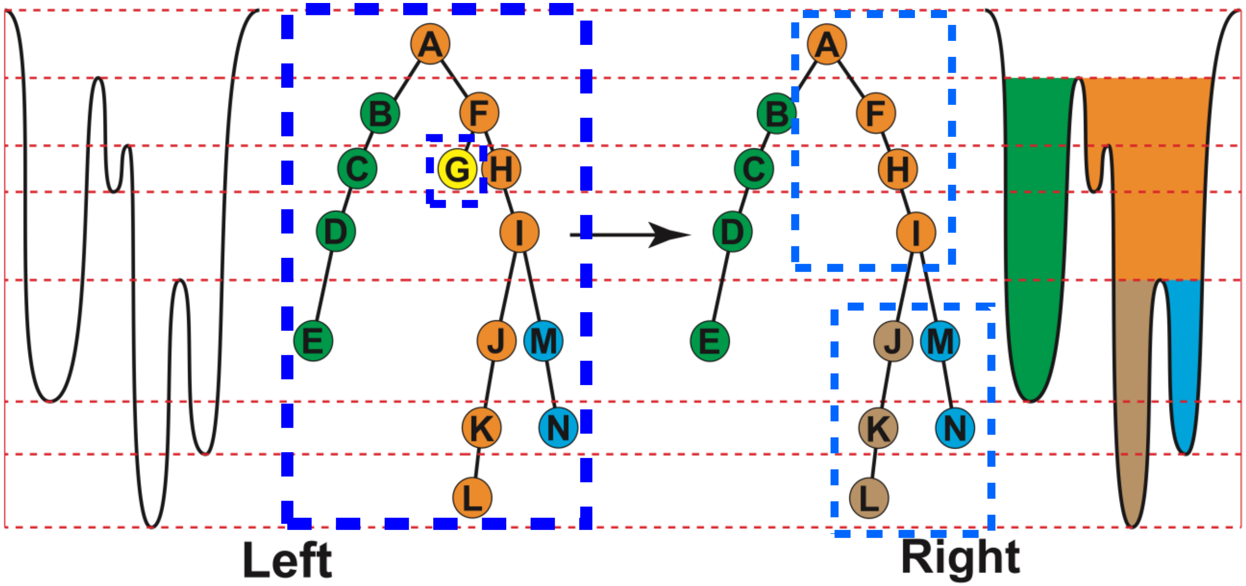
Connected Components Splitting

$$\mathcal{H}_i = \frac{\mathrm{Ker}(\partial_i)}{\mathrm{Im}(\partial_{i+1})}$$

(Zhao, Cang, Tong, and Wei, Bioinformatics, 2018)

Hierarchy (Reeb Graph)

Rundong Zhao

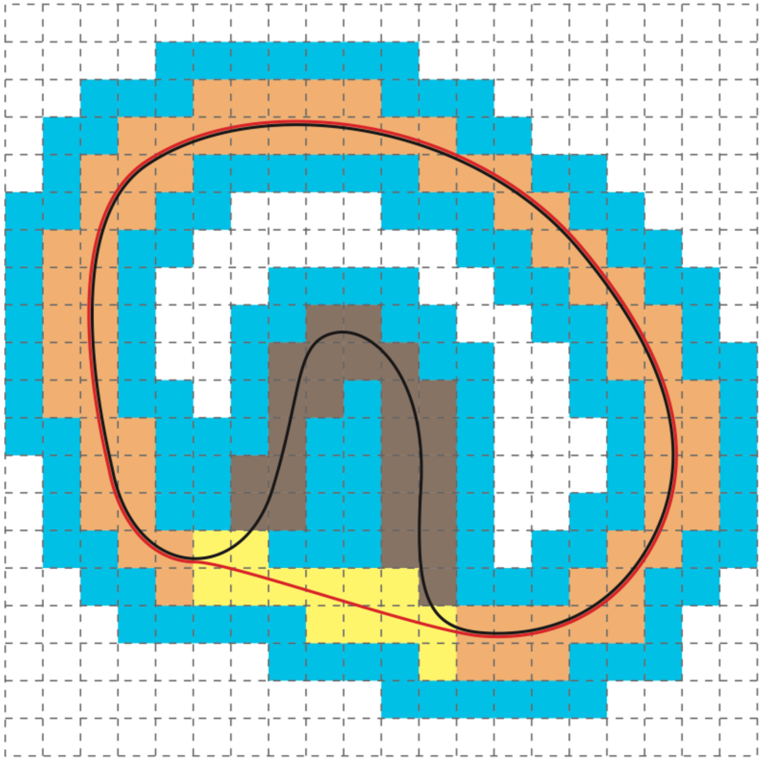


(Zhao, Cang, Tong, and Wei, Bioinformatics, 2018)

Data Structure

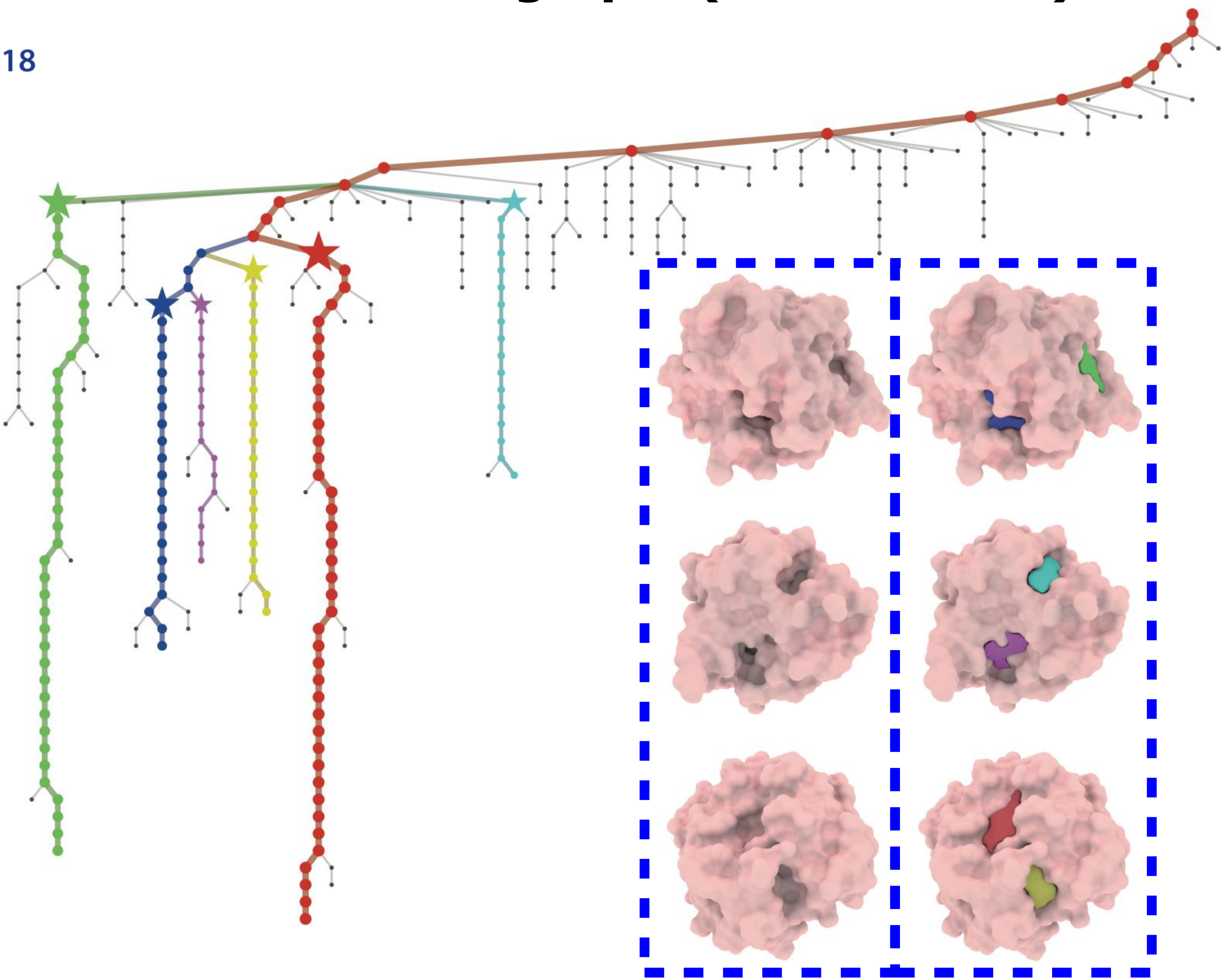
Narrow band (Colored) voxels are considered.

Straightforward calculation of pocket area and volume

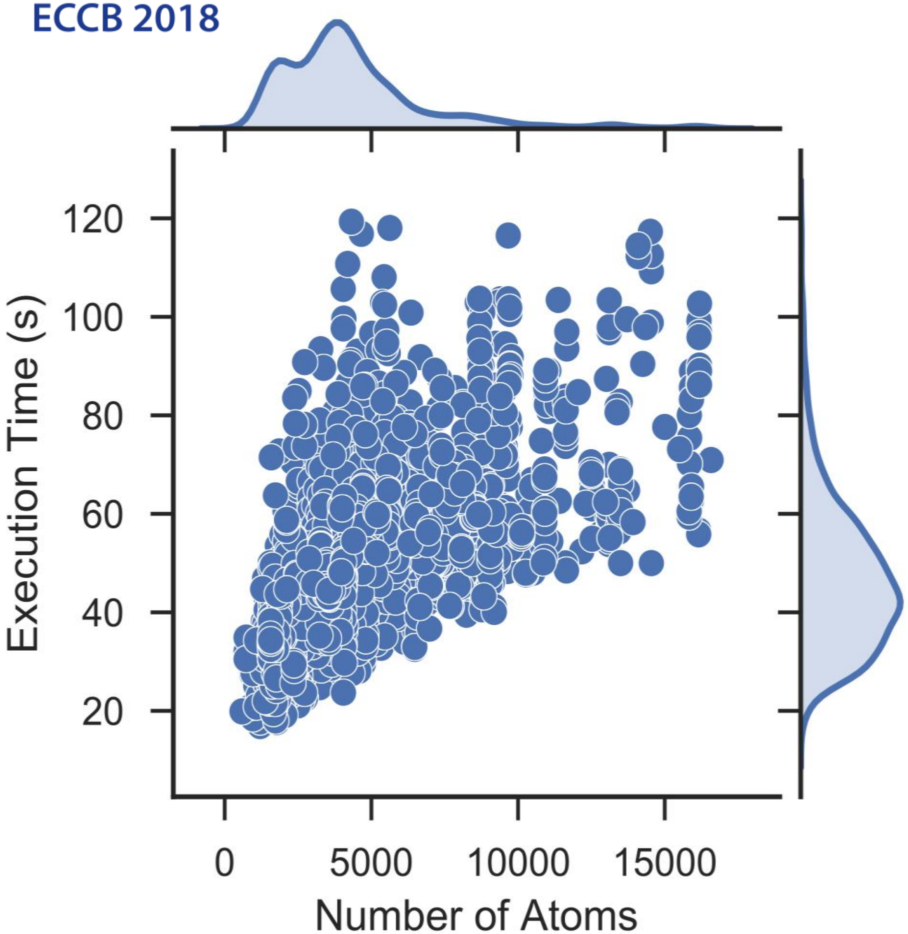


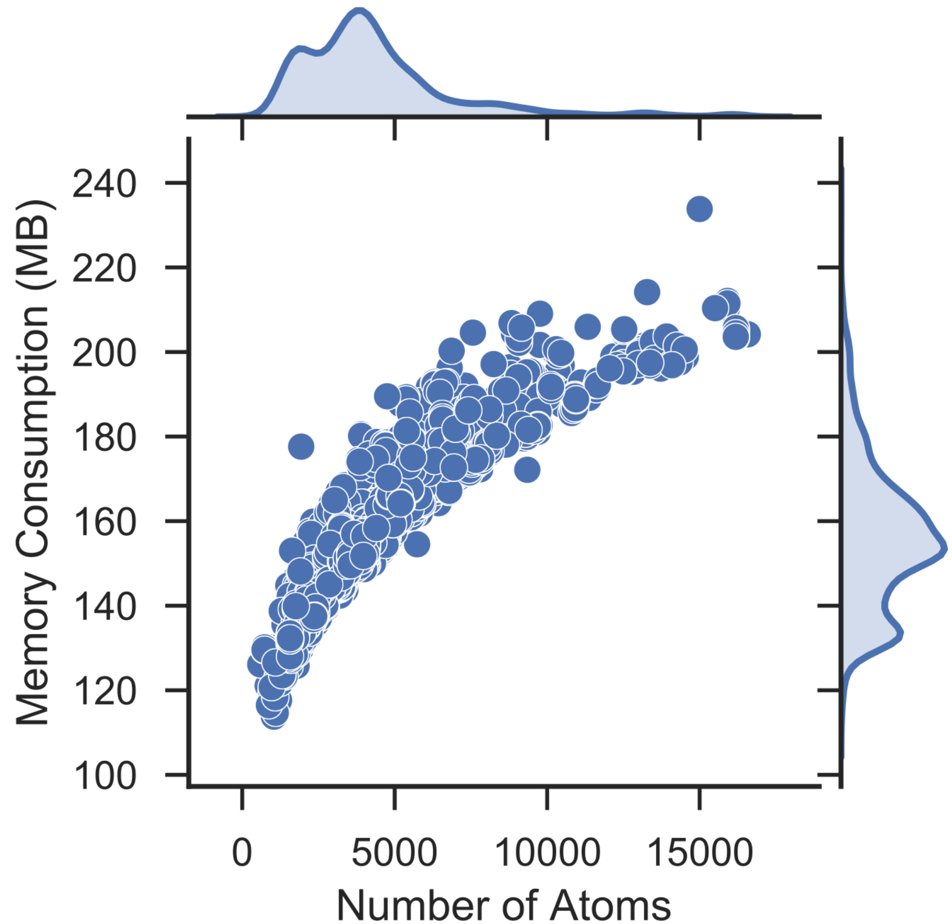
(Zhao, Cang, Tong, and Wei, Bioinformatics, 2018)

Protein pockets detected by curvature flow and Reeb graph (Protein 3ao4)



Time and Memory





Accuracy

ECCB 2018

	Whole proteins			Single chain proteins			Core-set proteins		
	P_{all} (4414)			P_{sc} (2430)			P_{cr16} (290)		
d	0.25	0.5	0.75	0.25	0.5	0.75	0.25	0.5	0.75
3Å	0.91	0.86	0.78	0.94	0.89	0.83	0.95	0.89	0.81
4Å	0.91	0.86	0.76	0.94	0.89	0.80	0.95	0.89	0.77
5Å	0.91	0.86	0.68	0.94	0.89	0.71	0.94	0.90	0.71

Accuracy ~80%, Best 95%

(Zhao, Cang, Tong, and Wei, Bioinformatics, 2018)

Further topics and future directions

- Phase field modeling of membrane structure and dynamics.
- Canham-Helfrich curvature, Ginzburg-Landau, and Cahn-Hilliard models for endoplasmic reticulum and mitochondrial ultrastructure, etc.
- Frenet–Serret frame analysis of protein structure.
- Mathematical modeling of subcullular organelle structure and dynamics.
- Element specific molecular manifolds.
- Atom specific molecular manifold.
- Interaction molecular manifolds.
- Curvature tensor based modeling of biomolecules.
- The de Rham-Hodge theory analysis of biomolecules.
- Convergence of Riemannian manifolds for biomolecular data.
- Gromov-Hausdorff and Gromov-Wasserstein metrics for biomolecular systems.


thank you