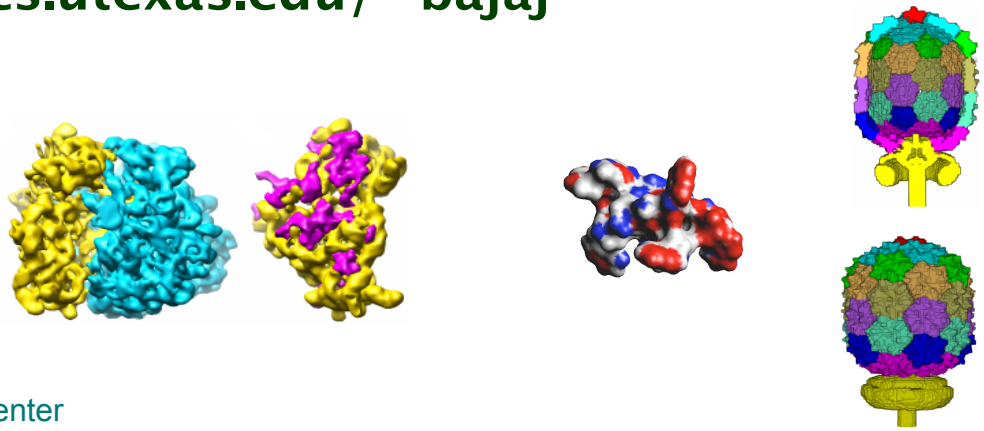
Molecular Electron Microscopy
to
Biophysical Analysis

Chandrajit Bajaj

http://www.cs.utexas.edu/~bajaj



Computational Visualization Center
Institute for Computational and Engineering Sciences
Department of Computer Sciences
University of Texas at Austin

January 2008

Acknowledgements

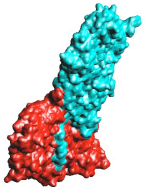
- **Group Members**
  - **Albert Chen (CS, Ph.D)**
  - **Rezaul Chowdhury (PostDoc)**
  - **Ajay Gopinath (BME,Ph.D)**
  - **Andrew Gillete (Math, Ph.D.)**
  - **R. Inkulu (PostDoc)**
  - **Alex Kim (CS, Ugrad)**
  - **Bongjune Kwon (BME,Ph.D)**
  - **Oliver Martin (CS, Ugrad)**
  - **Maysam Mousallem (CS,PhD)**
  - **Joe Rivera (Res. Scientist)**
  - **Aditi Saha (CS, MS)**
  - **Jesse Sweet (Math, MS)**
  - **Dongjin Suh (CS, MS)**
  - **Samrat Goswami (PostDoc)**
  - **Zeyun Yu (UCSD)****
  - **Xiaoyu Zhang (CSU)****
  - **Wenqi Zhao (CAM, Ph.D.)**

- **Senior Collaborators**
  - Manfred Auer (LBL,UCB)
  - Tim Baker, M. Ellisman (UCSD)
  - Nathan Baker (WashU)
  - Wah Chiu, Steve Ludtke (Baylor)
  - Inderjit Dhillon (UT, CS)
  - Ron Elber, Peter Rossky (UT, ICES)
  - Joachim Frank (Suny Albany, Columbia)
  - Kristen Harris , Dan Johnston (UT, ILM)
  - Tom Hughes, Tinsley Oden (UT,ICES)
  - Justin Kiney, Tom Bartol, Terry Sejnowski (Salk)
  - Andy McCammon, Michael Holst (UCSD)
  - Art Olson, Michel Sanner, David Goodsell (TSRI)
  - Alberto Paoluzzi, Antonio DiCarlo(Roma)
  - Sriram Subramaniam (NIH-NCI)

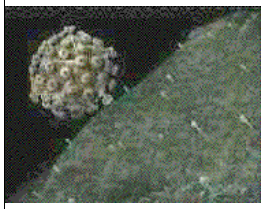
- **Sponsored by**
  - **NIH: P20-RR020647, R01-GM074258, R01-0783087, R01-EB004873**

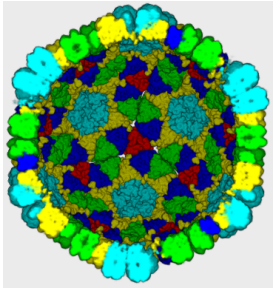
    NSF: ITR-IIS-032550, DDDAS-CNS-0540033

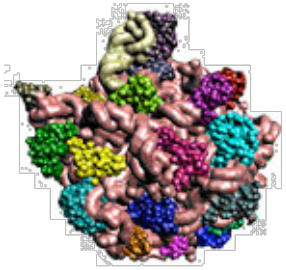
    UT: HTDS

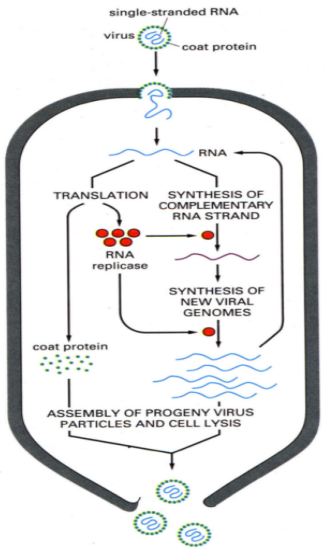


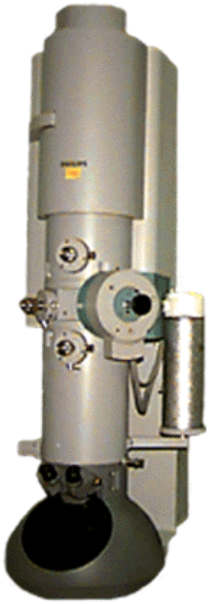
Seek, Analyze and Destroy: Imaging to BioPhysical Modeling to Drug Discovery

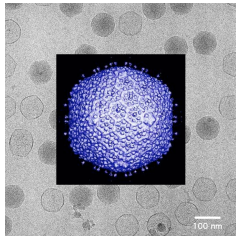


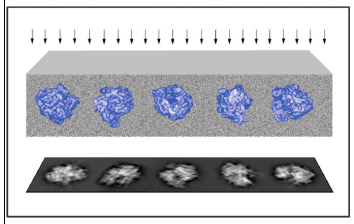




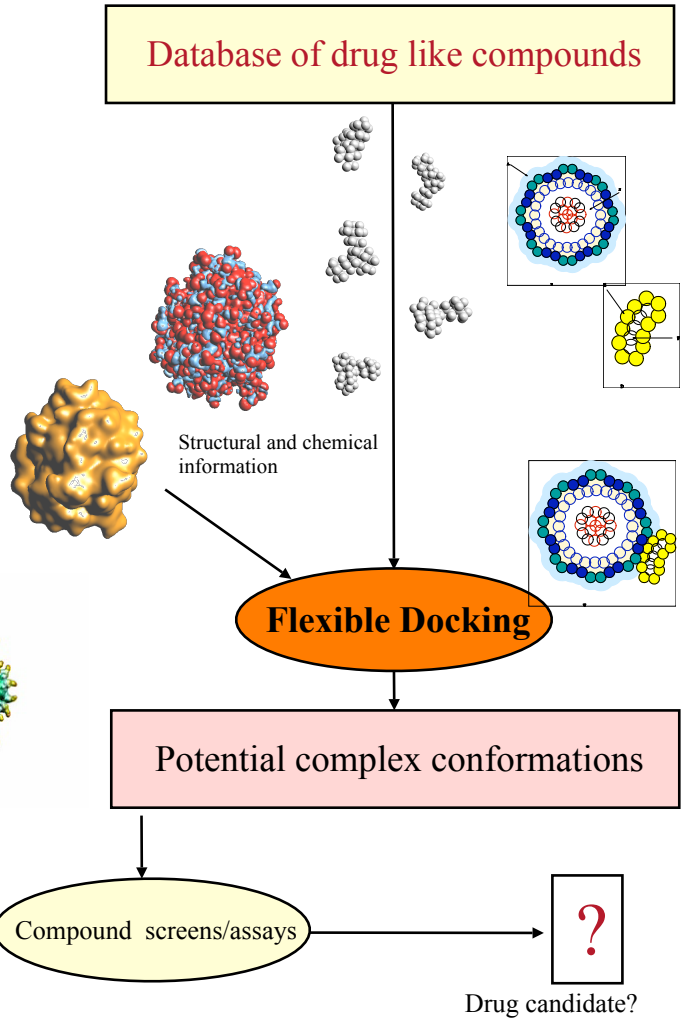








Single Particle Cryo-EM



HRV with IG

Collaborators: BCM, TSRI **Sponsored by NIH, TI-3D

Center for Computational Visualization http://www.ices.utexas.edu/CCV
Institute of Computational and Engineering Sciences
Department of Computer Sciences University of Texas at Austin

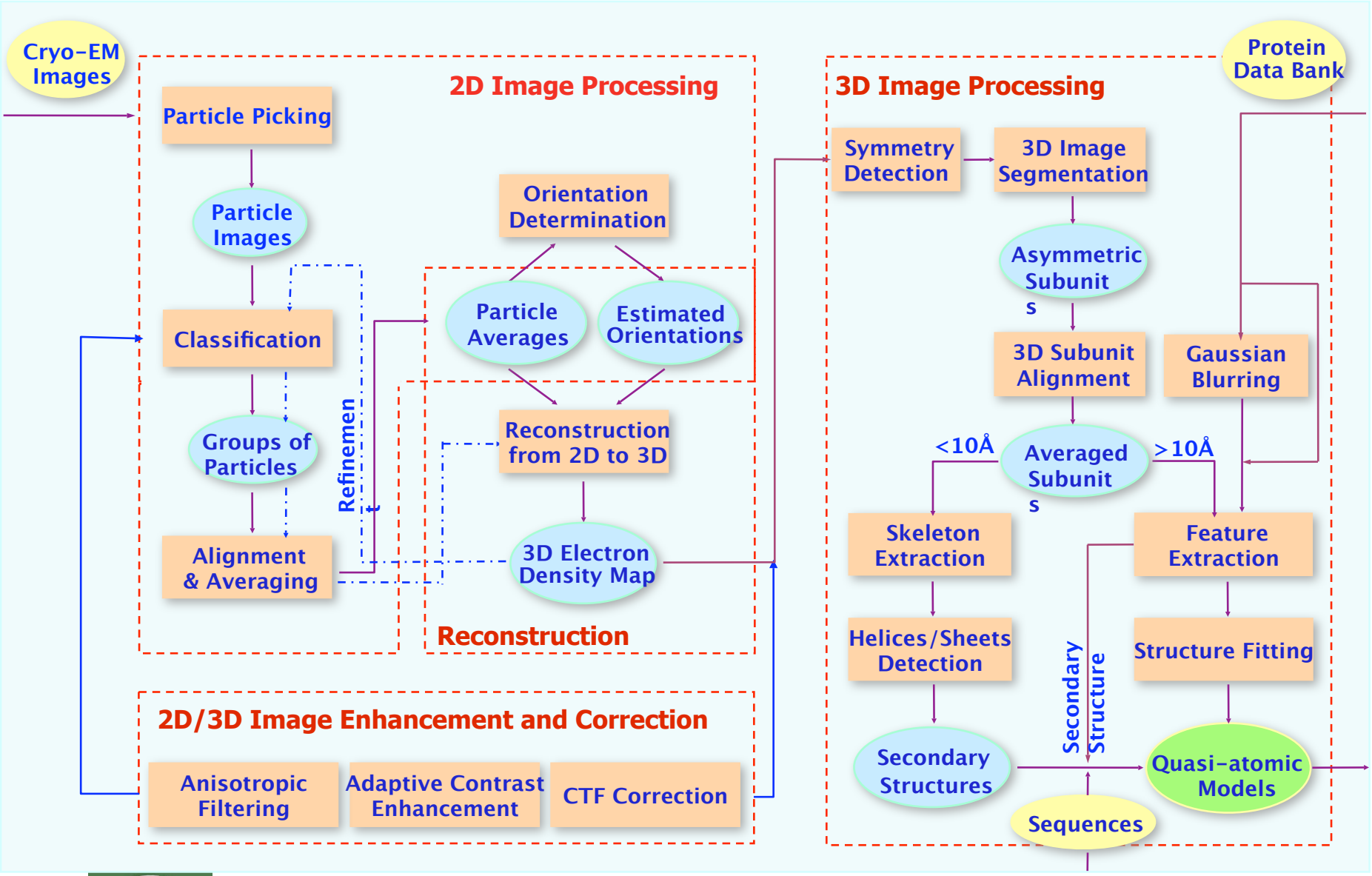
January 2008

Molecular Structure Elucidation from Electron Microscopy

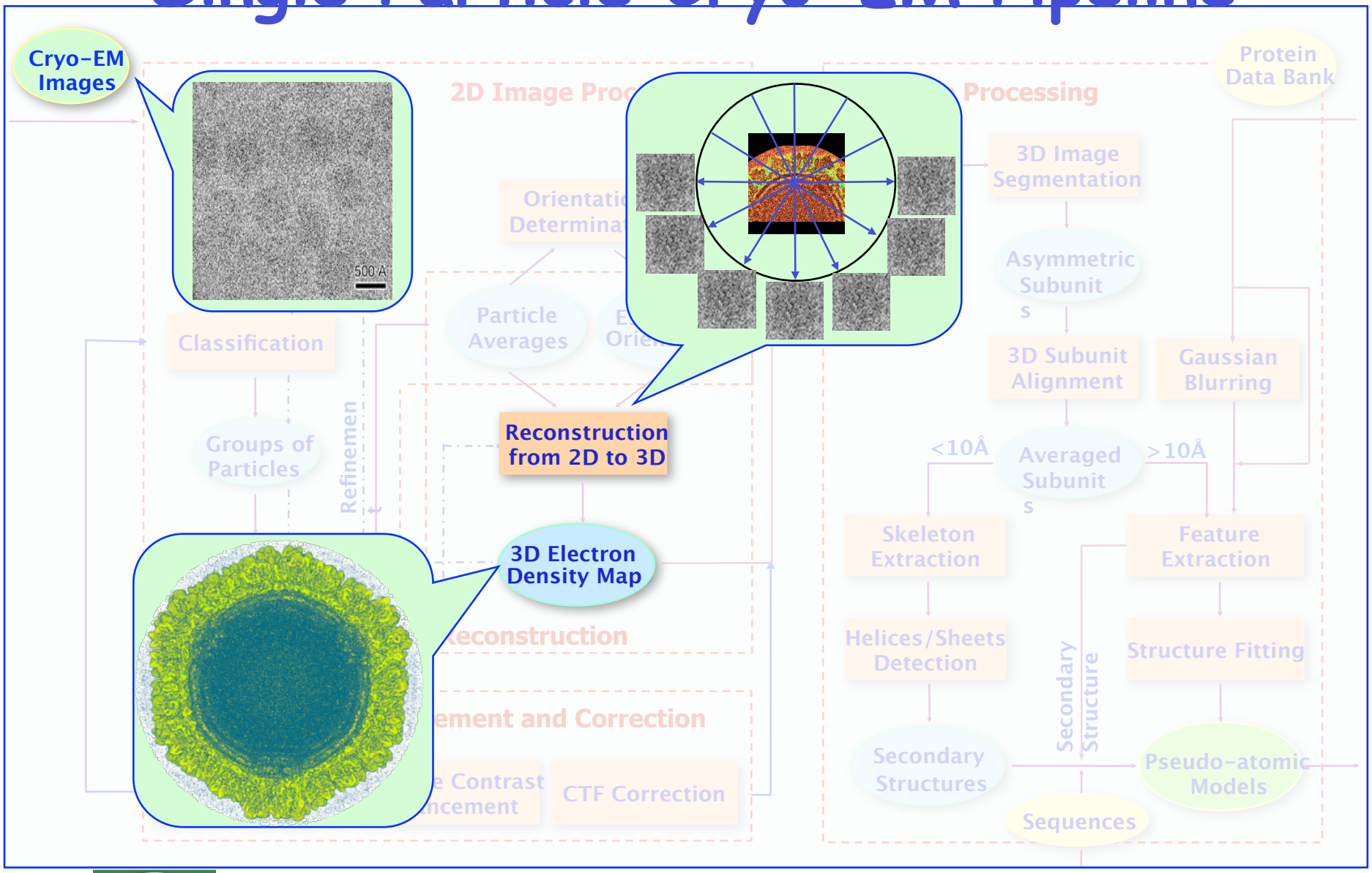
Center for Computational Visualization
Institute of Computational and Engineering Sciences
Department of Computer Sciences
University of Texas at Austin

January 2008

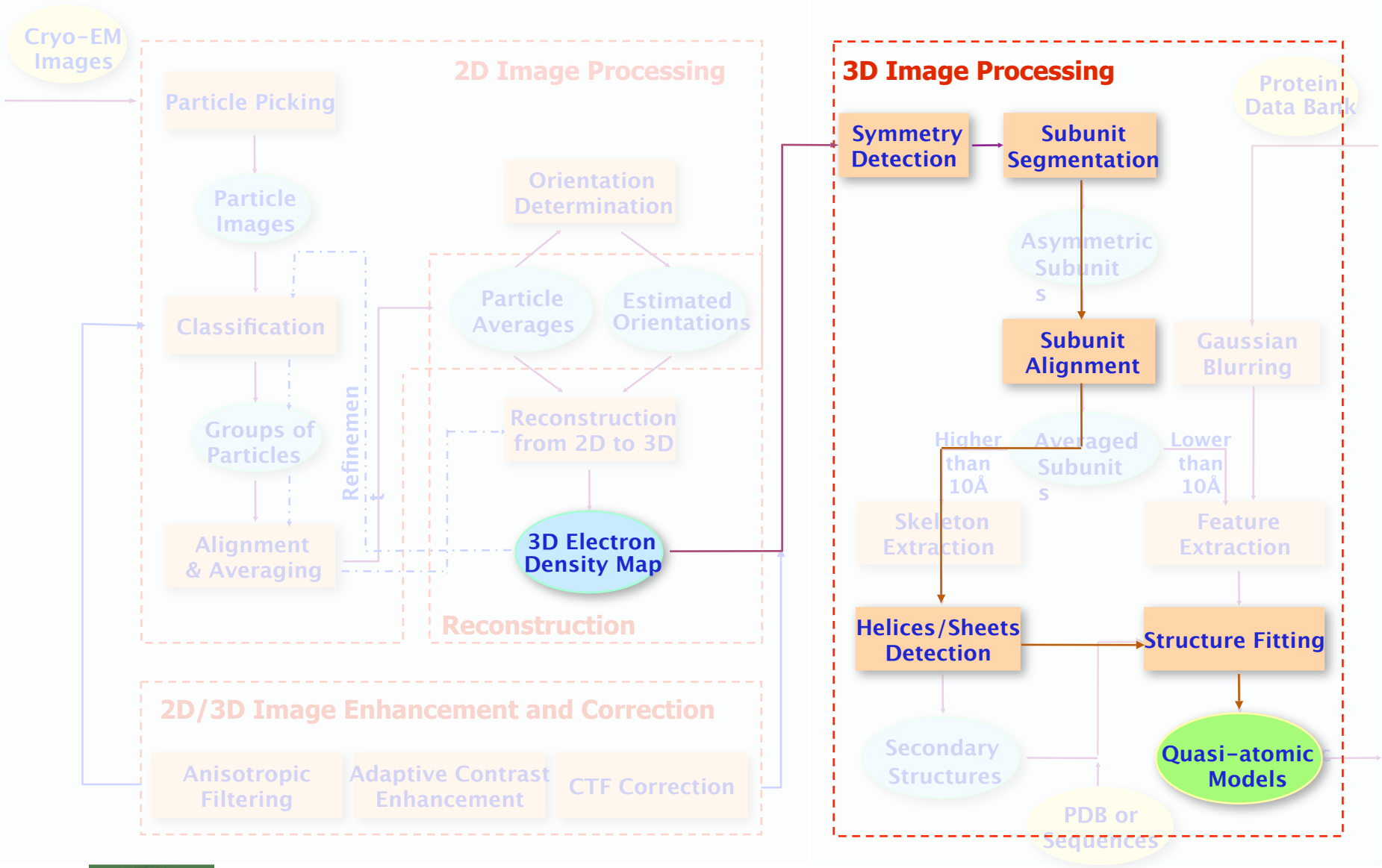
Quasi Atomic Models from Single Particle Cryo-EM



Single Particle Cryo-EM Pipeline



Single Particle Cryo-EM: Automatic Structure Analysis



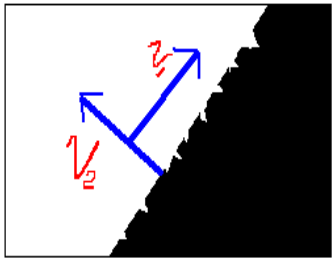
Step #1 : Anisotropic Filtering

Bilateral filtering

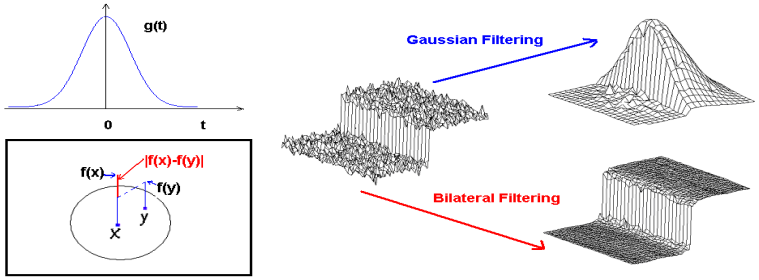
$$h(x,\xi) = e^{-\frac{(x-\xi)^2}{2\sigma_d^2}} \times e^{-\frac{(f(x)-f(\xi))^2}{2\sigma_r^2}}$$

where σ_d and σ_r are parameters

and f(.) is the image intensity value.



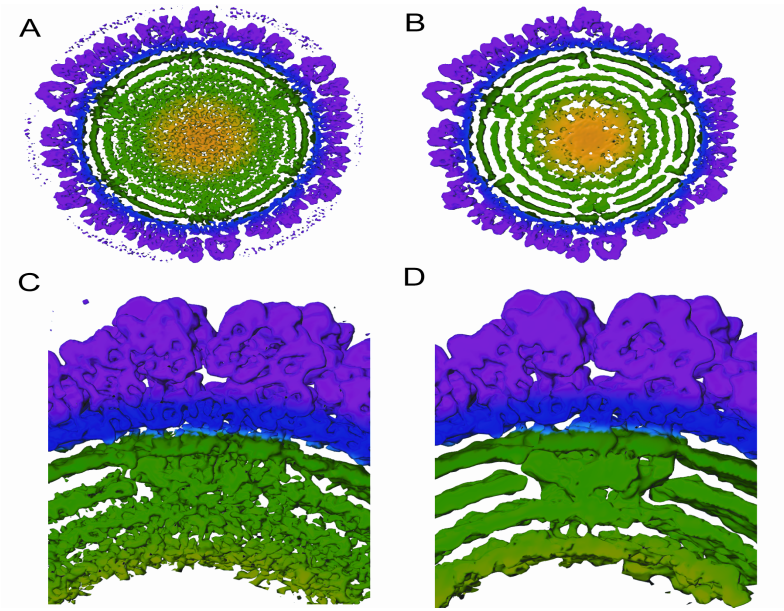
C. Bajaj, G. Xu, ACM Transactions on Graphics, (2003),22(1), pp. 4- 32.



Anisotropic diffusion filtering

$$\partial_t \phi - \mathrm{div}(a(|\nabla \phi_\sigma|)\nabla \phi) = 0$$

where **a** stands for the diffusion tensor determined by local curvature estimation.



W. Jiang, M. Baker, Q. Wu, C. Bajaj, W. Chiu, Journal of Structural Biology, 144, 5,(2003), Pages 114-122

Step #2: Critical Point Detection

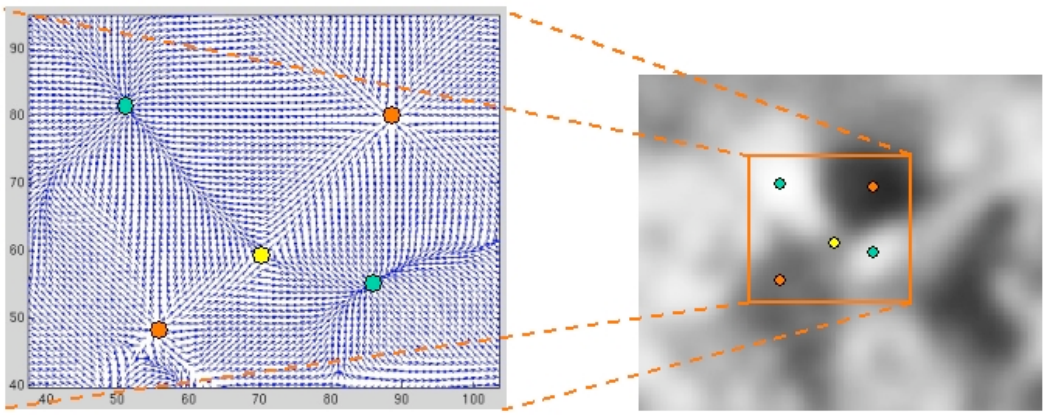
- For smooth data:
  - zeroes of the gradient vector field
  - simple, easy to implement
- For noisy data:
  - Gradient vector diffusion
  - higher time complexity but robust to noise

- Gradient vector diffusion:
  - smoothing the vector fields
  - diffusion to flat regions

$$\begin{cases} \dfrac{\partial u}{\partial t} = \mu \times div(g(\alpha)\nabla u) \\ \dfrac{\partial v}{\partial t} = \mu \times div(g(\alpha)\nabla v) \end{cases}$$

where $g(\alpha)$ is a decreasing function α is the angle between the central pixel and its surrounding pixels.

Z. Yu, C. Bajaj Proc. ICPR 2002, vol 2, 941- 944 and IEEE Transactions on Image Processing, 2005, 14, 9, 1324-1337



minimum (0) maximum (3) saddle (1, 2)

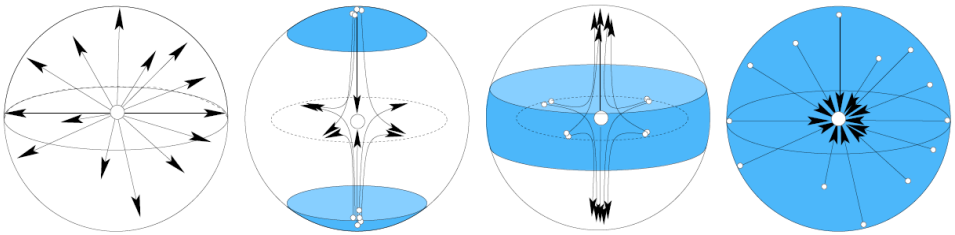
Critical Points, their Indices, and their Manifolds

Critical Point of a smooth function is a point where the gradient of the function vanishes.

Index of a critical point is the number of independent directions in which the function decreases.

In 3D, four types of critical points

1. Minima – index 0
2. Saddle of index 1
3. Saddle of index 2
4. Maxima – index 3



From EHNP SoCG'03

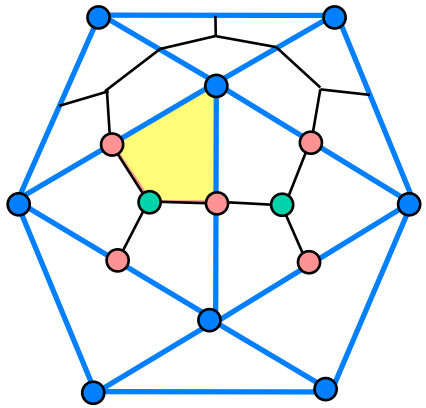
<u>Integral curve</u> : A path in the domain of the function on which at every point the tangent to the curve equals the gradient of the function.

<u>Stable Manifold</u> of a critical point is the union of all integral curves ending at the critical point.

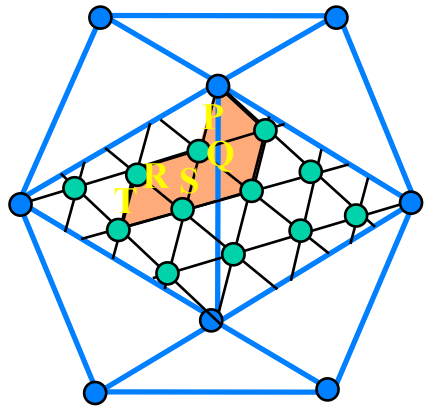
<u>Unstable Manifold</u> of a critical point is the union of all integral curves starting at the critical point.

Step #3: Symmetry Detection

- Asymmetric subunits in an icosahedra

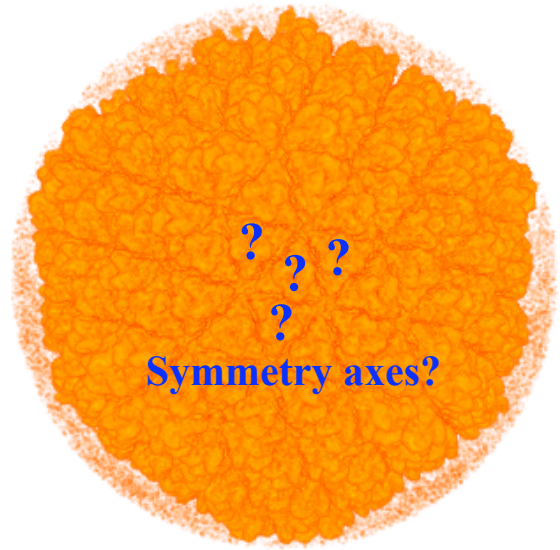


- Two-fold vertices
- Three-fold vertices
- Five-fold vertices



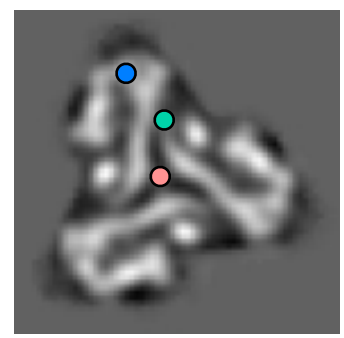
Local symmetry (RDV)

(260 trimers or 720 proteins)



- Critical Point Correlation search, addtly sped up by Spherical FFT:

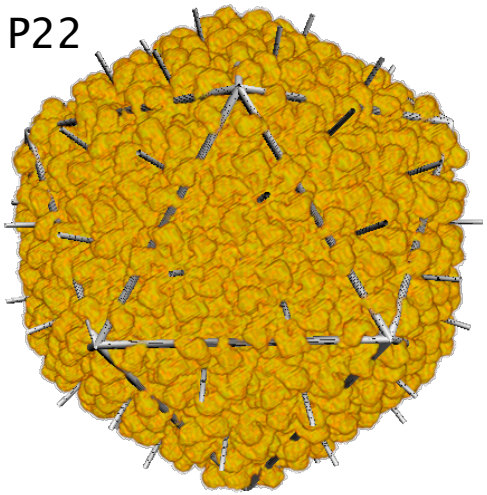
Find best c, minimizing: $$\sum_{\vec{r}\in D_0}\left(f(\vec{r})-f(R_{2\pi/n}(c)\times\vec{r})\right)^2$$



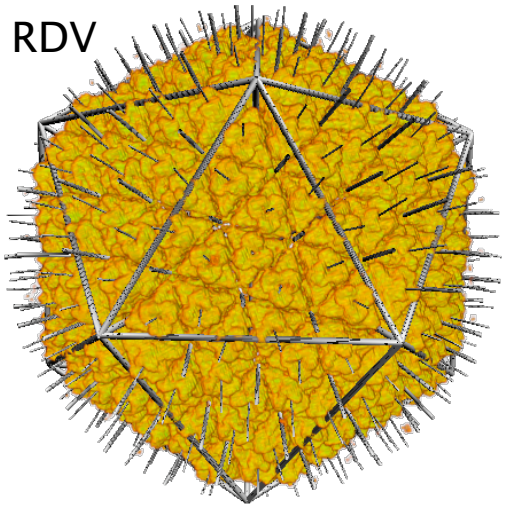
Z. Yu, C. Bajaj IEEE Transactions on Image Processing, 2005, 14, 9, 1324-1337

Results of Automatic Symmetry Detection in Virus and Phage Capsid Shells

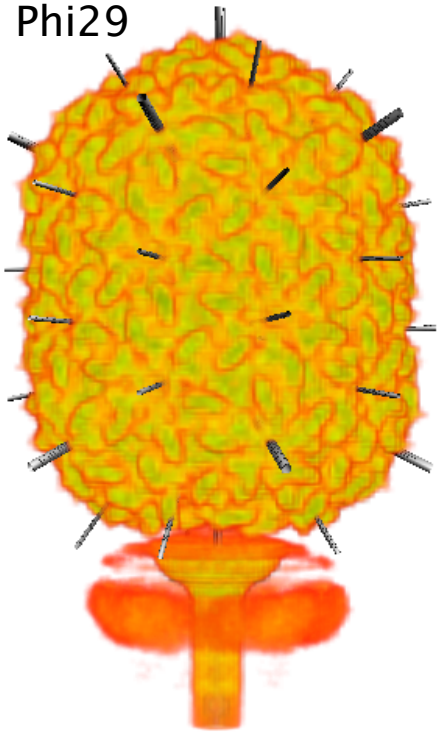
P22



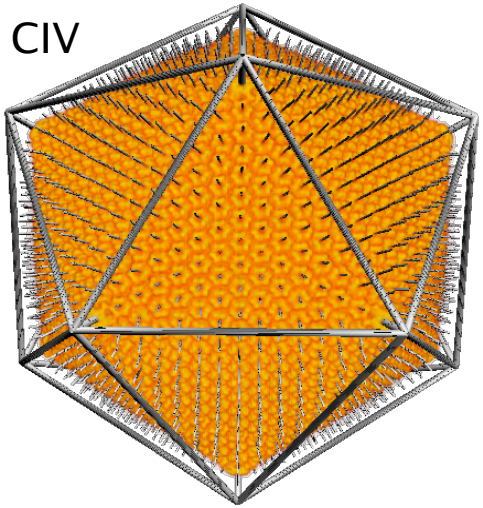
RDV



Phi29



CIV



Collaborators: Wah Chiu (BCM), Tim Baker (UCSD)

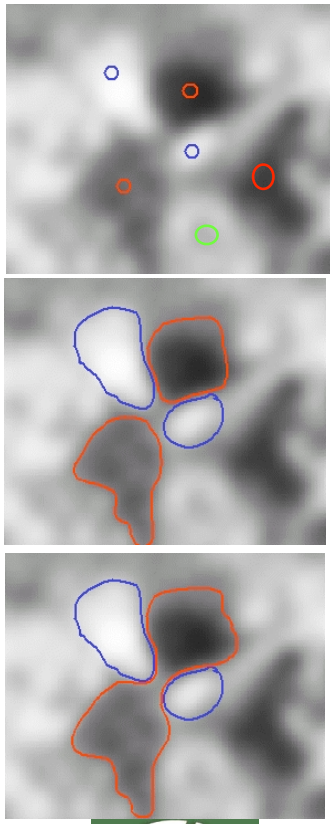
Center for Computational Visualization
Institute of Computational and Engineering Sciences
Department of Computer Sciences
University of Texas at Austin

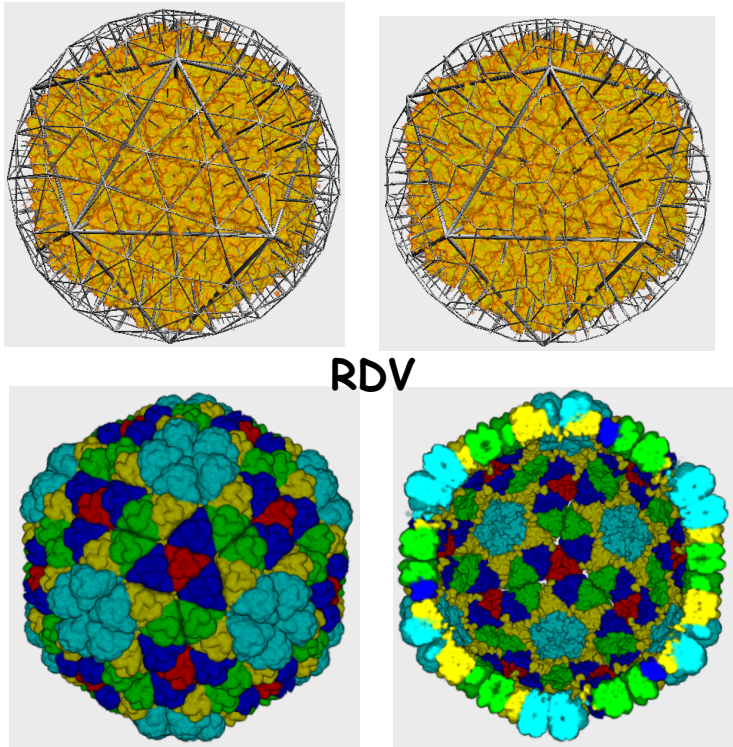
January 2008

Step #4: Subunit Segmentation

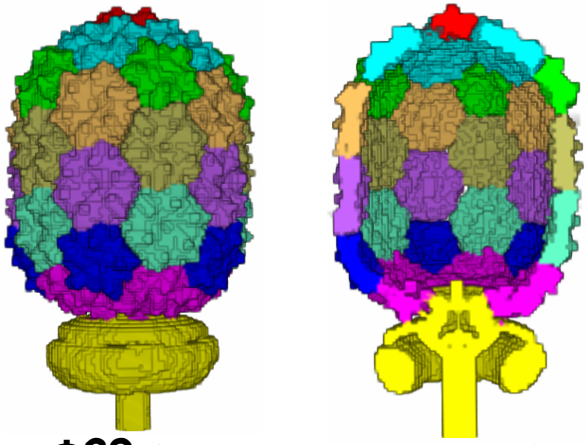
- Multi-seed Fast Marching Method
  - Classify the critical points based on local symmetry into separate groups.
  - Each seed initializes one contour, with its group's membership.
  - Contours march simultaneously. Contours with same membership are merged, while contours with different membership stop each other.

Zeyun, Bajaj IEEE Trans on Imag. Proc.,2005, Baker, Yu, Chiu, Bajaj, J S B 2006

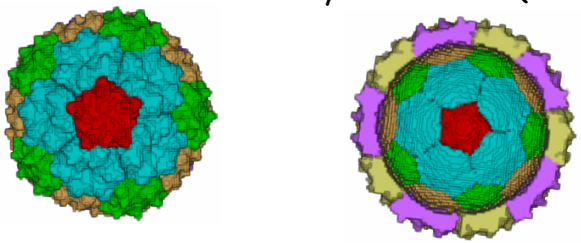




RDV

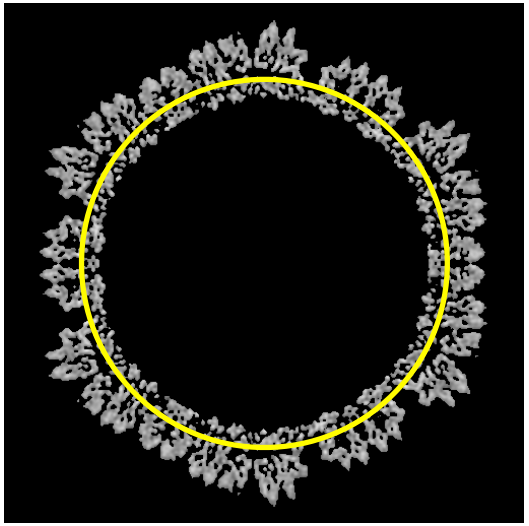


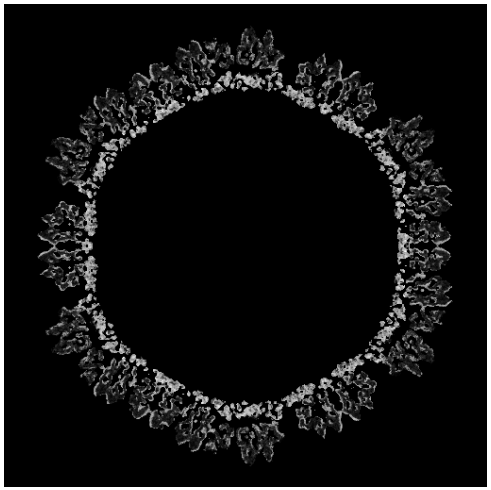
Φ29 data courtesy: Tim Baker (UCSD)

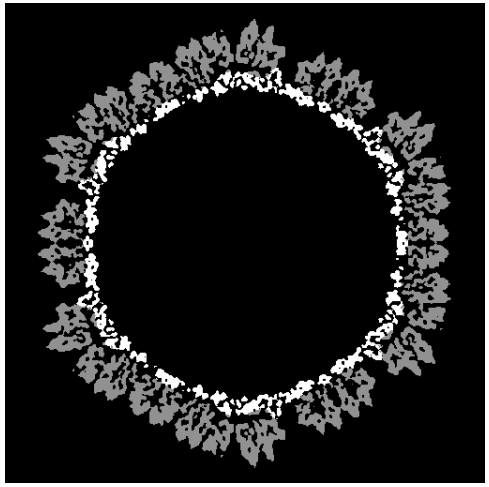


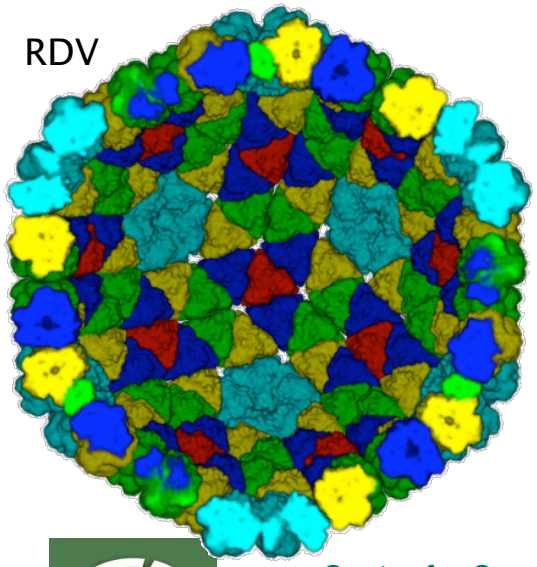
Bajaj, Xu 2007 "A Higher Order Level Set Method" ICES-TR

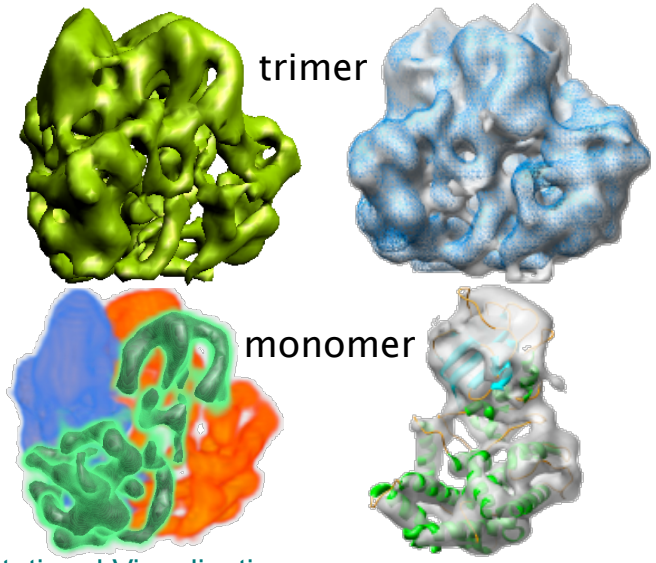
Capsid and Subunit Segmentation











Correlation Score with manually segmented subunits:

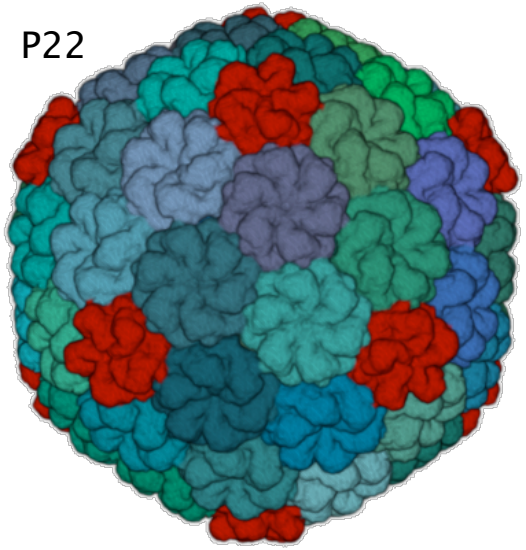
trimer: **0.74**
monomer: **0.80**

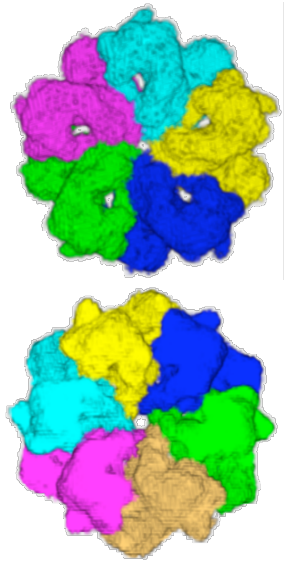
Correlation Score with blurred PDB density map:

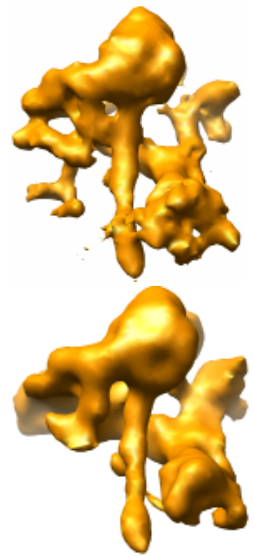
trimer: **0.85**
monomer: **0.84**

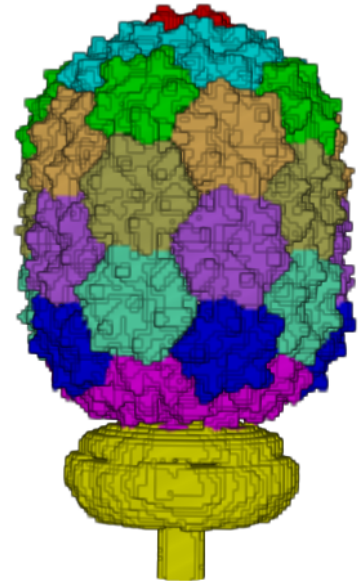
Subunit Segmentation

P22



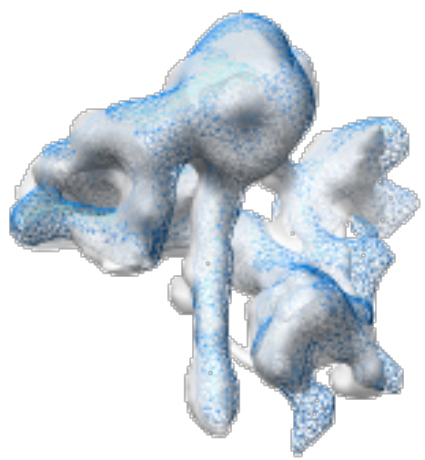


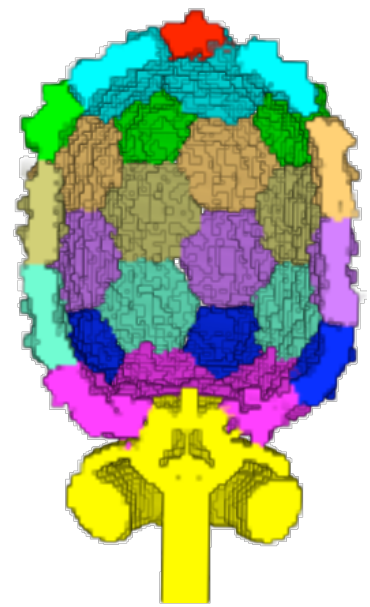




Correlation Score with manually segmented subunits:

5-fold: **0.72**
6-fold: **0.79**





Step #4b: Subunit Alignment

- Motivation
  - Similarity scores
  - Averaged subunit
  - Map Refinement
  - Structural fitting

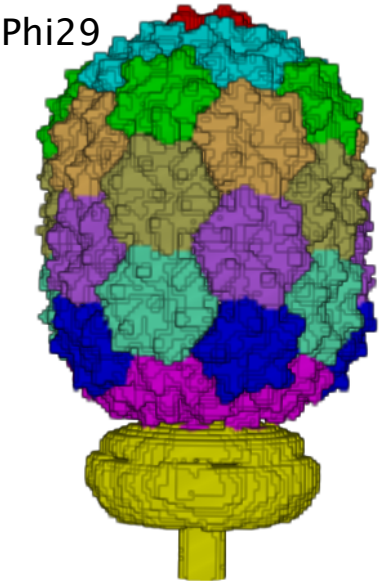
- General alignment problem
  - Six degrees of freedom

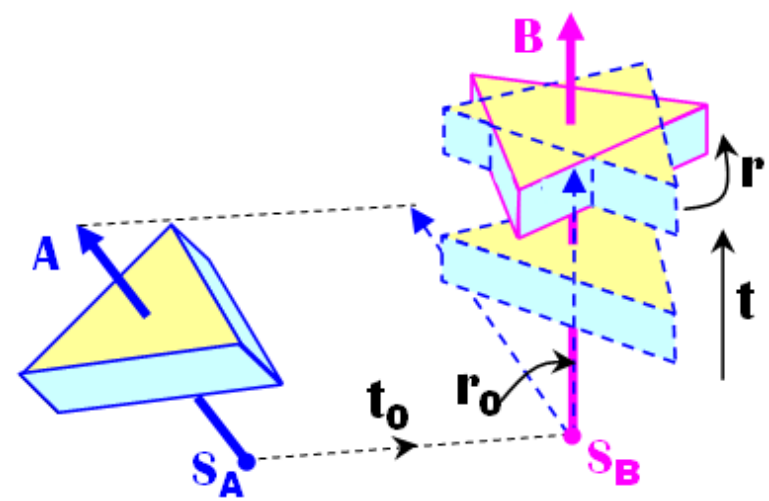
- Assumptions
  - Subunits are segmented
  - Subunits have an n-fold symmetry
  - Symmetry axes are known

Rotation **r** ⇩ **0 ~ 2π/n**

Translation **t** ⇩ **-5 ~ 5 pixels**

Phi29

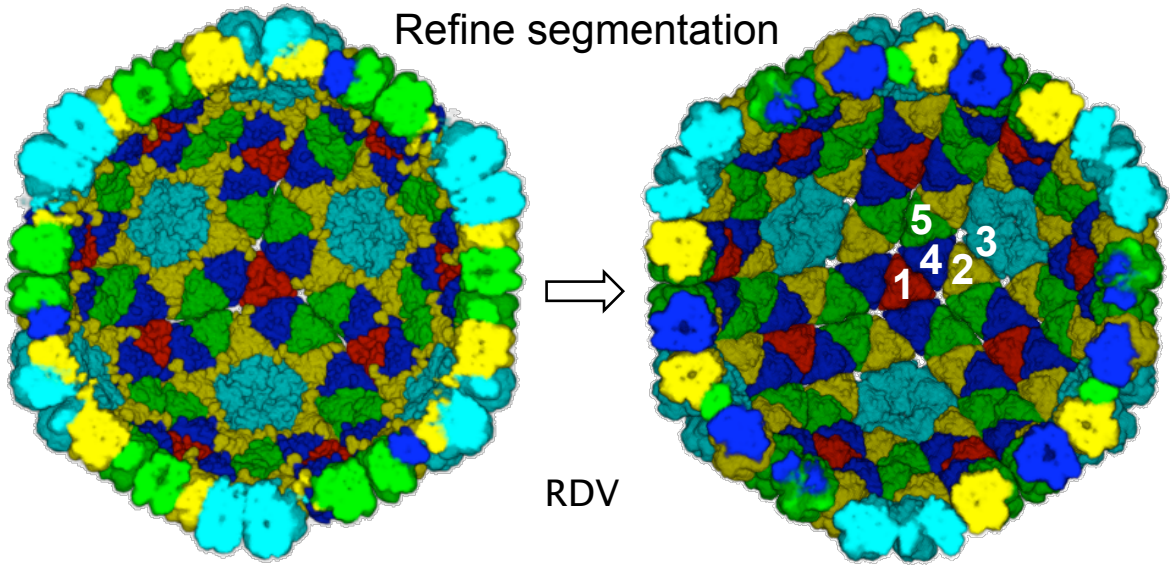




Two degrees of freedom

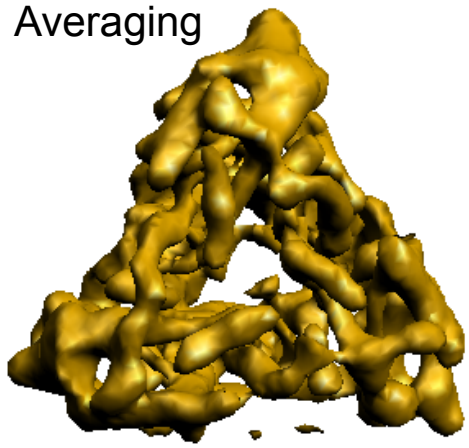
Subunit Alignment (RDV map refinement)

Refine segmentation



RDV

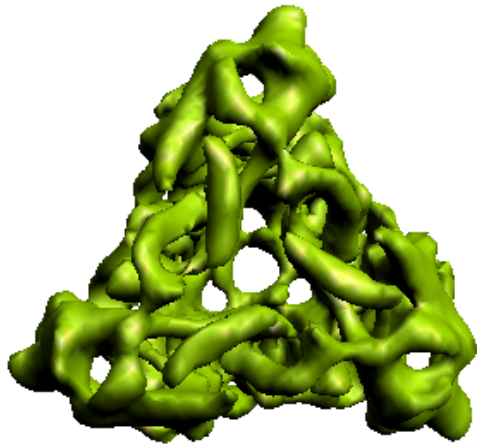
Averaging



A single RDV trimer

Similarity table

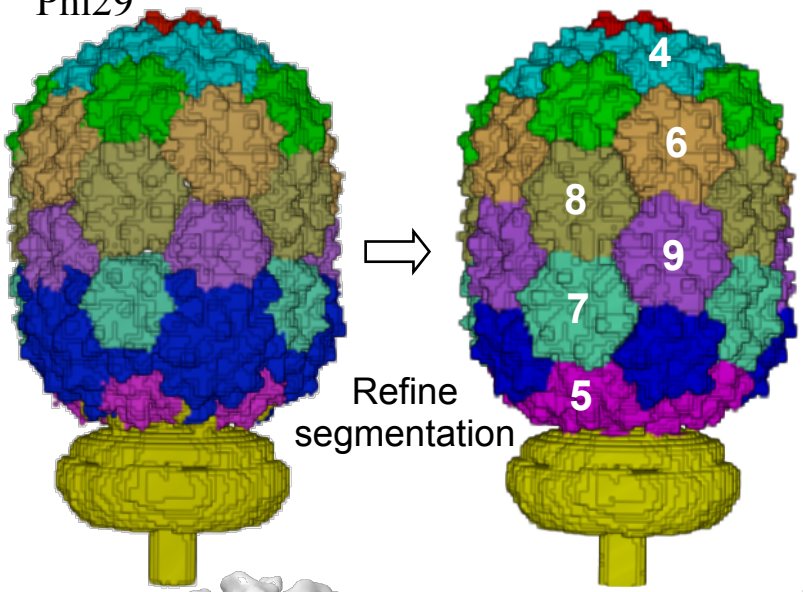
Similarity scores	1	2	3	4	5
1	**0.955**	0.900	0.894	0.911	0.848
2		**0.934**	0.885	0.889	0.848
3			**0.856**	0.880	0.845
4				**0.926**	0.854
5					**0.872**



Averaged trimer

Subunit Alignment (Simul. Structure Fitting)

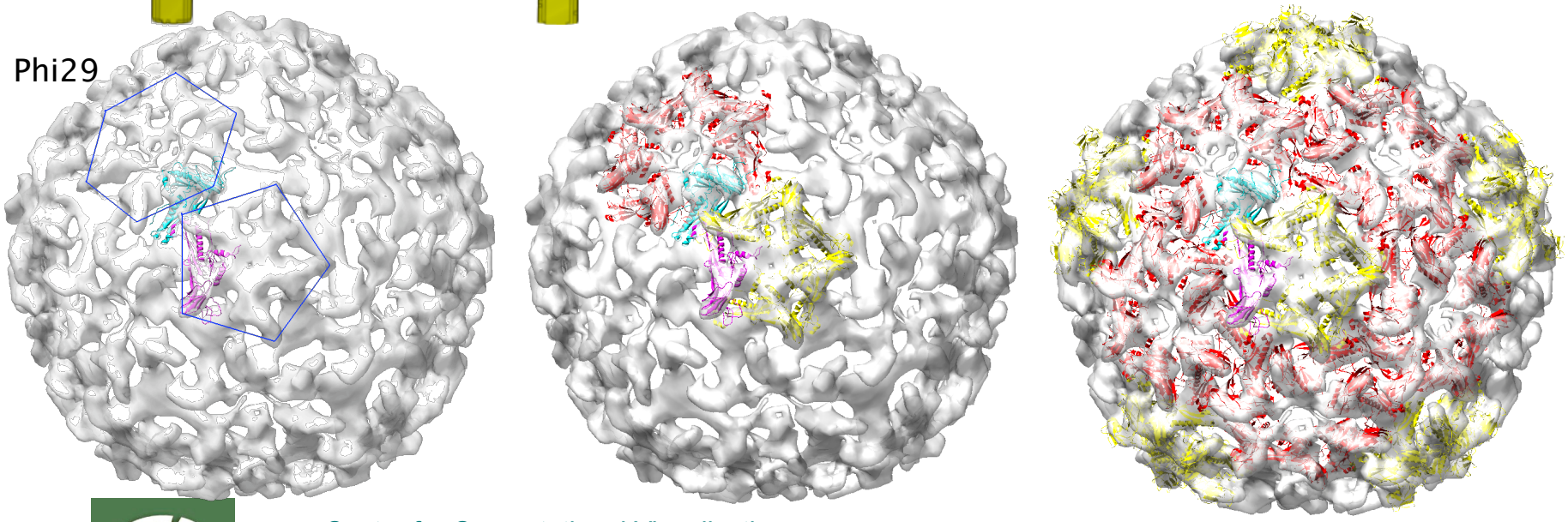
Phi29



Similarity table

	#4	#5	#6	#7	#8	#9
#4	**0.97**	0.79	0.95	0.94	0.87	0.88
#5		**0.81**	0.79	0.78	0.77	0.79
#6			**0.98**	0.96	0.88	0.88
#7				**0.96**	0.89	0.88
#8					**0.87**	0.94
#9						**0.81**

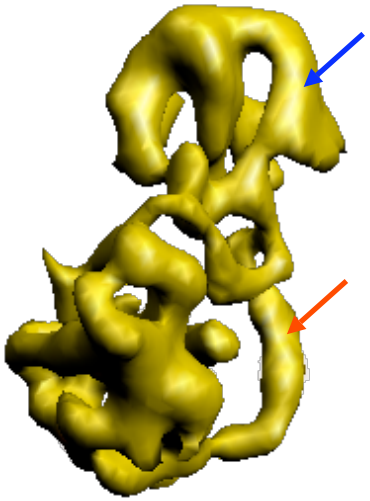
Phi29



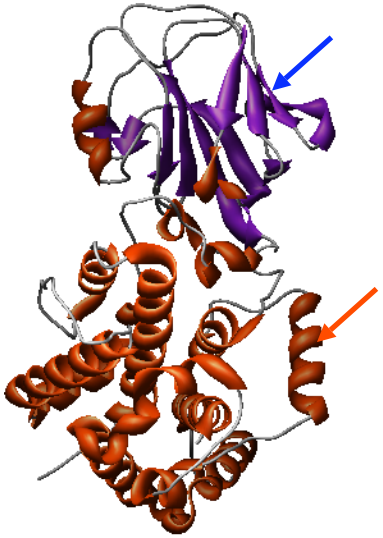
Step #5: Secondary Structure Identification

- Motivation
  - One more step towards quasi-atomic modeling

- Two types of secondary structures
  - Alpha-helices (cylindrical shape)
  - Beta-sheets (planar shape)

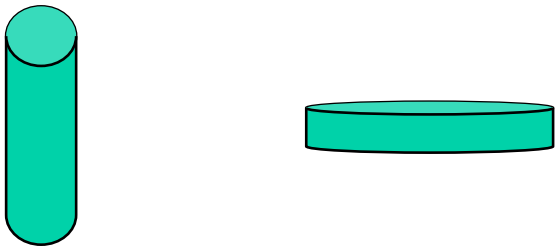


RDV P8 protein
(6.8Å)



RDV P8 protein
(PDB: 1UF2)

- Prior work
  - Alpha-helix detection (Jiang et al, 2001)
  - Beta-sheet detection (Kong et al, 2003)

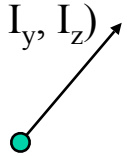


Prototypes for alpha-helices and beta-sheets

Step #5A: Secondary Structure Identification

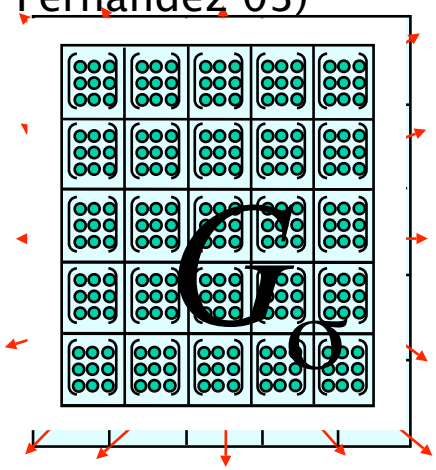
- Gradient tensor

(I_x, I_y, I_z)



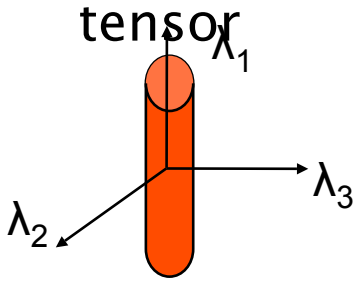
$$\begin{pmatrix} I_x^2 & I_x I_y & I_x I_z \\ I_x I_y & I_y^2 & I_y I_z \\ I_x I_z & I_y I_z & I_z^2 \end{pmatrix}$$

- Local structure tensor (Weickert'98, Fernandez'03)



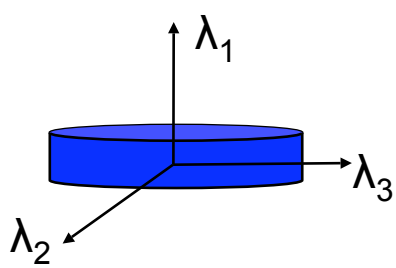
$$\begin{pmatrix} I_x^2 * G_\sigma & I_x I_y * G_\sigma & I_x I_z * G_\sigma \\ I_x I_y * G_\sigma & I_y^2 * G_\sigma & I_y I_z * G_\sigma \\ I_x I_z * G_\sigma & I_y I_z * G_\sigma & I_z^2 * G_\sigma \end{pmatrix}$$

- Property of local structure tensor



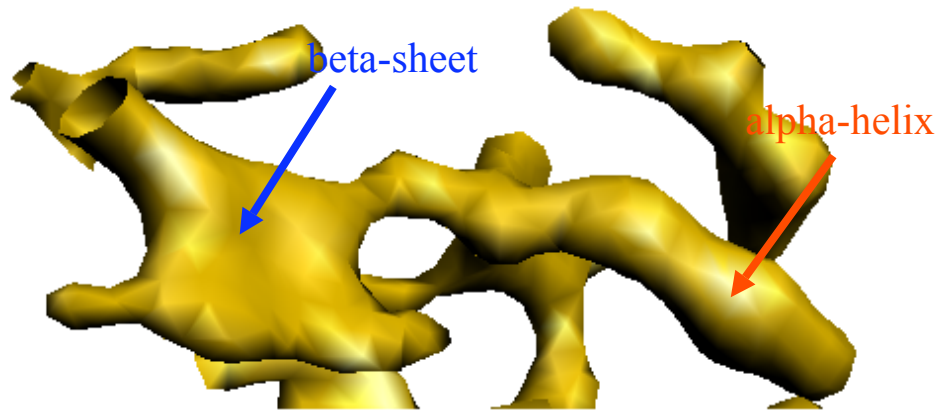
Line structure

$$\lambda_2 \approx \lambda_3 >> \lambda_1 \approx 0$$



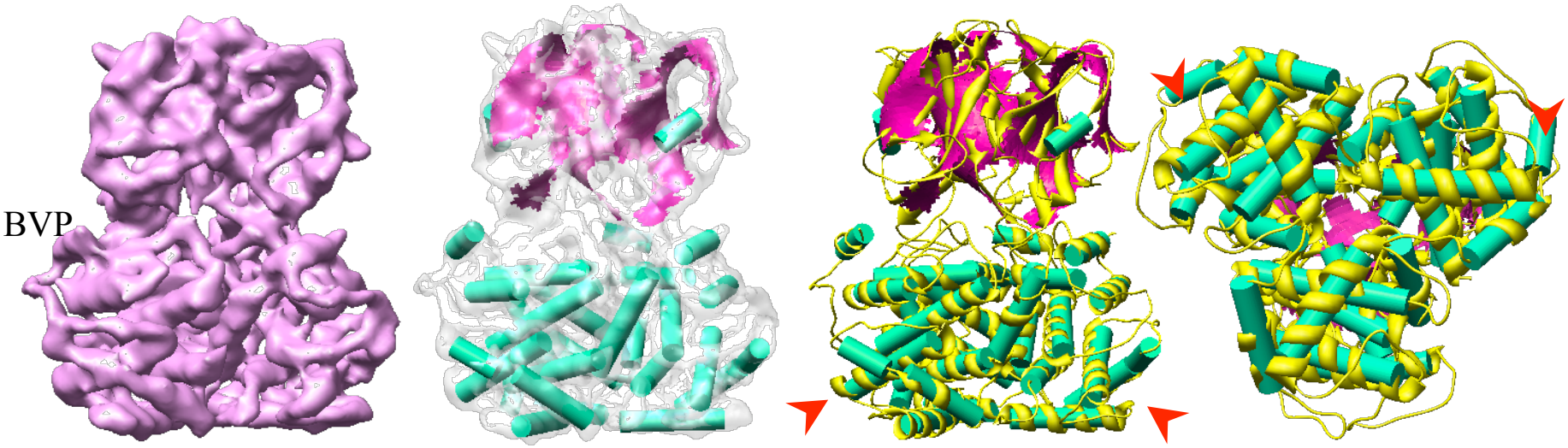
plane structure

$$\lambda_1 >> \lambda_2 \approx \lambda_3 \approx 0$$



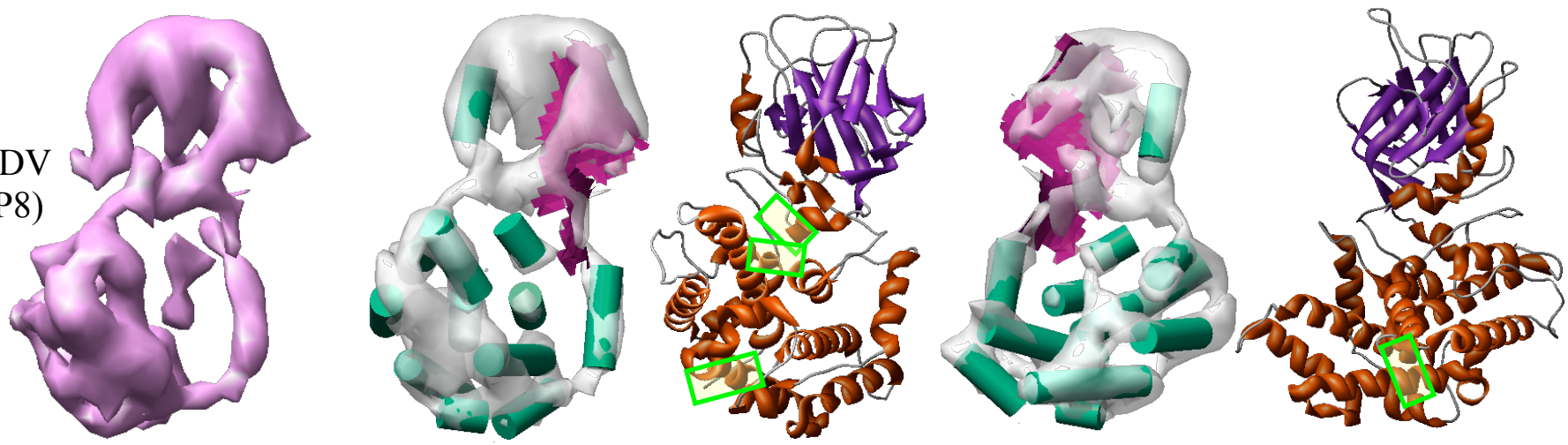
Step #5A: Secondary Structure Identification

1BVP



Blurred map (8Å) | Helices-Sheets with map | Helices-Sheets with PDB (all 27 helices detected)

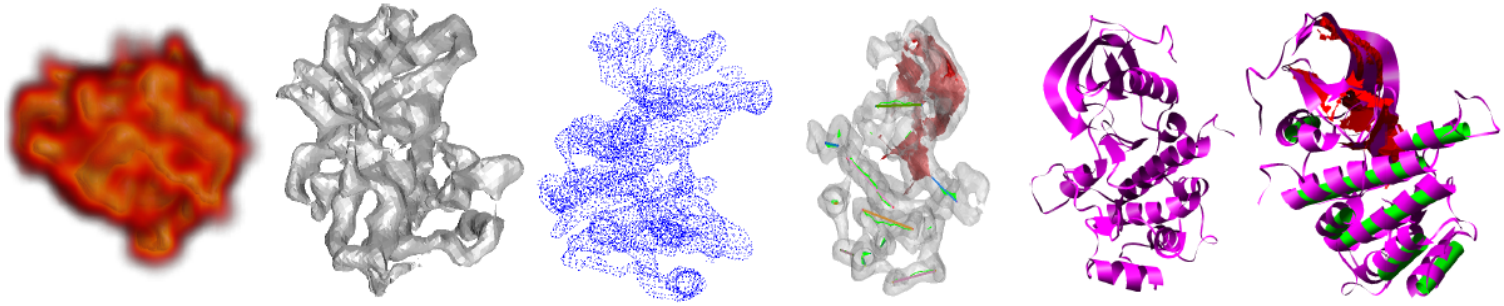
RDV (P8)



Zeyun, Bajaj, IEEE/ACM Trans. on CompBio&BioInf., 2007

12 out of 16 helices detected

Step #5B: Secondary Structure Elucidation



Step 1: Vor/Del(P) computation (using *CGAL*).

Goswami,Dey, Bajaj, ICVGIP'2006

Step 2: Identification of Interior Medial Axis M.

Step 3:

3.a: Identification of Critical points of distance function from Vor/Del(P).

3.b: Selection of Critical points only on M.

[By DGRS2005, Critical points are either near S or near M]

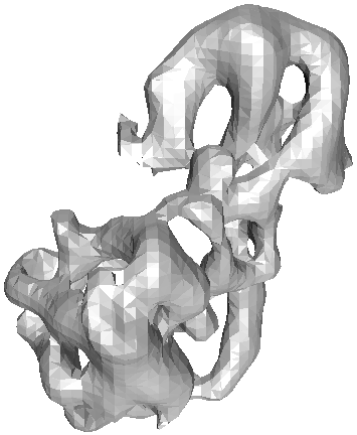
Step 4: Classification of Medial Axis via

4.a: U_1 – <u>Unstable Manifold of index 1 saddle</u> point on M

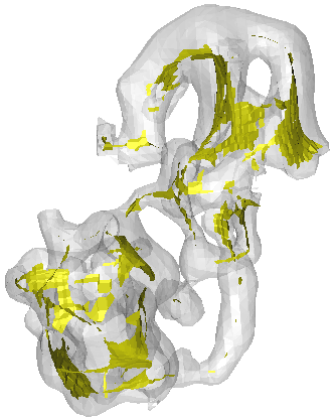
4.b: U_2 – <u>Unstable Manifold of index 2 saddle</u> point on M.

Step 5: Width Test to select the subsets of U_1 (β-sheets) and U_2 (α-helices).

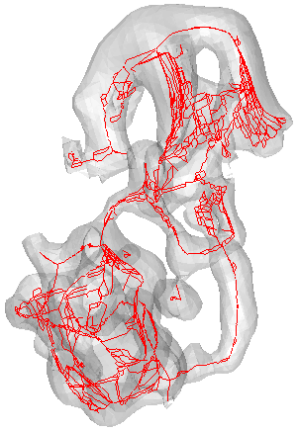
Secondary Structure of RDV Outer Capsid Coat Protein P8



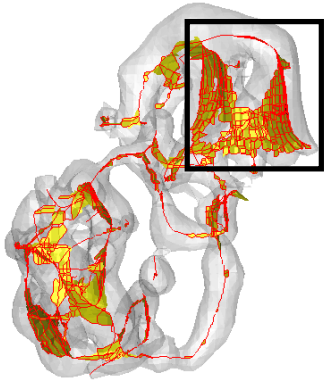
Surface



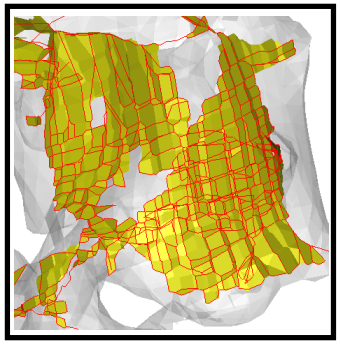
U_1



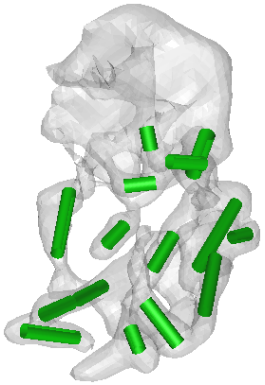
U_2



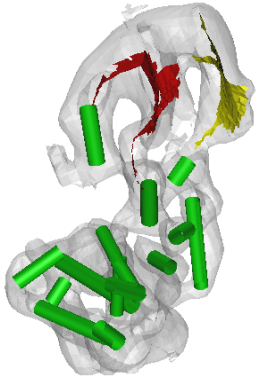
U_1 and U_2



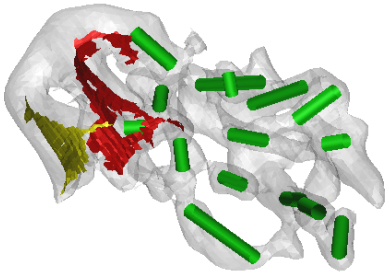
Closeup



Helices



Helices and Sheets



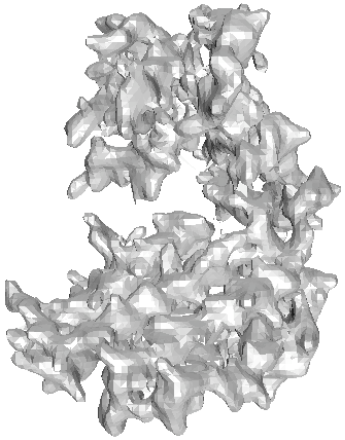
Another View

Ribbon Diag. of PDB

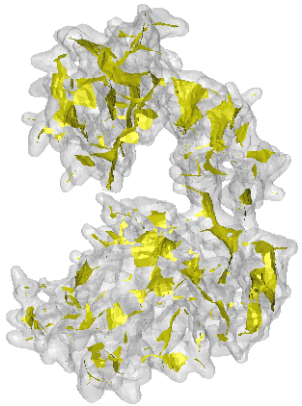
α-helices 15/16
β-sheets 3/3

3D EM map of P8 segmented out from cryo-EM map of Rice Dwarf Virus (RDV) at 6.8 A resolution

Secondary Structure of Bacterial Chaperonin GroEL

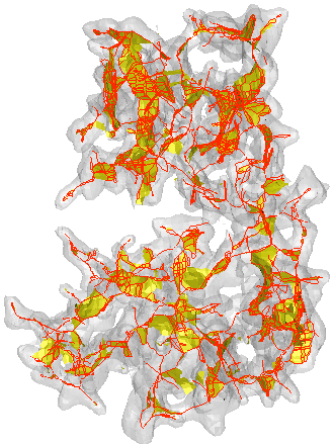


Surface

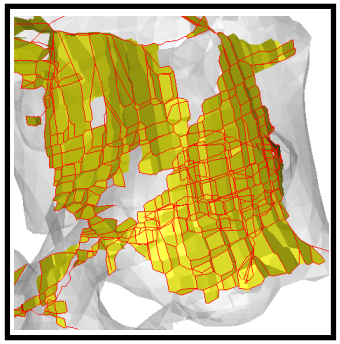


U_1

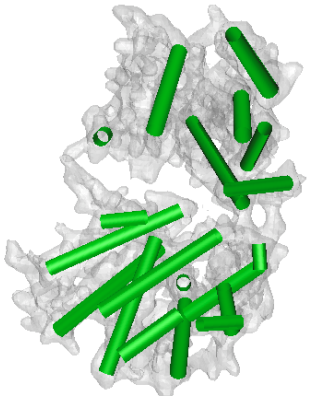
U_2



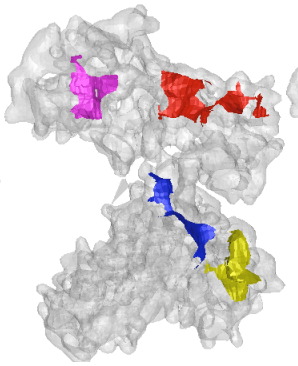
U_1 and U_2



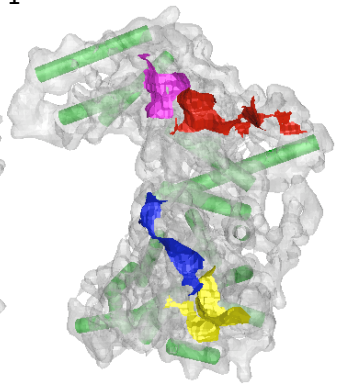
Closeup



Helices

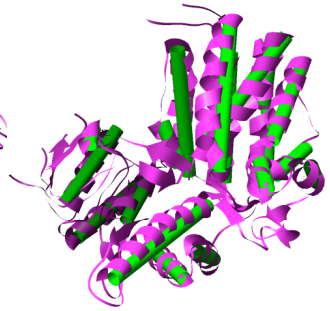


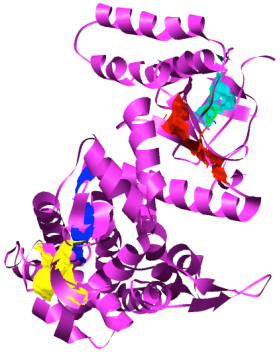
Sheets



Helices and Sheets

Ribbon Diag. of PDB



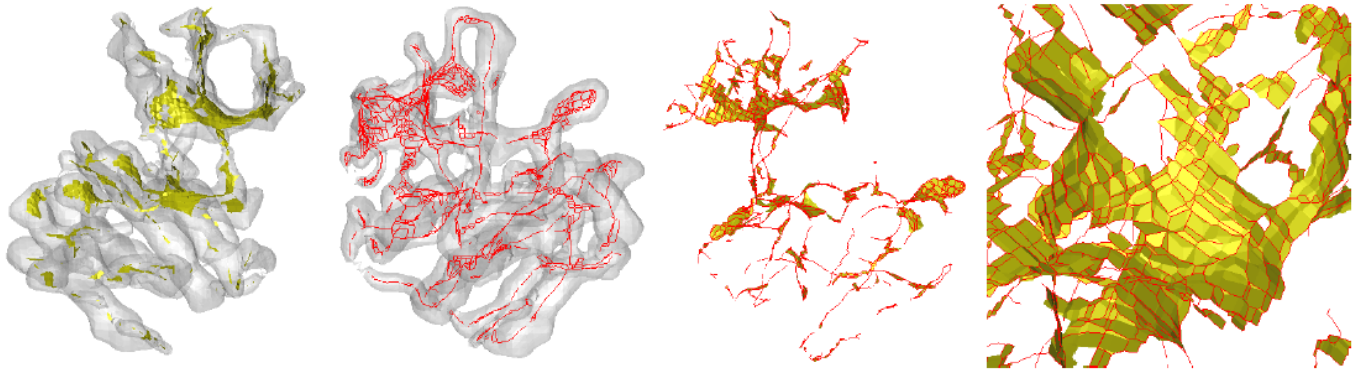


Superimposed

α-helices 15/16

β-sheets 3/3

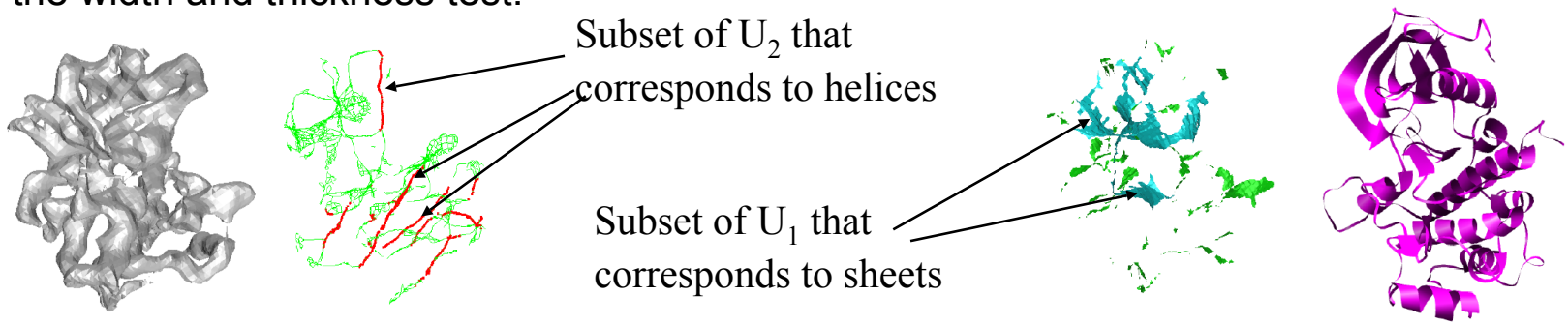
Width Test



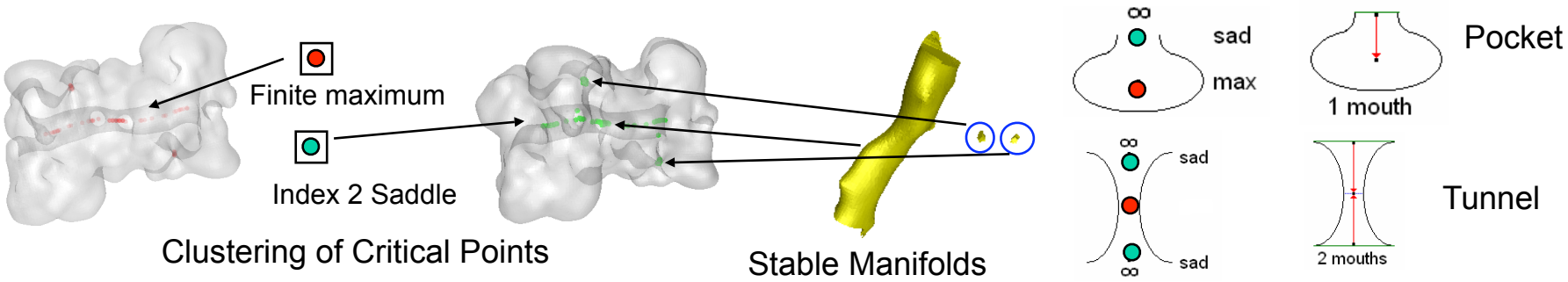
U_1 U_2 U_1 and U_2 Closeup: U_1 is bounded by U_2

- U_1 and U_2 give superset of sheets and helices.
- α-helix: width 2.5 A and pitch-length 1.5 A [Branden-Tooze]
- β-sheet: thickness 1.5 A [Branden-Tooze]

- h_P values of Voronoi elements constituting U_1 and U_2 help select the subset that passes the width and thickness test.



Molecular Surface Curation & Detection of Pockets/Tunnels

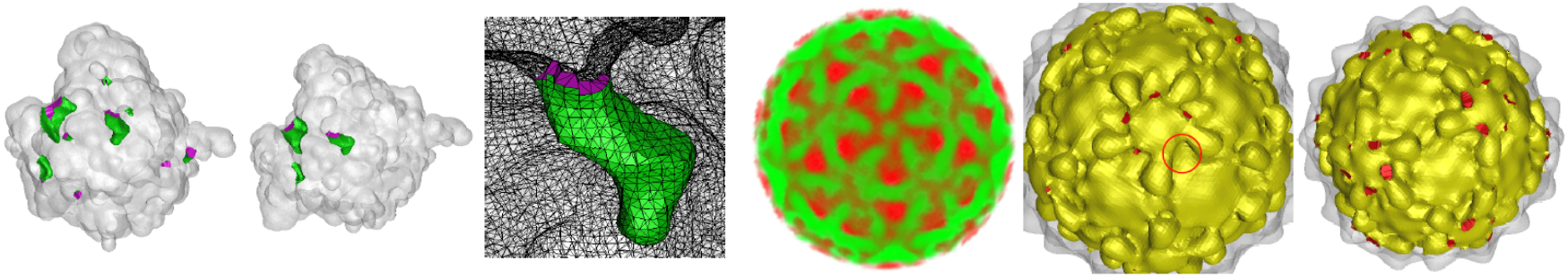


- Build adjacency relation between two finite maxima if they share a common boundary.

- Compute the Transitive Closure of this relation. This gives a clustering of the maxima and the index-2 saddles.

- Compute S_3 for each cluster.

Nodavirus

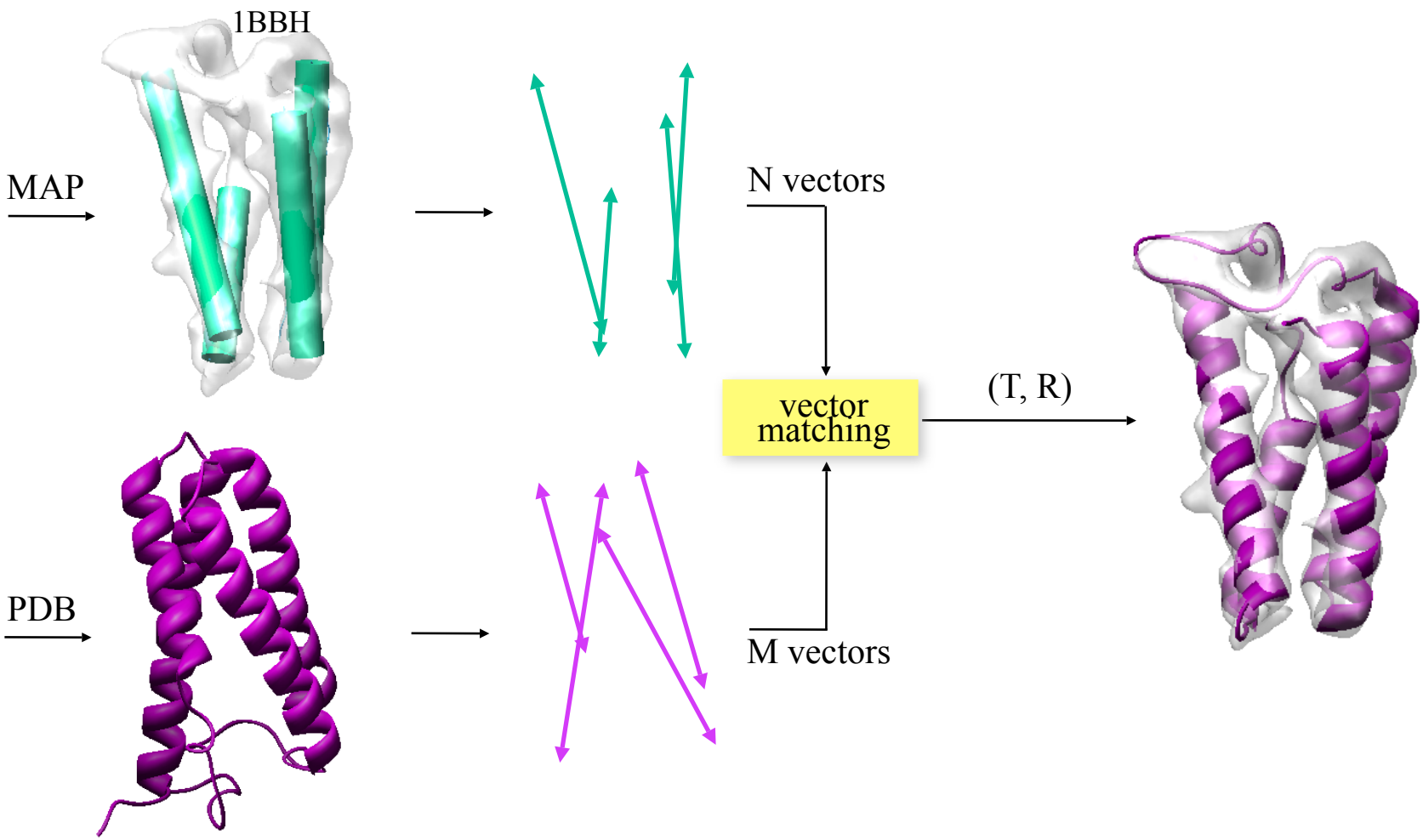


AcetylCholineEsterase

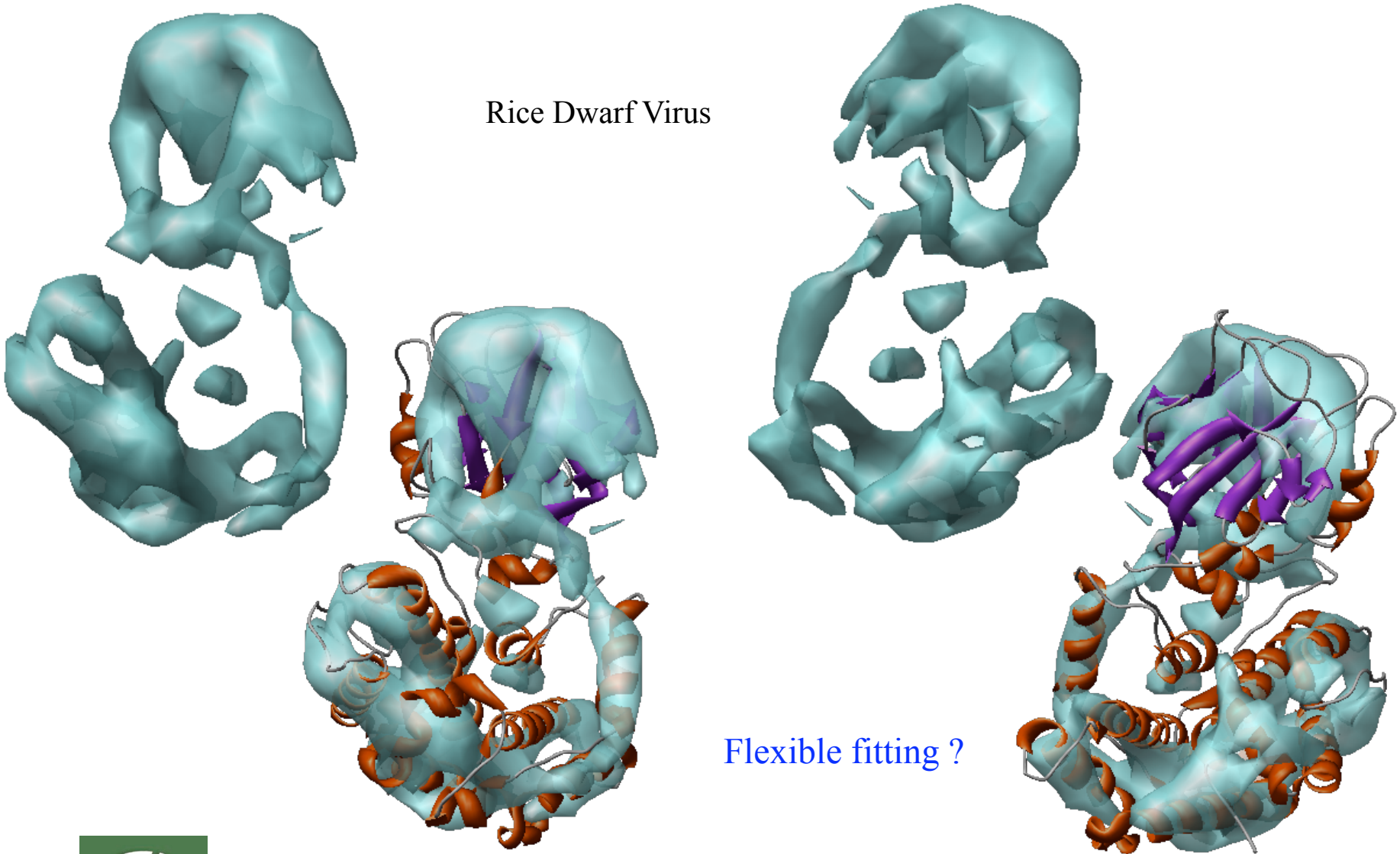
Bajaj, Gillette, Goswami
TopoInVis'2007

Step #6: Structure Fitting/Modeling

- PDB-based approach— vector matching



Step #6: Structure Fitting/Modeling

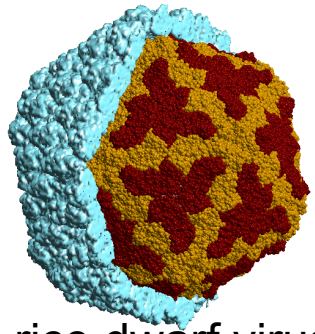


Analytic Molecular Models with extensions to Flexible Models/Coarse Grained Models

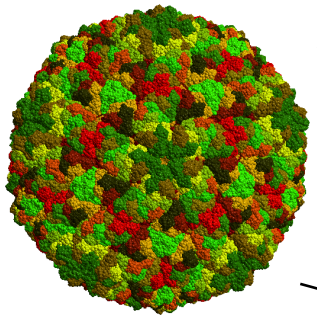
Center for Computational Visualization
Institute of Computational and Engineering Sciences
Department of Computer Sciences
University of Texas at Austin

January 2008

Rice Dwarf Virus Capsid (1UF2)

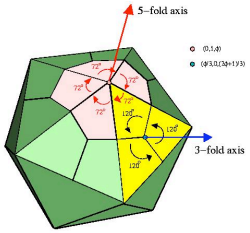


rice dwarf virus

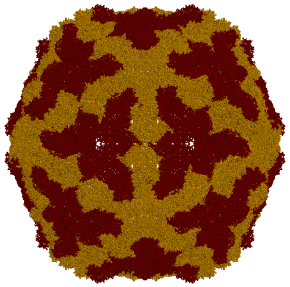


P8

outer layer capsid

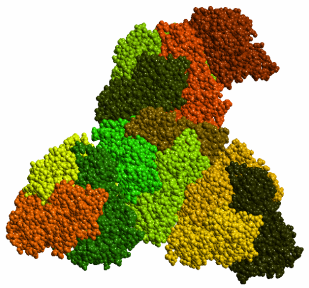


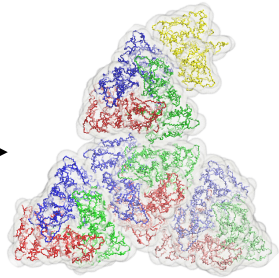
global symmetry



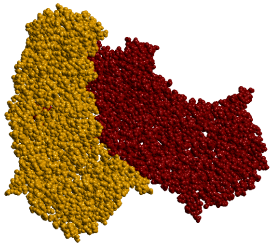
P3

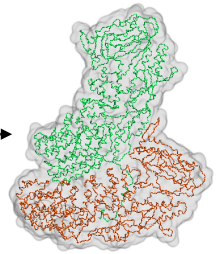
inner layer capsid





asymmetric subunit

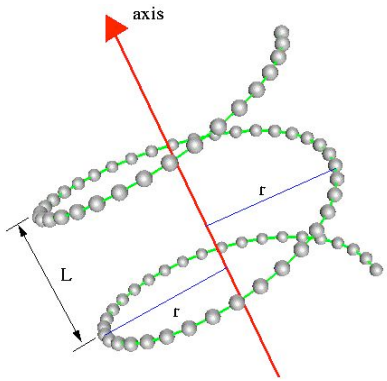


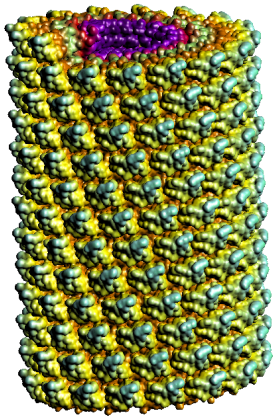


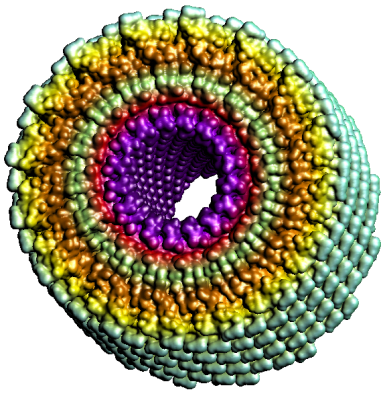
dimer

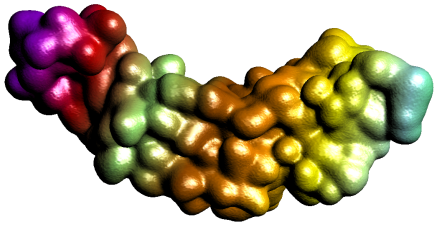
Computational Visualization Center
Institute for Computational and Engineering Sciences
Department of Computer Sciences
University of Texas at Austin

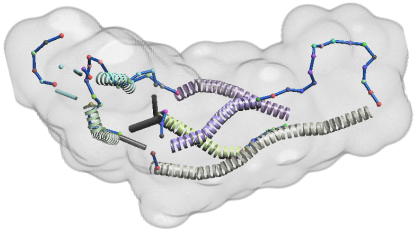
Helical: Tobacco Mosaic Virus (1EI7)

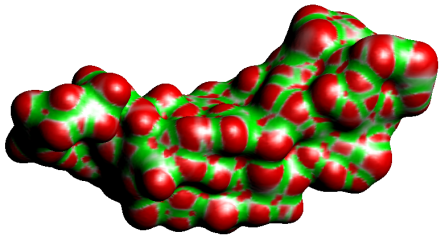




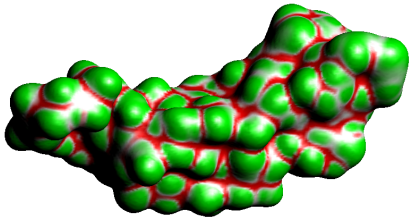








Mean

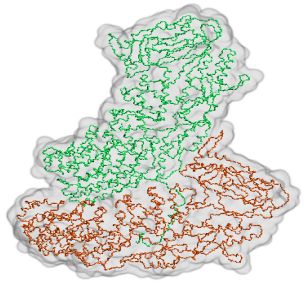


Gaussian

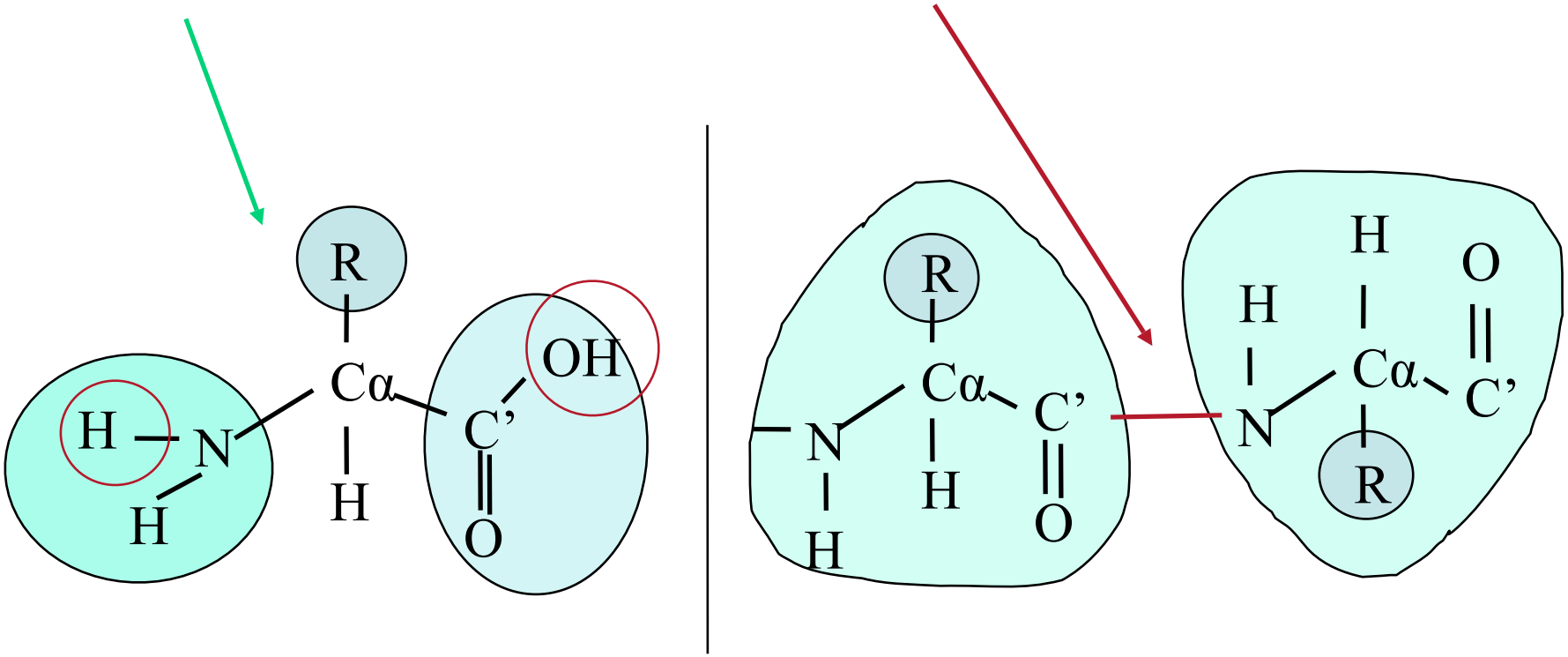
C. Bajaj, Modeling of Viruses, Chap in Modeling Biology, MIT Press, (2007)

Center for Computational Visualization
Institute of Computational and Engineering Sciences
Department of Computer Sciences
University of Texas at Austin

Proteins



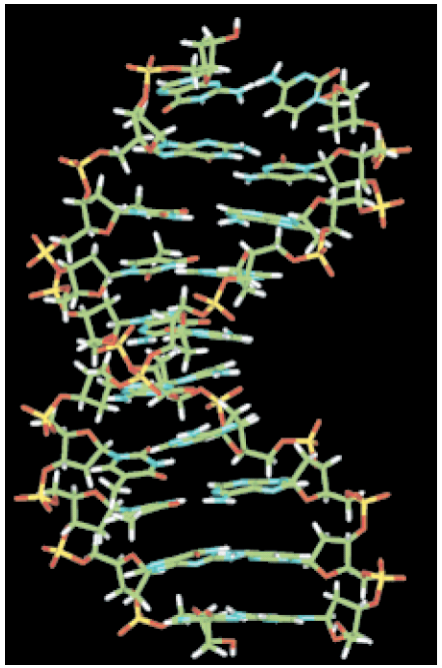
- Amino acids contain an amide, a residue and a carboxyl group
- Proteins are polypeptide chains, made from amino acids combined via peptide bonds.



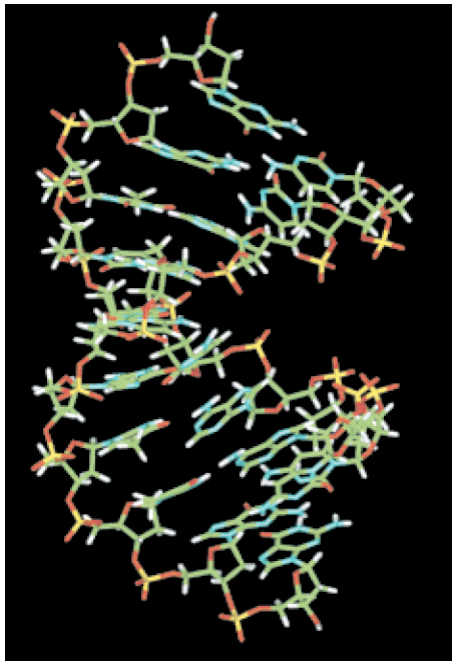
Computational Visualization Center
Institute for Computational and Engineering Sciences
Department of Computer Sciences
University of Texas at Austin

Stability of Macromolecules in Solution

d(CCAACGTTGG)



B-form DNA in water

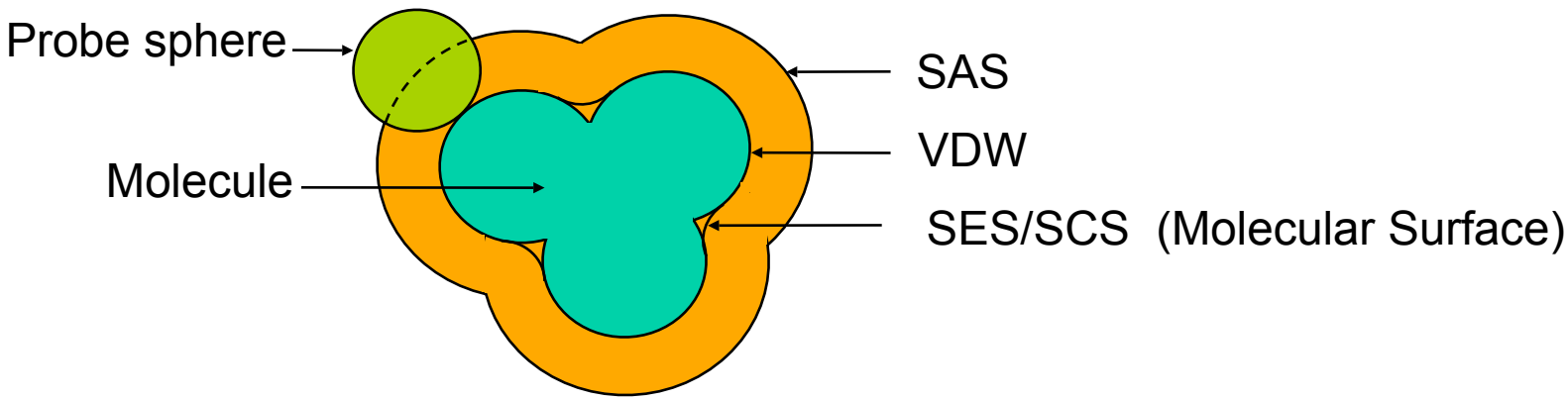


A-form DNA in ethyl

Tsui and Case, 2000

Molecules in Solvent (Implict Model)

- Solvent molecule modeled as a sphere. Water: radius 1.4A

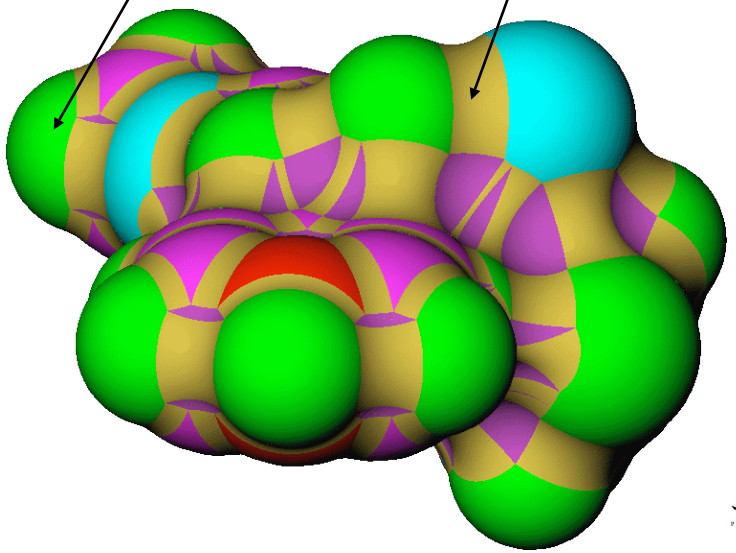


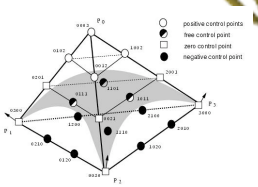
SAS: solvent accessible surface: locus of probe center
VDW: van der Waals surface: Union of spheres with VDW radii
SES/SCS: solvent excluded/contact surfaces

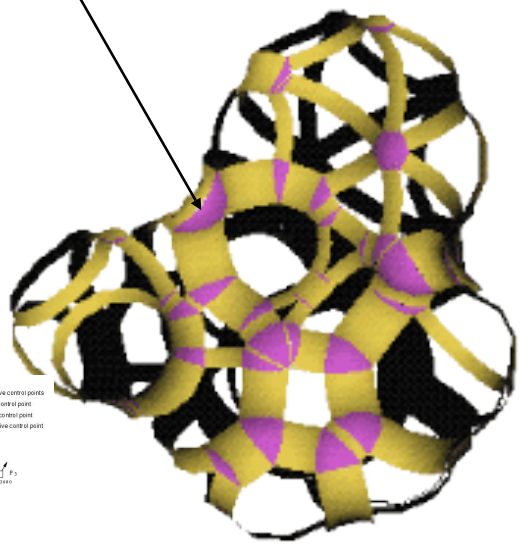
Molecular Models I: Implicit Solvation Surface for the Hard Sphere Model

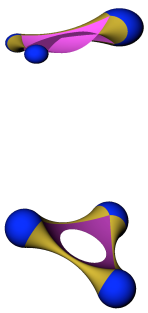
Lee-Richard (LR) surface is decomposed into three kinds of patches:

convex spherical, toroidal, and concave spherical patches









The LR surface can be represented as A-patches and NURBS

Bajaj et al, Discrete Appl Math. (2003), 23-51

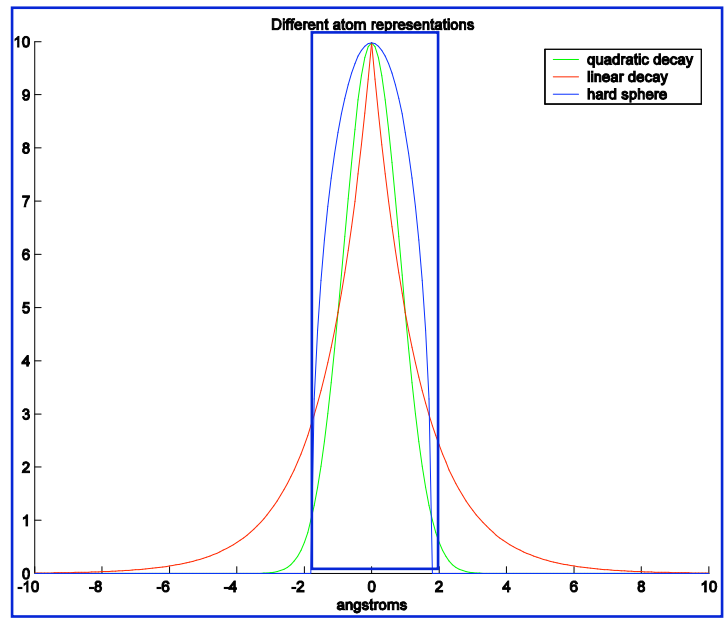
Problem with LR: Singularities

Molecular Models II: Analytic Atomic Shape Parameters

- **Isotropic Quadratic Kernel:** $G_i(x) = e^{-\frac{\beta}{r_i^2}\left((x-x_i)^2 - r_i^2\right)}$

- **Isotropic Linear Kernel:** $G_i(x) = e^{-\beta\left(|x-x_i| - r_i\right)}$

  - where
    - The decay β controls the shape of the Gaussian function.
    - The van der Waal's radius is $\mathbf{r_i}$
    - The center of the atom is $\mathbf{x_c}$.



$\beta = 2.3$, $r = 1.8$ Å

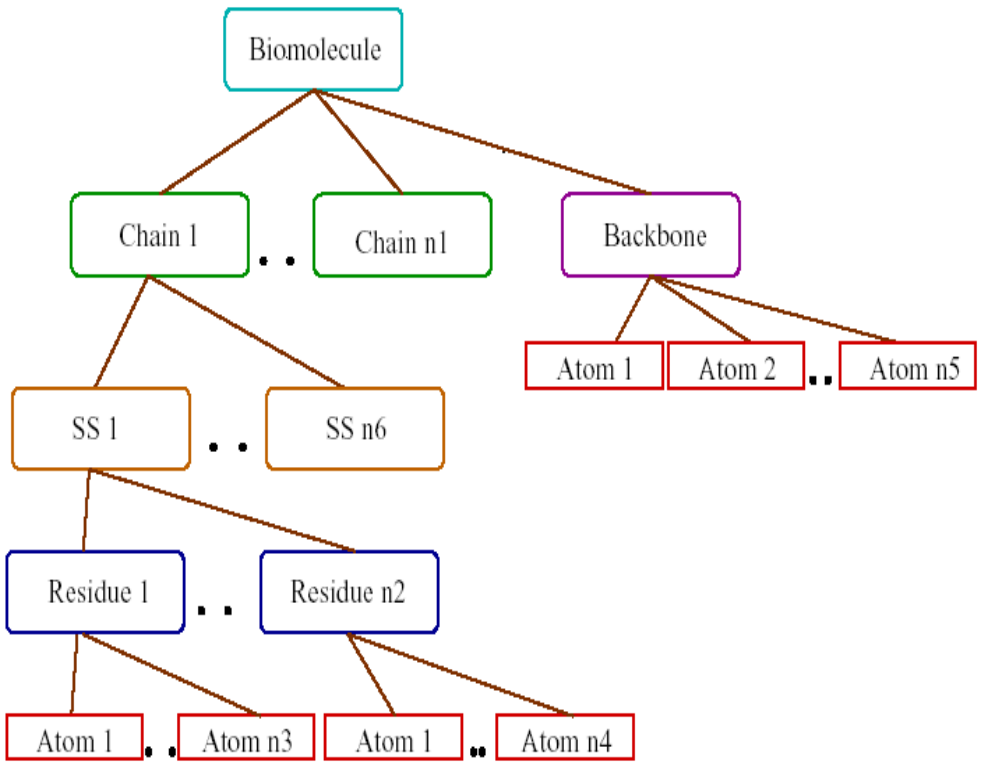
β values suggested in (Boys 50), (Grant Pickup 99)

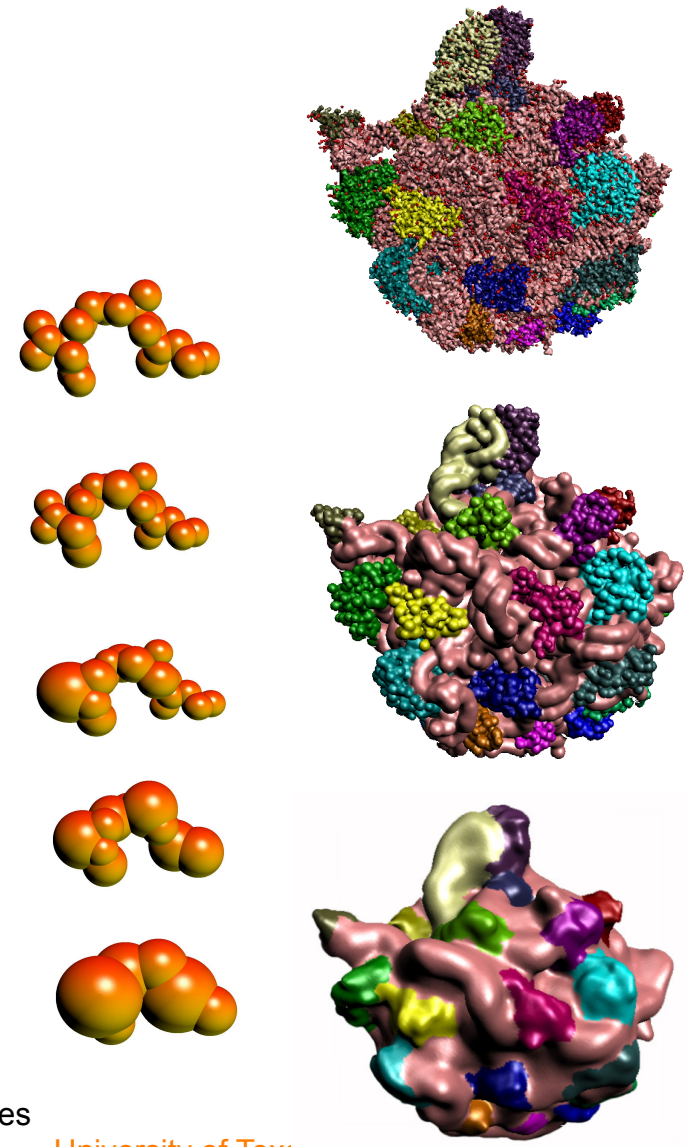
- **Anisotropic Kernels**

Modeling Flexibility: FCC Cluster Hierarchy

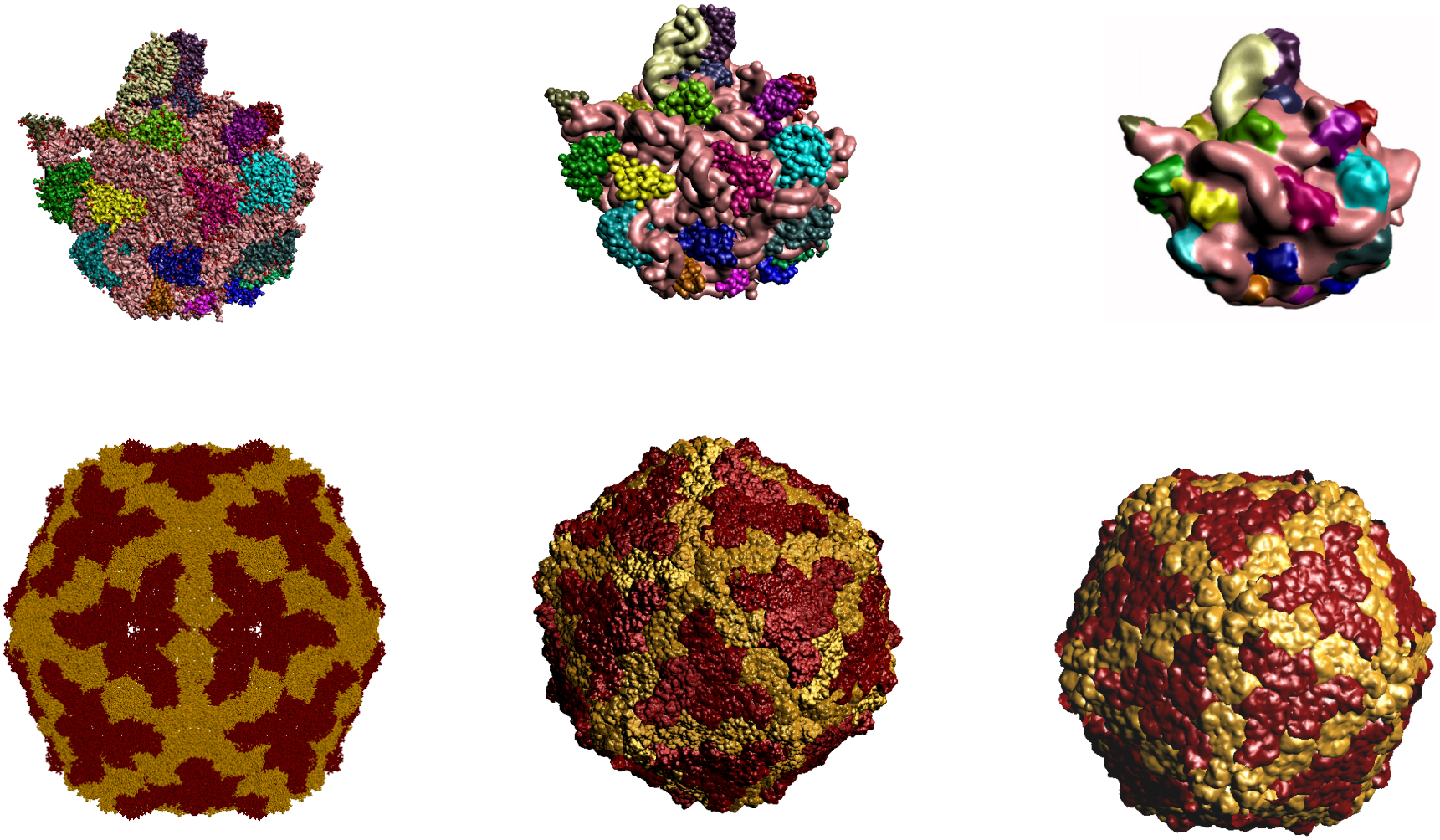
Clustering of atoms is based on

- biochemical units
- preserving molecular shape feature



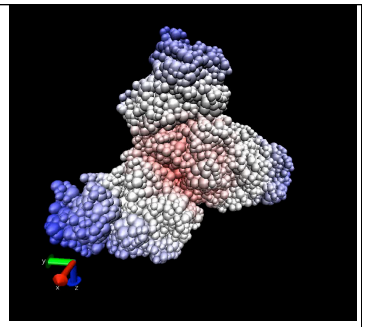


FCC Models of Ribosome and Viral Capsid



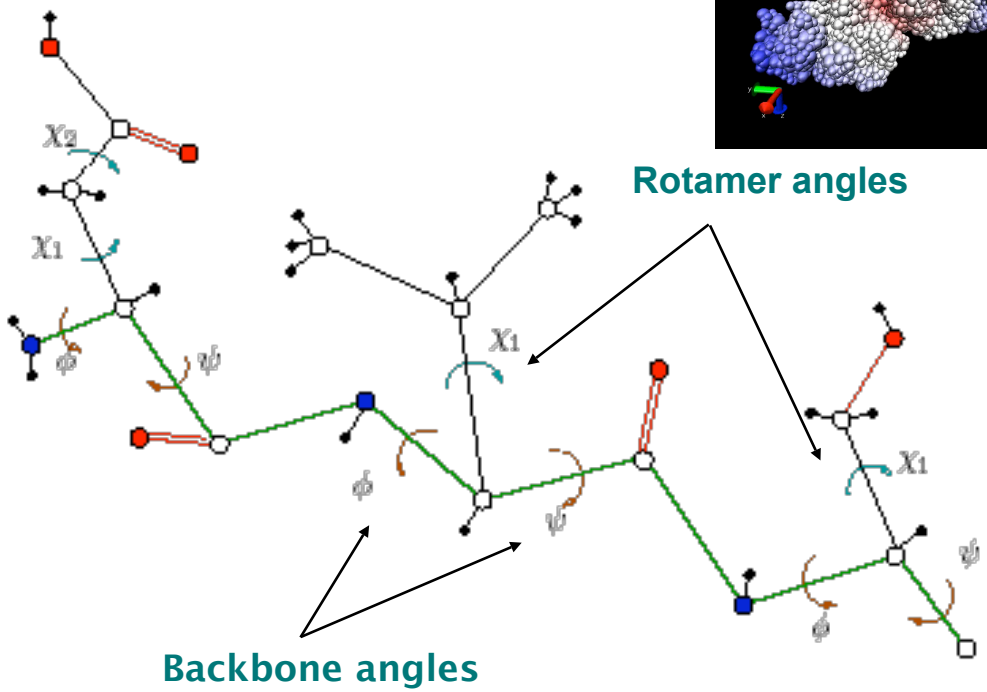
Computational Visualization Center
Institute for Computational and Engineering Sciences
Department of Computer Sciences
University of Texas at Austin

Modeling Flexibility

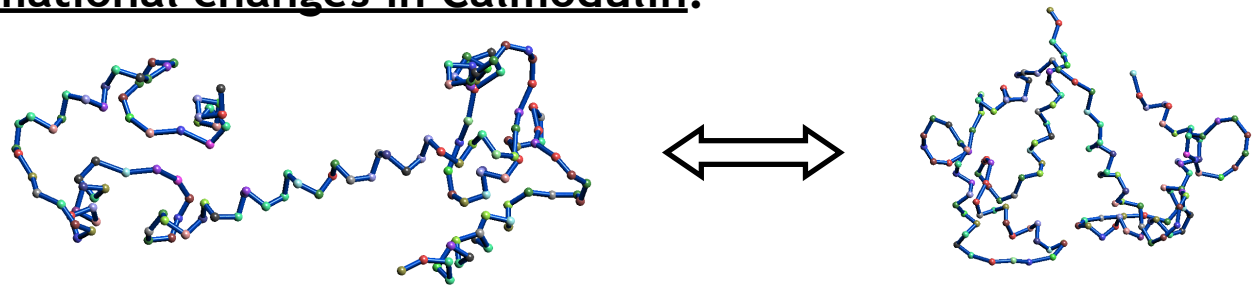


Conformational changes occur due to

- changes in torsional angles: φ, ψ and χ
- hinge-type bending and shearing movements

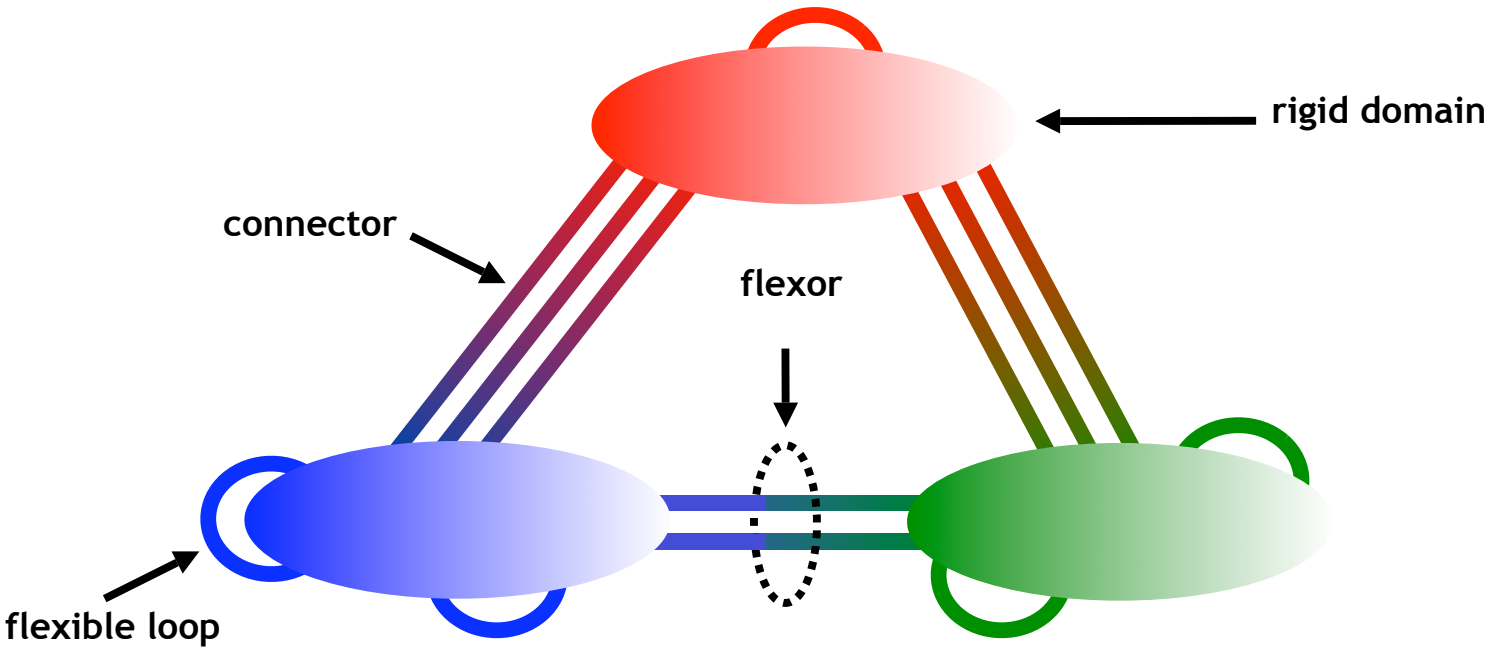


Conformational changes in Calmodulin:



open (1CLL.PDB) closed (2BBM.PDB)

Modeling Flexibility:
The Flexible Chain Complex (FCC)



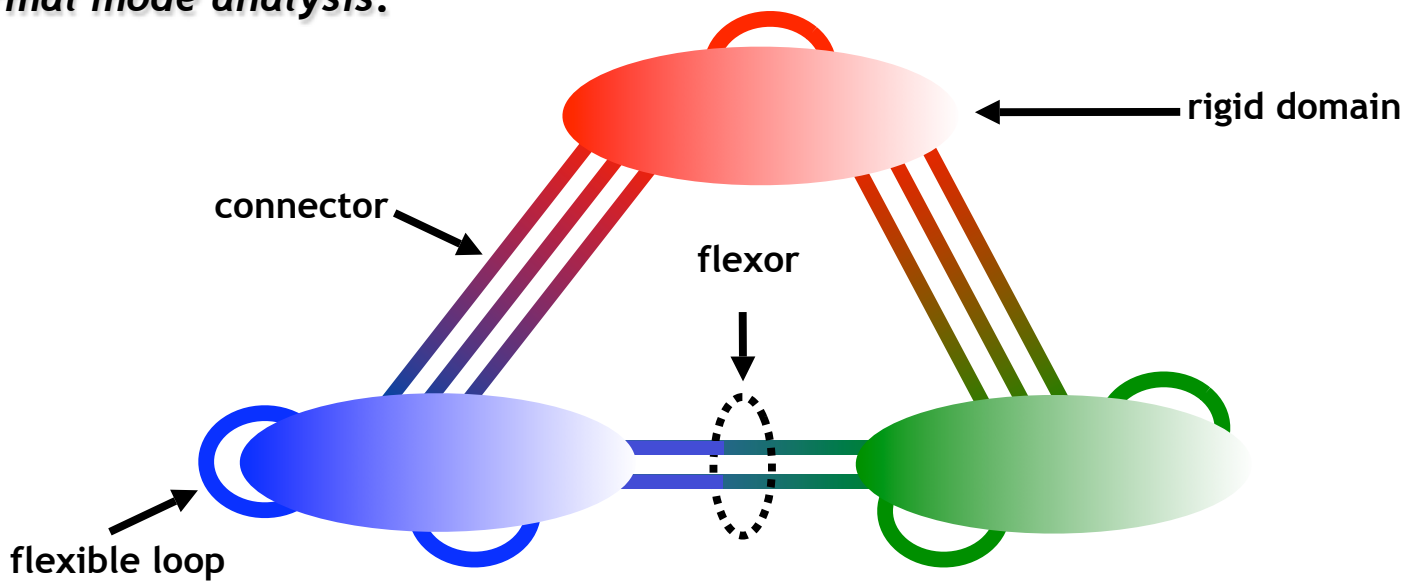
Flexible Chain Complex (FCC)

Hierarchical

By relaxing the threshold of rigidity domains can be hierarchically decomposed into subdomains

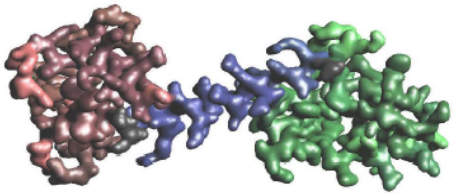
Modeling Flexibility: Identifying Domains using Normal Mode Analysis

Domains are identified by capturing domain movements using *normal mode analysis*.

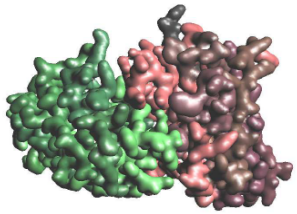


Computing normal modes:

- Minimum energy conformation
- Hessian of internal energy
- Eigen-decomposition gives motion



Calmodulin : 2BBM



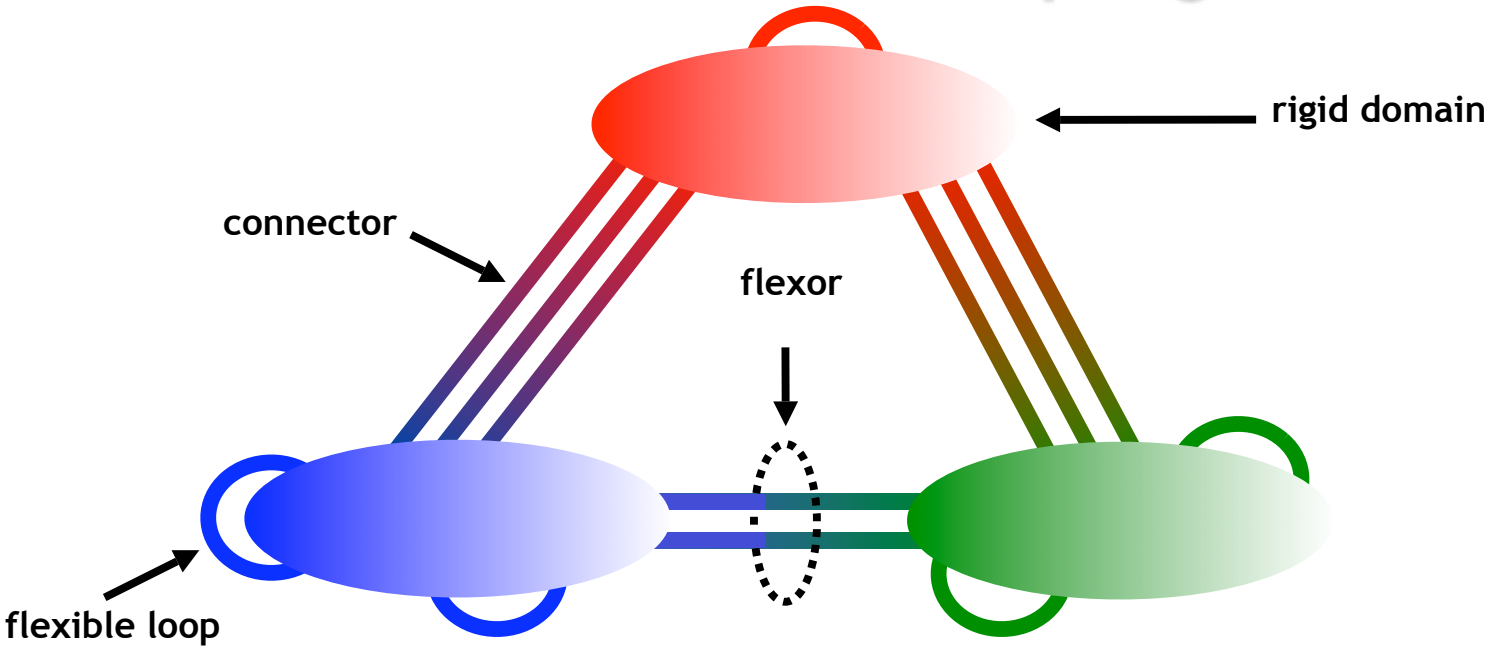
Immunoglobulin: 1VFB

Modeling Flexibility: Motions at Flexors

We apply three types of motions at flexors:

- Shear: a lateral movement applied along the interface of two domains if
  - they share a large number of short connectors, and / or
  - they have a large interface area
- Bending: applied when the domains are connected by at least one connector
  - applied around three orthogonal axes
  - hinge point: geometric center of the shortest connector
  - primary axis: normal to the plane containing the geometric centers of the two domains and of the shortest connector
  - secondary axis: orthogonal to the primary axis and the line connecting the two domain centers
  - third axis: line through the domain centers
- Twist: applied when only a single connector exists between the domains
  - the torsion angles along the backbone are updated

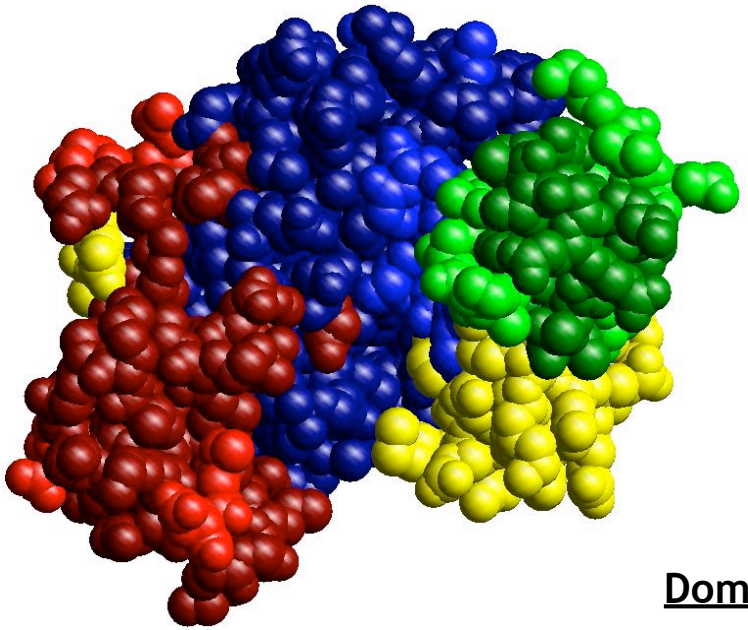
Modeling Flexibility: Conformation Sampling

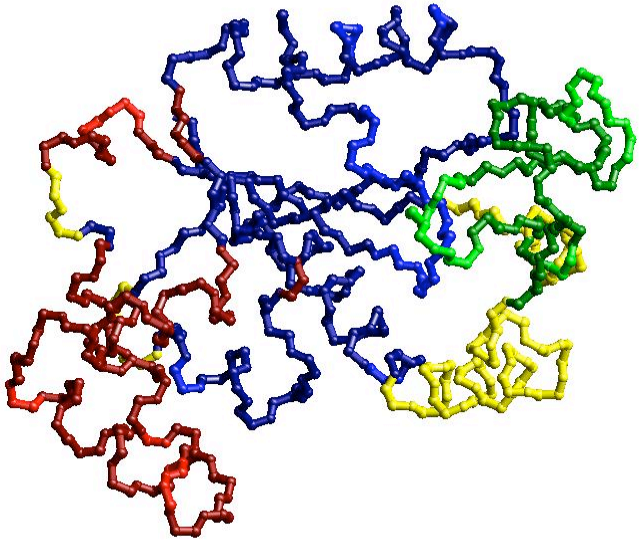


Suppose we want to generate N global conformations.

- conformations are allocated to flexors
- each flexor f is assigned a weight w_f (= sum of the domain weights)
- number of conformations allocated to f is $N^{w_f / \sum w_f}$
- conformations are generated by applying motions to the flexors

Modeling Flexibility: Domain Decomposition of Adenylate Kinase (4AKE.PDB)





- Domains: red / green / blue
- Flexible Loops: light domain color
- Connectors: yellow

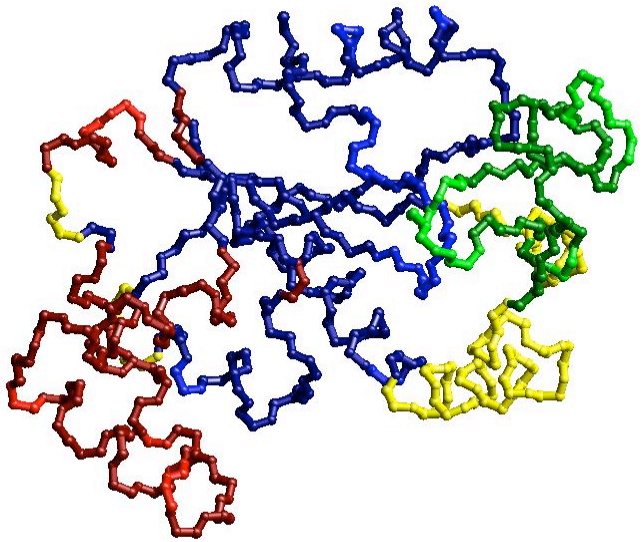
Domains:

- Core: 93 residues, 12 segments, 12 loops
- AMP-binding: 52 residues, 9 segments, 4 loops
- ATP-lid: 36 residues, 4 segments, 3 loops

Connectors and Interface Area:

- Core and AMP-binding: 3, 508 Å
- Core and ATP-lid: 2, 2 Å

Modeling Flexibility: Motion Graph of Adenylate Kinase

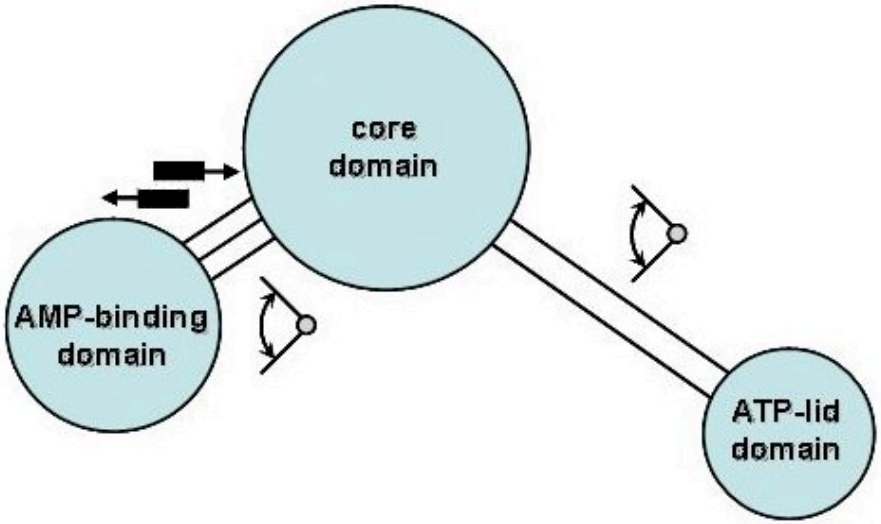


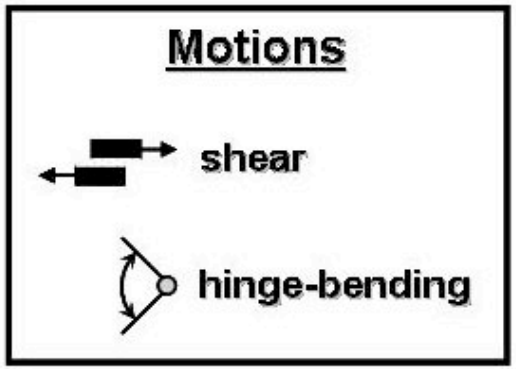
Domains:

- **Core: 93 residues, 12 segments, 12 loops**
- **AMP-binding: 52 residues, 9 segments, 4 loops**
- **ATP-lid: 36 residues, 4 segments, 3 loops**

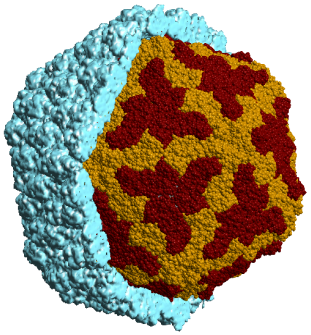
Connectors and Interface Area:

- **Core and AMP-binding: 3, 508 Å**
- **Core and ATP-lid: 2, 2 Å**

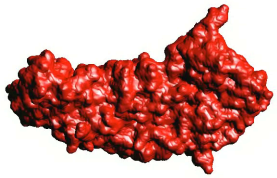


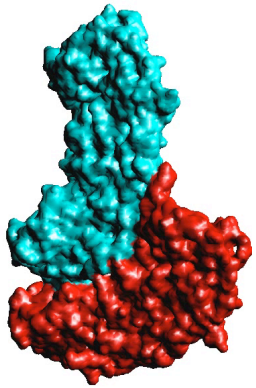


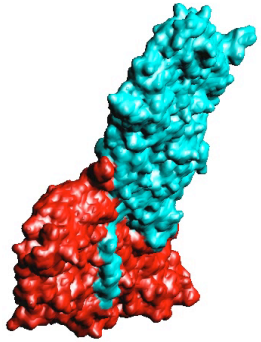
Coarse Grained Flexibility of RDV capsid Proteins



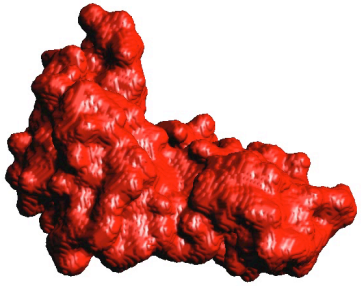
P3

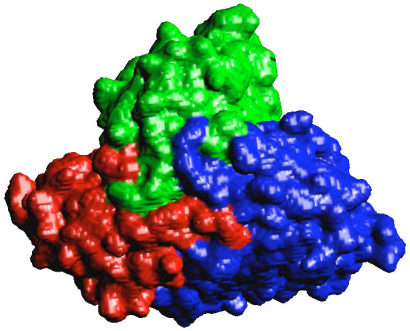


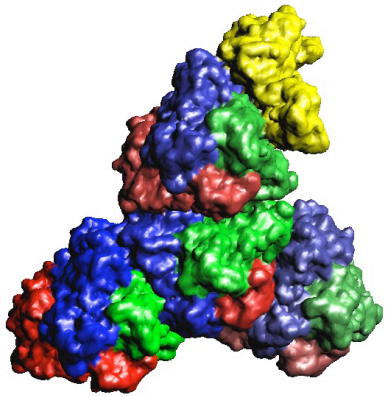




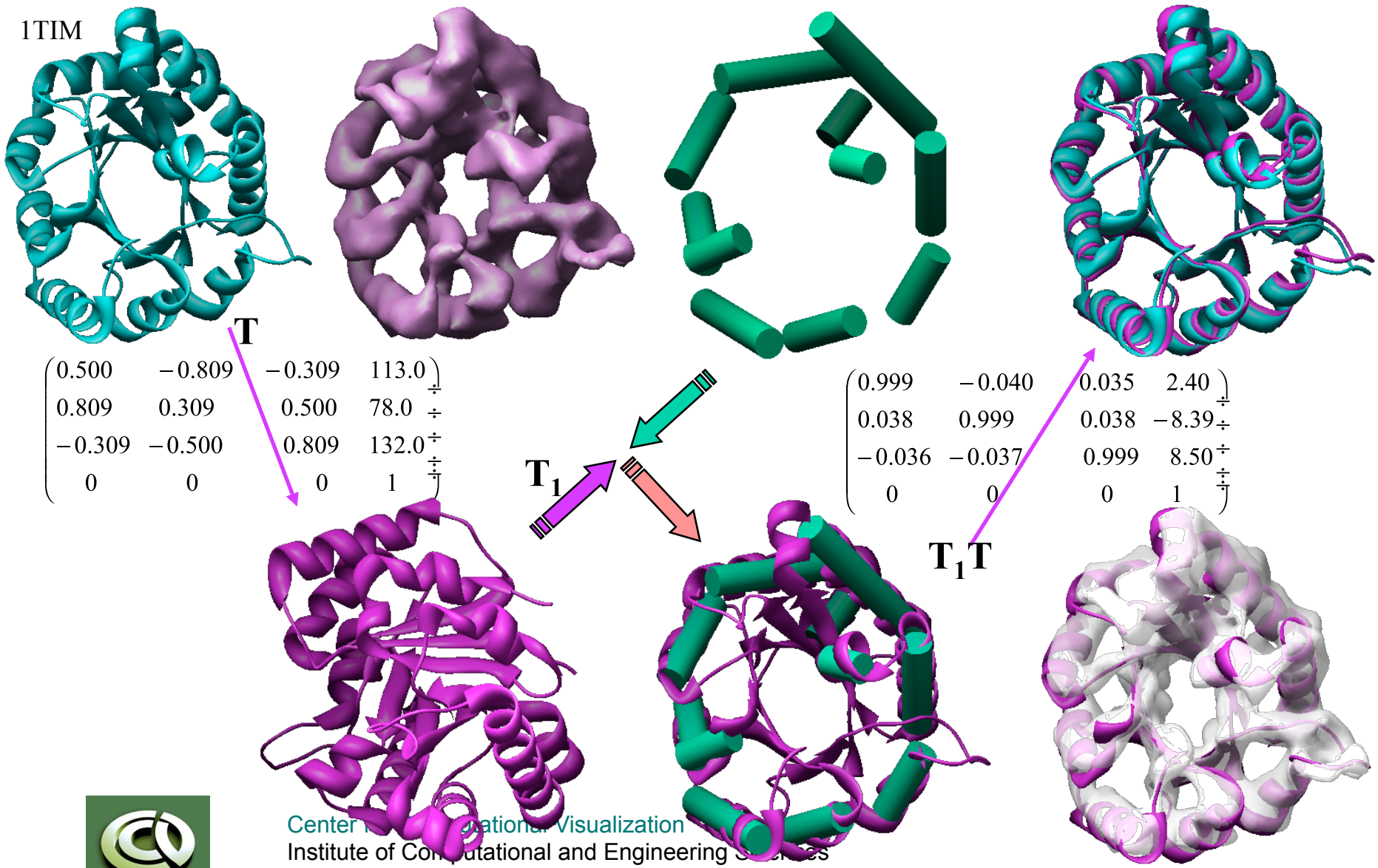
P8







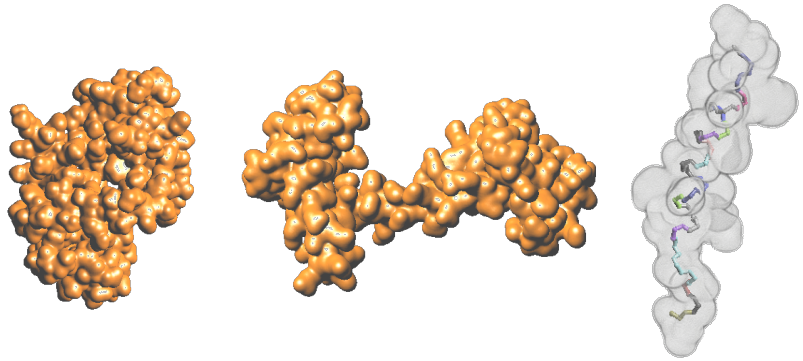
Low Res. Flexible Structure Fitting in 3DEM



Search and Scoring for Flexible Docking (F3Dock)

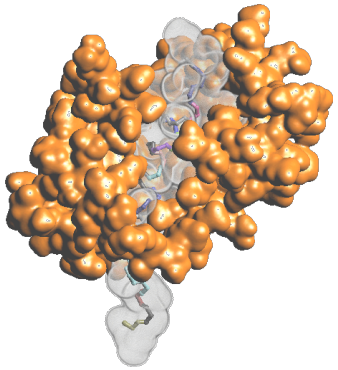
Docking score is based on

- Shape complementarity:
  - Lennard Jones potential
  - Curvature complementarity
- Electrostatic interactions
  - Long range
  - Short range
    - Hydrogen bonds
    - Salt bridges
  - Hydrophobic patches
- Desolvation free energy
- Free Energy change due to conformational changes



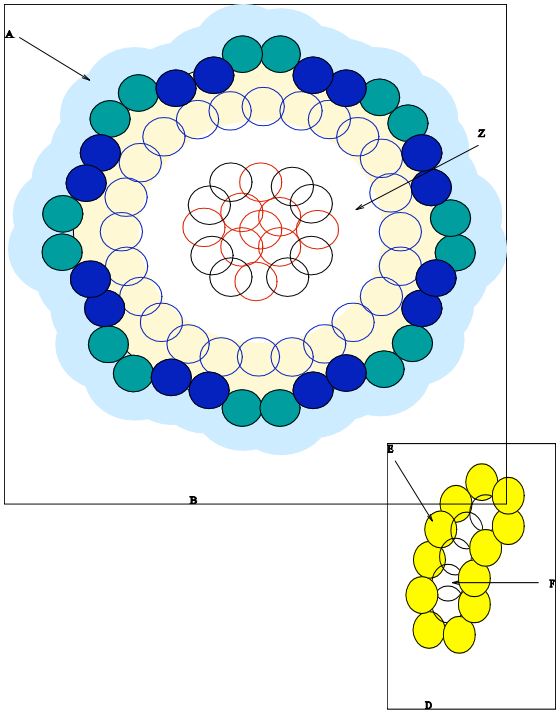
Closed Open Peptide

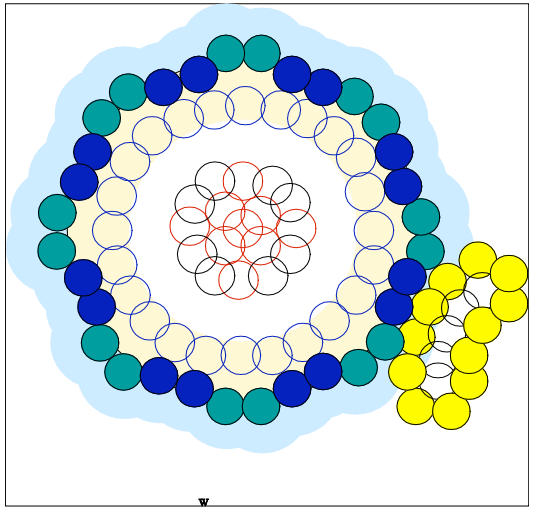
Calmodulin
(1CLL.PDB)



Calmodulin complexed with peptide
(2BBM.PDB)

Fast Computation of Protein Binding Energetics and Forces





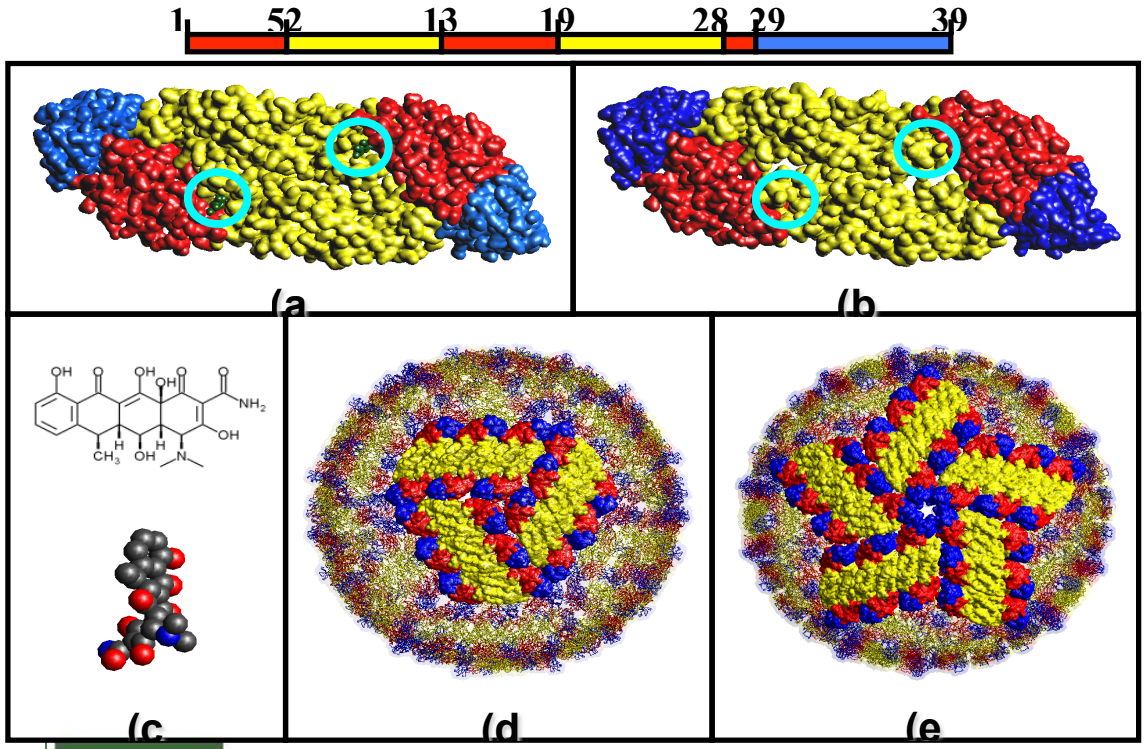
Center for Computational Visualization
Institute of Computational and Engineering Sciences
Department of Computer Sciences
University of Texas at Austin

January 2008

Binding Energy of Macromolecules

In structure based drug design, binding of a drug (ligand) to a receptor (protein/nucleic acid), usually causes the ligand to either enhance or inhibit the activity of the receptor.

Binding energy: $\Delta G_{\text{bind}} = G_{\text{protein+ligand}} - (G_{\text{protein}} + G_{\text{ligand}})$



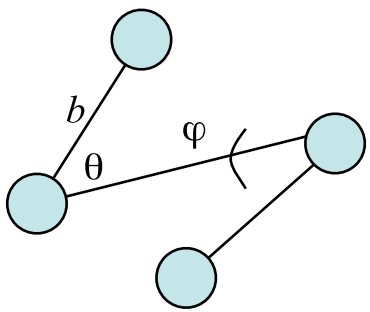
Docking Inhibitors in a Flexible Pocket of Dengue Virus Envelope Protein

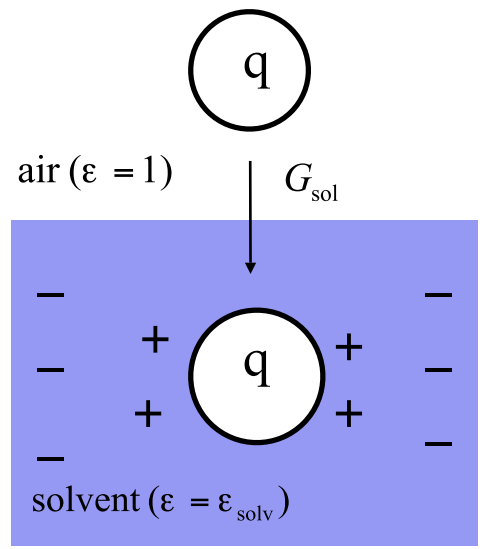
Free Energy of a Single Molecule in Solvent

Total free energy : $G = E_{MM} + G_{\text{sol}} - TS$

$$E = \underbrace{E_b + E_\theta + E_\varphi}_{\text{bonded}} + \underbrace{E_{vdw} + E_{elec}}_{\text{non-bonded}}$$

$$G_{\text{sol}} = G_{\text{cav}} + G_{\text{vdw}} + \boxed{G_{\text{pol}}}$$





Free Energy of a Single Molecule in Solvent

- bonded

$$E_b = \sum_b k_b (r_b - r_b^0)^2$$ r, r_0 : covalent bond variation

$$E_\theta = \sum_a k_a (\theta_a - \theta_a^0)^2$$ θ, θ_0: valence angle variation

$$E_\varphi = \sum_t k_t (1 + \cos n(\varphi_t - \varphi_t^0))$$ φ, φ_0: torsion angle variation

- non-bonded

$$E_{vdw} = \sum_{i<j} \frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^6}$$ C, D : Lennard-Jones parameters

r_{ij} : distance between atoms

$$E_{elec} = \sum_{i<j} \frac{q_i q_j}{r_{ij}}$$ q : atomic charge

- $G_{\text{cav}} + G_{\text{vdw}} = \gamma S$ γ, S : surface tension and surface area

How to Compute G_{pol} ?

- Poisson-Boltzmann (PB) Theory

$$G_{\text{pol}} = \frac{1}{2}\int [\phi_{\text{solvent}}(\mathbf{r}) - \phi_{\text{air}}(\mathbf{r})]\rho(\mathbf{r})\, dV$$

$$-\nabla \times [\varepsilon(\mathbf{r})\nabla\phi(\mathbf{r})] = 4\pi\rho(\mathbf{r}) + 4\pi\lambda(\mathbf{r})\sum_{j=1} c_j^{\infty} q_j \exp(-q_j\phi(\mathbf{r})/k_B T)$$

finite difference, finite element -- APBS, DELPHI, ...

- ASMS Boundary Element [Bajaj, Chen 2007]

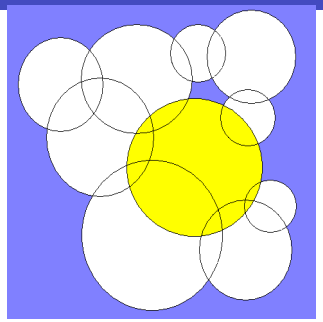
ε	dielectric constant
ϕ	electrostatic potential
ρ	solute charge density
λ	ion accessibility parameter
c_j^{∞}	ion bulk concentration
q_j	ion charge
k_B	Boltzmann's constant
T	temperature

- Generalized Born (GB) Theory
  - Born formula (Born 1920)
    generalized Born formula (Still 1990)
  - methods to compute the Born radii:
    - pairwise summation : fast but not easy for force calculation
    - ASMS, nFFT, and higher order quadrature [Bajaj, Zhao 2006]

$$G_{\text{pol}} = -\frac{\tau}{2}\sum_{ij}\frac{q_i q_j}{[r_{ij}^2 + R_i R_j \exp(-\frac{r_{ij}^2}{4R_i R_j})]^{\frac{1}{2}}}$$

$$R_i^{-1} = \frac{1}{4\pi}\int_{\Gamma}\frac{(\mathbf{r}-\mathbf{x}_i)\cdot\mathbf{n}(\mathbf{r})}{|\mathbf{r}-\mathbf{x}_i|^4}\, dS$$

Generalized Born Energy



Generalized Born method

$$G_{\mathrm{pol}} = -\frac{\tau}{2}\sum_{i,j}\frac{q_i q_j}{[r_{ij}^2 + R_i R_j \exp(-\frac{r_{ij}^2}{4R_iR_j})]^{\frac{1}{2}}} \quad (1)$$

where q is the atomic charge, r is the distance between atom pairs, and R is the atom's **effective Born radius** which reflects how deeply the atom is buried in a molecule:

$$R_i^{-1} = \frac{1}{4\pi}\int_{\Gamma}\frac{(\mathbf{r}-\mathbf{x}_i)\cdot\mathbf{n}(\mathbf{r})}{|\mathbf{r}-\mathbf{x}_i|^4}\,dS \quad (2)$$

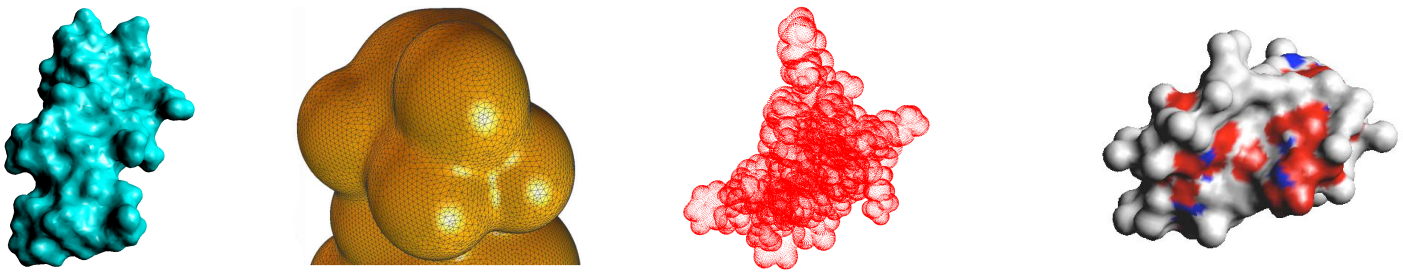
where Γ is the molecular surface of the solute, $\mathbf{x}$ is the atomic center, and $\mathbf{n}$ is the unit surface normal.

Fast Computation of Born Radii R_i

$$R_i^{-1} = \frac{1}{4\pi}\int_\Gamma \frac{(\mathbf{r}-\mathbf{x}_i)\times\mathbf{n}(\mathbf{r})}{|\mathbf{r}-\mathbf{x}_i|^4}\,dS \approx \frac{1}{4\pi}\sum_{k=1}^{N} w_k \frac{(\mathbf{r}_k-\mathbf{x}_i)\times\mathbf{n}(\mathbf{r}_k)}{|\mathbf{r}_k-\mathbf{x}_i|^4}, \qquad \mathbf{r}_k \in \Gamma$$

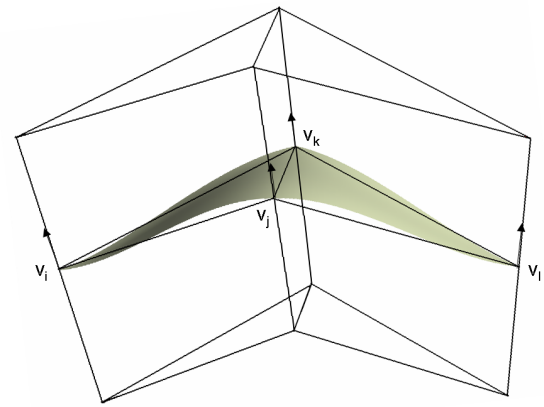
Algorithm:

1. Generate a smooth A-spline model for the molecular surface Γ.
2. Cubature: choose w_k and $\mathbf{r}_k$ properly so that higher order of accuracy can be obtained for small N.
3. Fast summation using non-uniform FFT to evaluate R_i , $i = 1,\ldots,M$.



•[Zhao, Xu, Bajaj, ACM SPM 2006]

Algebraic Splines (A-Splines)



$S := \{\mathbf{p}(b_1, b_2, b_3, \lambda) : F(b_1, b_2, b_3, \lambda) = 0, \mathbf{p} \in D_{ijk}\}$ where

- $D_{ijk} = \{\mathbf{p}(b_1, b_2, b_3, \lambda) : \mathbf{p} = b_1 \mathbf{v}_i(\lambda) + b_2 \mathbf{v}_j(\lambda) + b_3 \mathbf{v}_k(\lambda), \lambda \in I_{ijk}\}$
- I_{ijk} is an interval containing 0
- $F(b_1, b_2, b_3, \lambda) = \sum_{i+j+k=n} b_{ijk}(\lambda) B^n_{ijk}(b_1, b_2, b_3)$
- $B^n_{ijk}(b_1, b_2, b_3) = \frac{n!}{i!\, j!\, k!} b_1^i b_2^j b_3^k$
- n>2 so that S is smooth. We consider the case n=3. $b_{ijk}(\lambda)$ are determined so that C^1 continuity is obtained across the patch boundaries.

•[Bajaj, Xu CAGD 2002]

Fast Summation via Error bounded Non-uniform Fast Fourier Transforms

To evaluate forms like $I(x_i) = \sum_{k=1}^{N} c_k K\ (x_i - r_k), \quad i = 1, \cdots, M, \quad K \in C^p(-\frac{1}{2}, \frac{1}{2})$

$$K(x_i - r_k) \approx \sum_{\omega=-n/2}^{n/2} b_\omega e^{2\pi i\omega(x_i - r_k)} := K_F$$

$$I(x_i) \approx \sum_{k=1}^{N} c_k K_F(x_i - r_k) = \sum_{k=1}^{N} c_k \sum_{\omega=-n/2}^{n/2-1} b_\omega e^{2\pi i\omega(x_i - r_k)} = \sum_{\omega=-n/2}^{n/2-1} b_\omega a_\omega e^{2\pi i\omega x_i}$$ NFFT

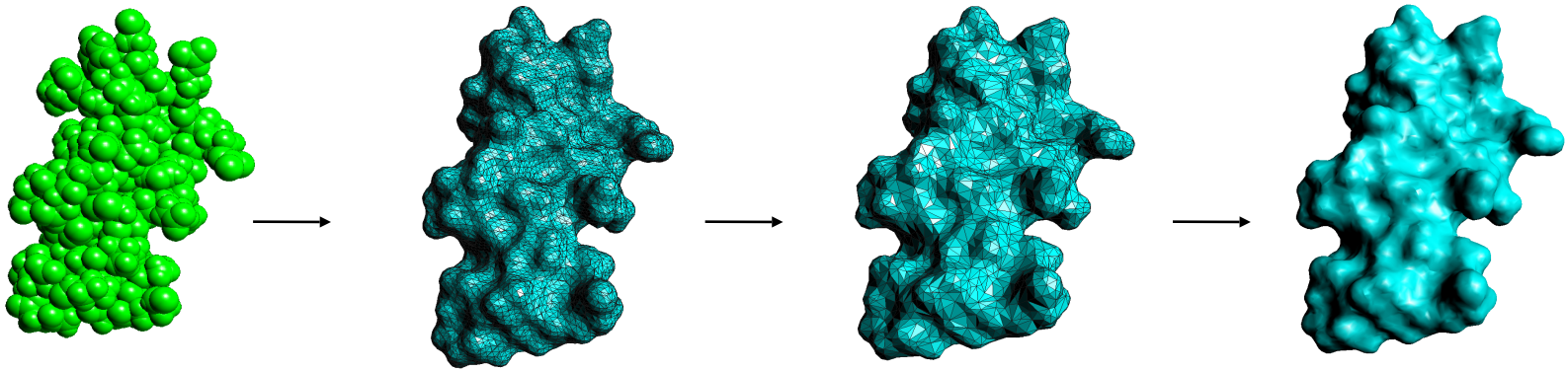
where $a_\omega = \sum_{k=1}^{N} c_k e^{-2\pi i\omega r_k}$ NFFT^T

Time complexity:

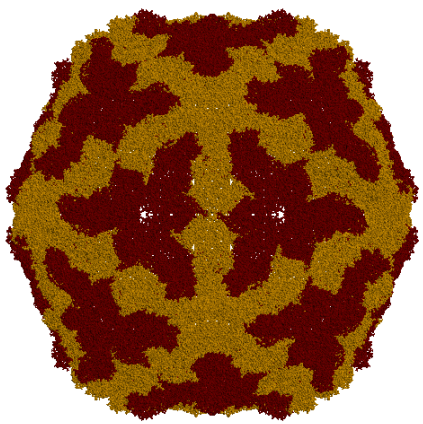
$a_\omega : O(mn + N\log N)$ $\quad b_\omega : O(n\log n)$ $\quad I : O(mM + n\log n)$ $\quad m < n/2$

overall: $O(mM + mn + N\log N + n\log n)$ (trivial method: $O(MN)$)

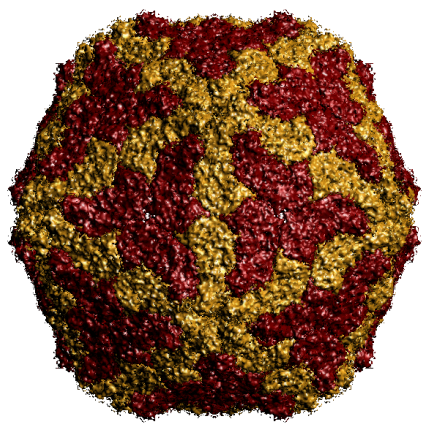
ASMS: Algebraic Spline Molecular Surfaces



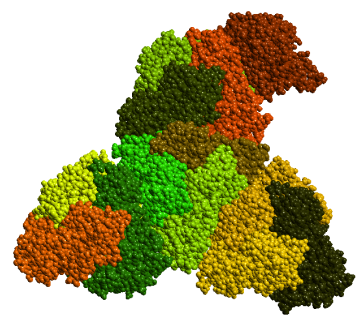
atomic/quasi-atomic structure → molecular surface triangulation → coarse mesh → A-Spline



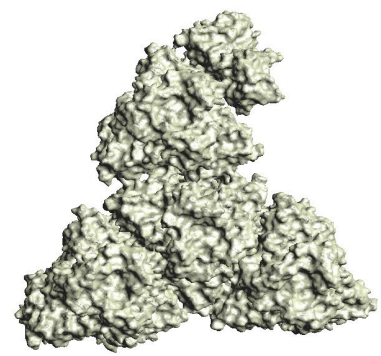
atomic structure



A-Spline



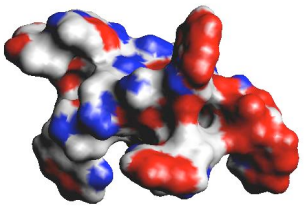
atomic structure asymmetric subunit P8



A-Spline

Force Calculation

The electrostatic force acting on atom α which is part of the forces driving Molecular dynamics is



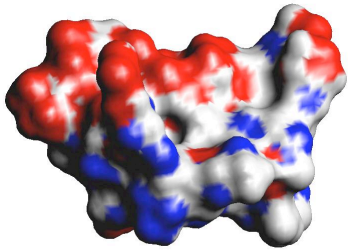
$$\mathbf{F}_\alpha^{\text{elec}} = -\frac{\partial G_{\text{pol}}}{\partial \mathbf{x}_\alpha}$$

$$\frac{\partial G_{\text{pol}}}{\partial \mathbf{x}_\alpha} = \sum_i \sum_{j \neq i} \frac{\partial G_{\text{pol}}}{\partial r_{ij}} \frac{\partial r_{ij}}{\partial \mathbf{x}_\alpha} + \sum_i \frac{\partial G_{\text{pol}}}{\partial R_i} \frac{\partial R_i}{\partial \mathbf{x}_\alpha}$$

$$R_i^{-1} = \frac{1}{4\pi} \int_{\text{ex}} \frac{1}{|\mathbf{r} - \mathbf{x}_i|^4} \, dV$$, or $$R_i^{-1} = \frac{1}{4\pi} \int_\Gamma \frac{(\mathbf{r} - \mathbf{x}_i) \cdot \mathbf{n}(\mathbf{r})}{|\mathbf{r} - \mathbf{x}_i|^4} \, dS$$

The integration domain depends on $\mathbf{x}_\alpha$.

Force Calculation



$$-\frac{1}{R_i^2}\frac{\partial R_i}{\partial \mathbf{x}_\alpha} = \frac{1}{4\pi}\int_{\mathbb{R}^3} \frac{\partial}{\partial \mathbf{x}_\alpha}\frac{\bar{\varrho}(\mathbf{r},\{\mathbf{x}_j\})}{|\mathbf{r}-\mathbf{x}_i|^4}\, dV$$

$$\frac{\partial R_i}{\partial \mathbf{x}_\alpha} = -\frac{R_i^2}{4\pi}\left(\boxed{\int_{\mathbb{R}^3} \frac{\frac{\partial}{\partial \mathbf{x}_\alpha}\bar{\varrho}(\mathbf{r},\{\mathbf{x}_j\})}{|\mathbf{r}-\mathbf{x}_i|^4}\, dV} + \int_{\mathrm{ex}} \frac{\partial}{\partial \mathbf{x}_\alpha}\frac{1}{|\mathbf{r}-\mathbf{x}_i|^4}\, dV\right)$$

$$\|$$

$$\boxed{-4\mathbf{x}_i 3\int_{\Gamma} \frac{(\mathbf{r}-\mathbf{x}_i)\cdot\mathbf{n}(\mathbf{r})}{|\mathbf{r}-\mathbf{x}_i|^6}\, dS}$$

$$\frac{\partial}{\partial \mathbf{x}_\alpha}\bar{\varrho} = -\frac{\partial}{\partial \mathbf{x}_\alpha}\tilde{\varrho} = \frac{\partial \varrho_\alpha}{\partial \mathbf{x}_\alpha}\left(1-\sum_j \varrho_j + \sum_{j<k}\varrho_j\varrho_k - \sum_{j<k<l}\varrho_j\varrho_k\varrho_l\right)$$

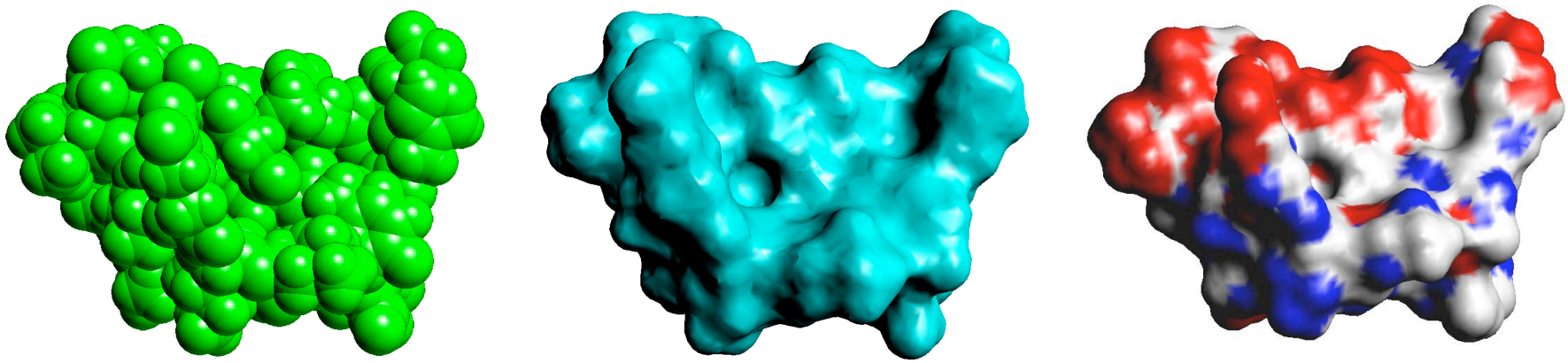
Molecular surface

Since $\frac{\partial \rho_\alpha}{\partial \mathbf{x}_\alpha} \neq 0$ only if $a_\alpha - w < |\mathbf{r}-\mathbf{x}_\alpha| < a_\alpha + w$

Molecular skin

$$\int_{\mathbb{R}^3} \frac{\frac{\partial}{\partial \mathbf{x}_\alpha}\bar{\rho}(\mathbf{r},\{\mathbf{x}_j\})}{|\mathbf{r}-\mathbf{x}_i|^4}\, dV = \boxed{\int_{|\mathbf{r}-\mathbf{x}_\alpha|=a_\alpha-w}^{|\mathbf{r}-\mathbf{x}_\alpha|=a_\alpha+w} \frac{\partial \rho_\alpha}{\partial \mathbf{x}_\alpha}\frac{1-\sum_j \rho_j + \sum_{j<k}\rho_j\rho_k - \sum_{j<k<l}\rho_j\rho_k\rho_l}{|\mathbf{r}-\mathbf{x}_i|^4}}$$

Molecular Solvation Forces



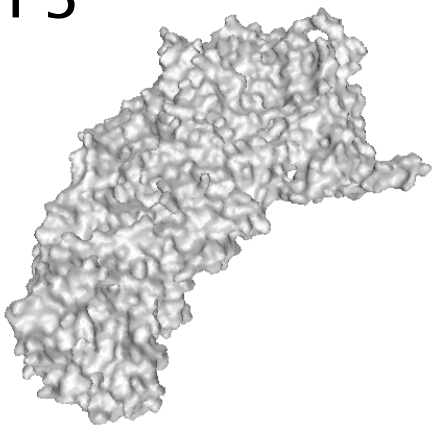
(a) (b) (c)

Figure: (a) The atomic model of protein 1PPE (436 atoms); (b) The piecewise algebraic surface of 1PPE. $G_{pol} = -825.33$ kcal/mol.

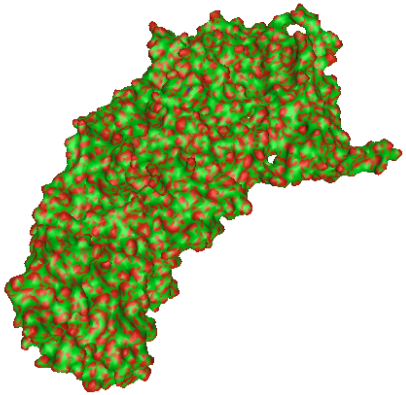
; (c) The solvation forces of protein 1PPE shown as function on surface

P3 and P8

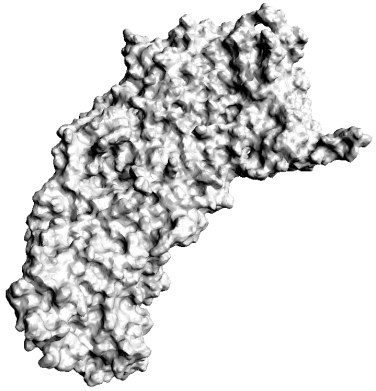
P3



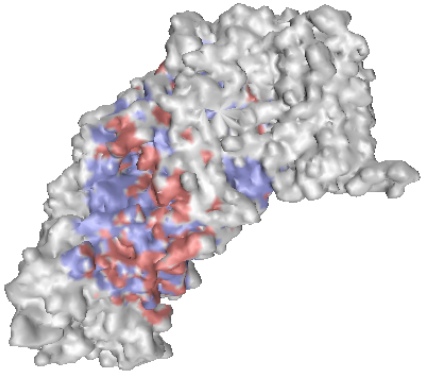
Molecular Surface



Gaussian Curvature

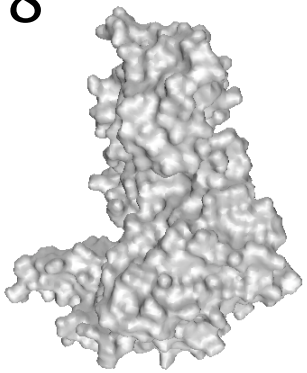


GB Energy:
-23892.33 kcal/Mol



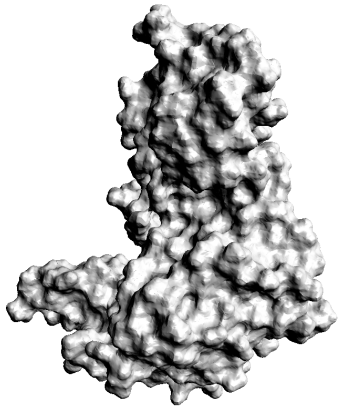
PB energy:
-17758.9 kJ/Mol

P8

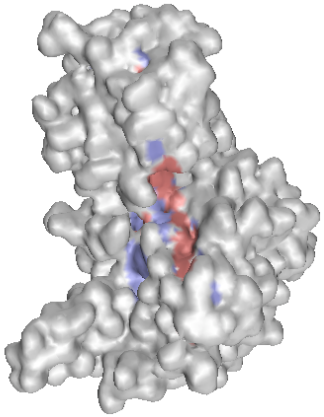


Molecular Surface

Gaussian Curvature



GB energy:
-8413.185 kcal/Mol



PB energy:
-7163.08 kJ/Mol

Center for Computational Visualization
Institute of Computational and Engineering Sciences
Department of Computer Sciences University of Texas at Austin

Acknowledgements

- **Group Members**
  - **Albert Chen (CS, Ph.D)**
  - **Rezaul Chowdhury (PostDoc)**
  - **Ajay Gopinath (BME,Ph.D)**
  - **Andrew Gillete (Math, Ph.D.)**
  - **R. Inkulu (PostDoc)**
  - **Alex Kim (CS, Ugrad)**
  - **Bongjune Kwon (BME,Ph.D)**
  - **Oliver Martin (CS, Ugrad)**
  - **Maysam Mousallem (CS,PhD)**
  - **Joe Rivera (Res. Scientist)**
  - **Aditi Saha (CS, MS)**
  - **Jesse Sweet (Math, MS)**
  - **Dongjin Suh (CS, MS)**
  - **Samrat Goswami (PostDoc)**
  - **Zeyun Yu (UCSD)****
  - **Xiaoyu Zhang (CSU)****
  - **Wenqi Zhao (CAM, Ph.D.)**

- **Senior Collaborators**
  - Manfred Auer (LBL,UCB)
  - Tim Baker, M. Ellisman (UCSD)
  - Nathan Baker (WashU)
  - Wah Chiu, Steve Ludtke (Baylor)
  - Inderjit Dhillon (UT, CS)
  - Ron Elber, Peter Rossky (UT, ICES)
  - Joachim Frank (Suny Albany, Columbia)
  - Kristen Harris , Dan Johnston (UT, ILM)
  - Tom Hughes, Tinsley Oden (UT,ICES)
  - Justin Kiney, Tom Bartol, Terry Sejnowski (Salk)
  - Andy McCammon, Michael Holst (UCSD)
  - Art Olson, Michel Sanner, David Goodsell (TSRI)
  - Alberto Paoluzzi, Antonio DiCarlo(Roma)
  - Sriram Subramaniam (NIH-NCI)

- **Sponsored by**
  - **NIH: P20-RR020647, R01-GM074258, R01-0783087, R01-EB004873**

    NSF: ITR-IIS-032550, DDDAS-CNS-0540033

    UT: HTDS