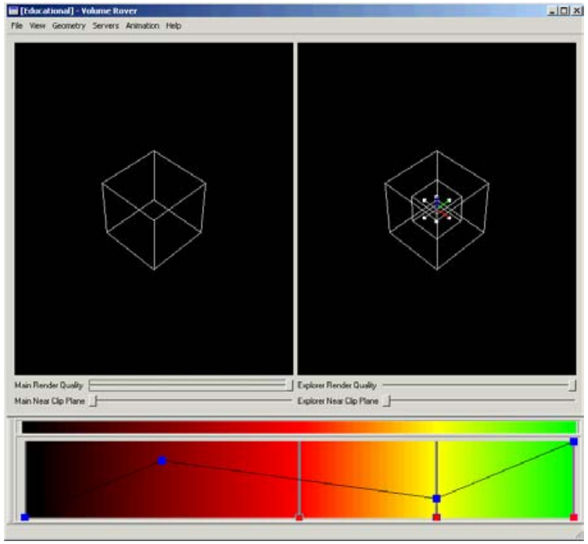
Geometric and Signal 3D Processing (and some Visualization)

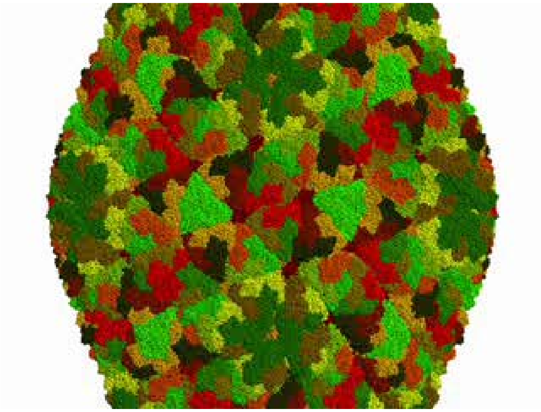
Chandrajit Bajaj

Algorithms & Tools

- Structure elucidation: filtering, contrast enhancement, segmentation, skeletonization, subunit identification
- Structure Modeling: finite element meshing, spline representations(A-spline,RBF representations) for structural fitting & complementary docking
- Visualization: multi-dimensional transfer functions, surface and volume texture rendering, wavelet compression, hierarchical representations, cluster based parallelism



VolRover



TexMol

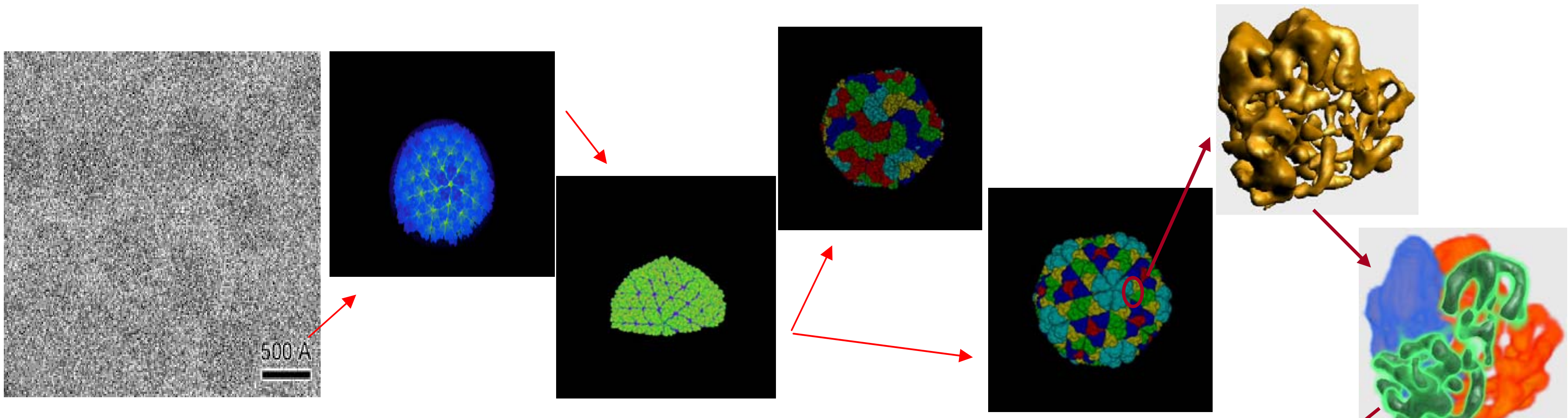
The CVC Team and Collaborators

- *Personnel*
  - **Inderjit Dhillon (Assoc. Director)**
  - **Albert Chen (CS, Ph.D)**
  - **Katherine Clarridge (MBE,M.S.)**
  - **KL Chandrasekhar (ME, Ph.D.)**
  - **Tamal Dey (OSU) ****
  - **Samrat Goswami (PostDoc, CS)**
  - **Rick Hankins (Res. Scientist)**
  - **Insung Ihm (SNU, S. Korea)****
  - **Sangmin Park (CS, Ph.D.)**
  - **Bong-June Kwon (CS, M.S.)**
  - **Bong-Soo Sohn (Stanford U)****
  - **Jason Sun (Res. Scientist)**
  - **John Wiggins (Res. Scientist)**
  - **Vinay Siddahanavalli (CS, Ph.D.)**
  - **Guoliang Xu (AS,China)****
  - **Zeyun Yu (CS, Ph.D)**
  - **Xiaoyu Zhang (CSU)****
  - **Jessica Zhang (PostDoc, ICES)**
  - **Wenqi Zhao (ICES, Ph.D.)**

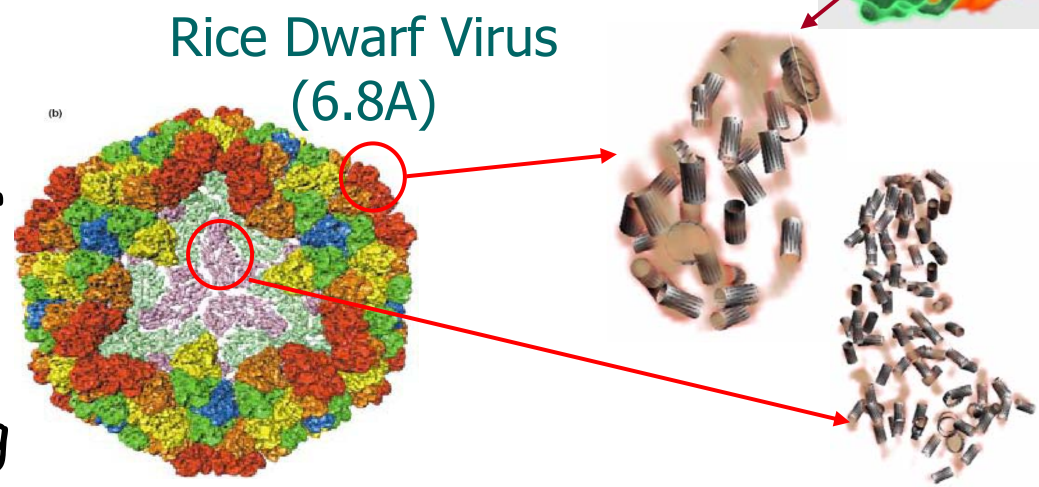
- *Senior Collaborators*
  - Manfred Auer (LBL)
  - Nathan Baker (Wash. U.)
  - Helen Baker, Cathy Lawson (Rutgers U)
  - Tim Baker (UCSD)
  - Tom Bartol (SALK)
  - Luis Caffarelli (ICES)
  - Wah Chiu, Matt. Baker (Baylor)
  - Leszek Demkowicz (ICES)
  - Gregory Gladish, J. Hazle (MD Anderson)
  - Tom Hughes (ICES)
  - Andy McCammon (USCD)
  - Tinsley Oden (ICES)
  - Art Olson, M. Sanner, D. Goodsell, Charlie Brooks, V. Reddy (Scripps)
  - Peter Rossky (ICES)
  - Andre Sali (UCSF)

- ***Funding***
  - **NIH: P20(planning), R01**
  - **NSF: ITR , DDDAS**
  - **Whitaker Foundation**

Sub-nanometer Structure Elucidation from 3D Cryo-EM



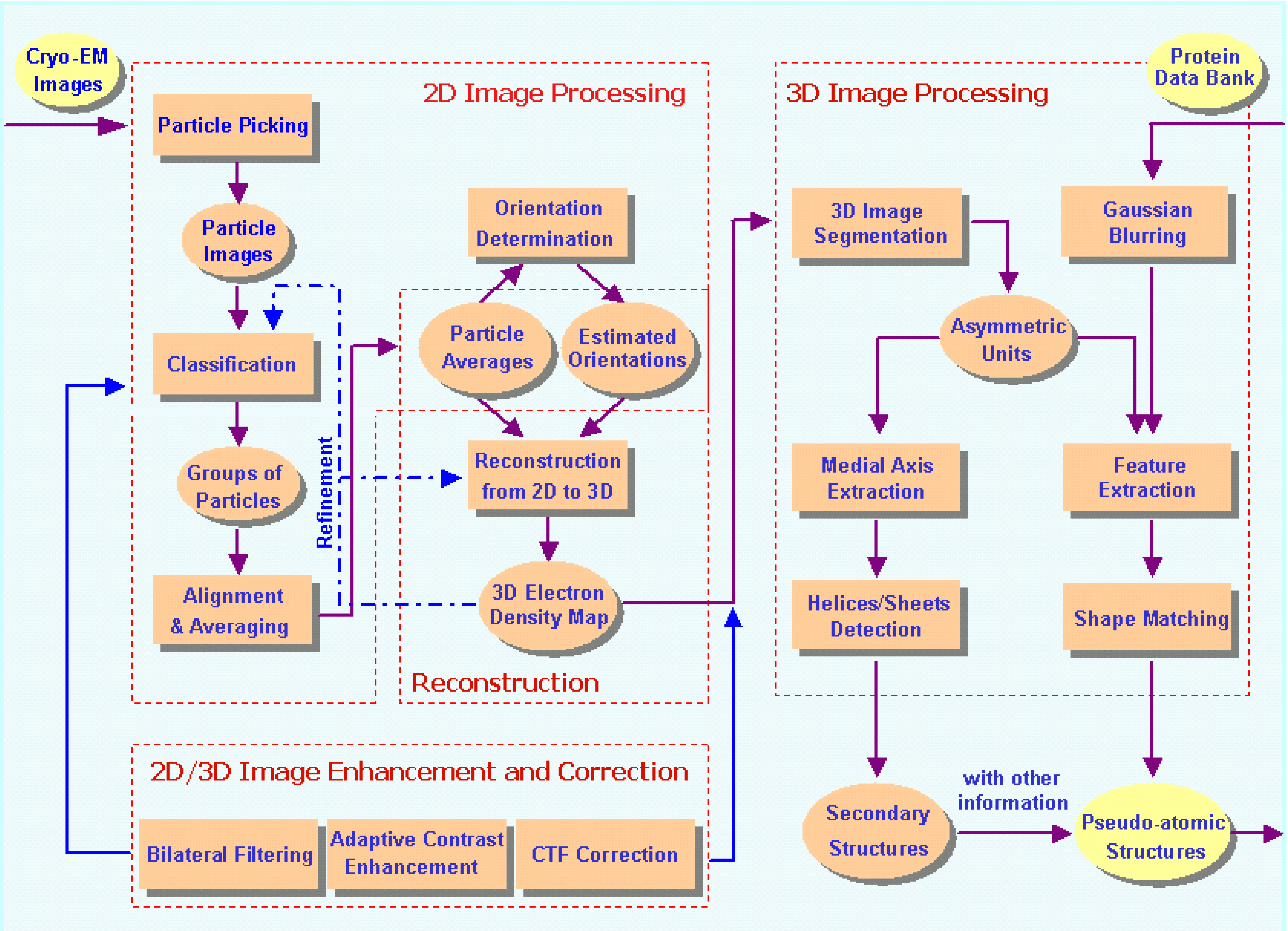
Cryo-EM →FFT based 3D Reconstruction →Anisotropic and Vector Diffusion Filtering → Structure Segmentation →Quasi-Atomic Modeling →Visualization



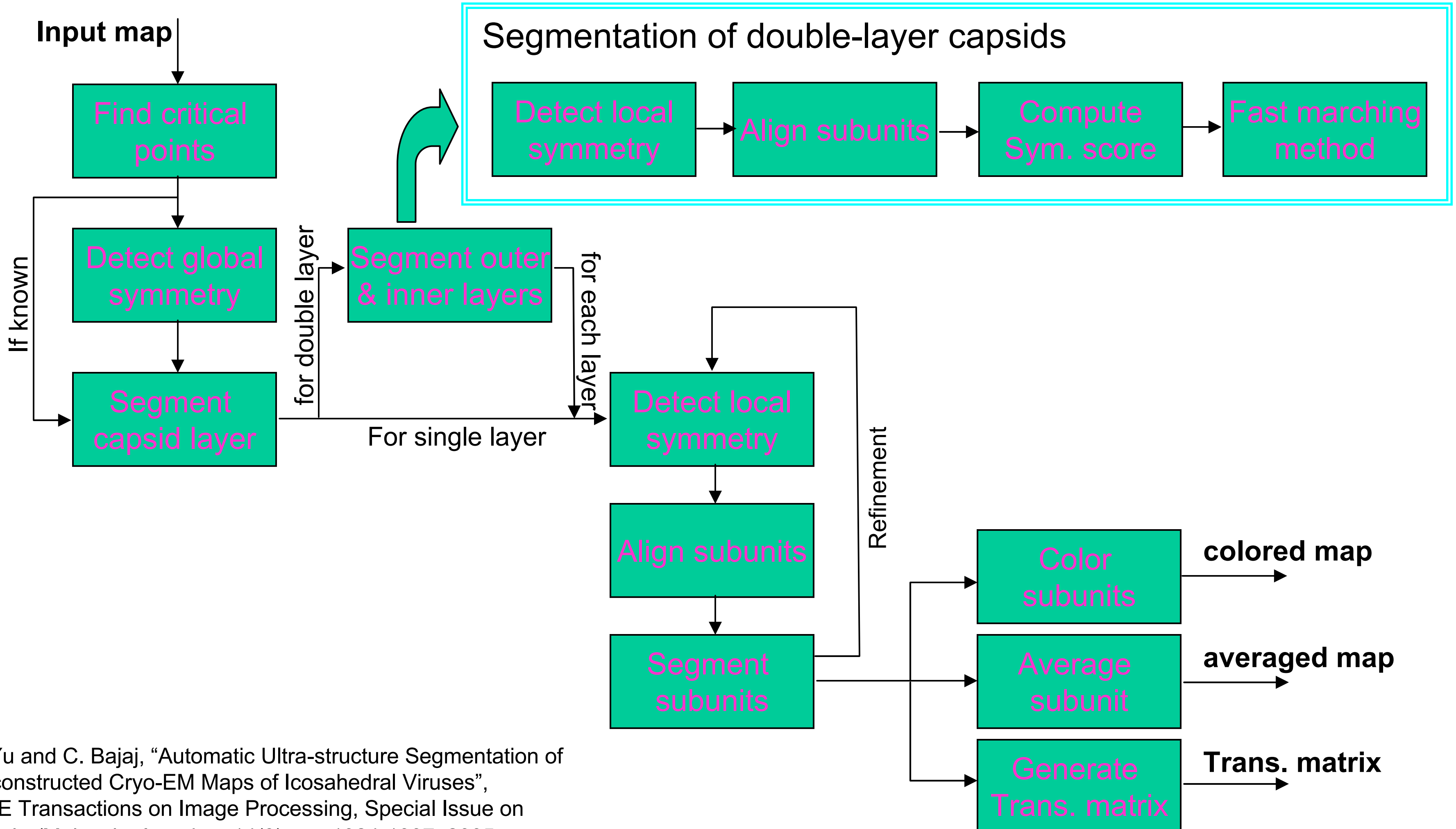
**Sponsored by NSF-ITR, NIH

(Collaborators: Wah Chiu,NCMI, Baylor College of Medicine, Andrej Sali, UCSF)

A Structure Determination Pipeline for single particle cryo-EM



Structure Elucidation for Icosahedral Viruses



Z. Yu and C. Bajaj, "Automatic Ultra-structure Segmentation of Reconstructed Cryo-EM Maps of Icosahedral Viruses", IEEE Transactions on Image Processing, Special Issue on Cellular/Molecular Imaging, 14(9), pp. 1324-1337, 2005.

Structure Elucidation 1(A)

- Adaptive contrast enhancement
- Bilateral filtering

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

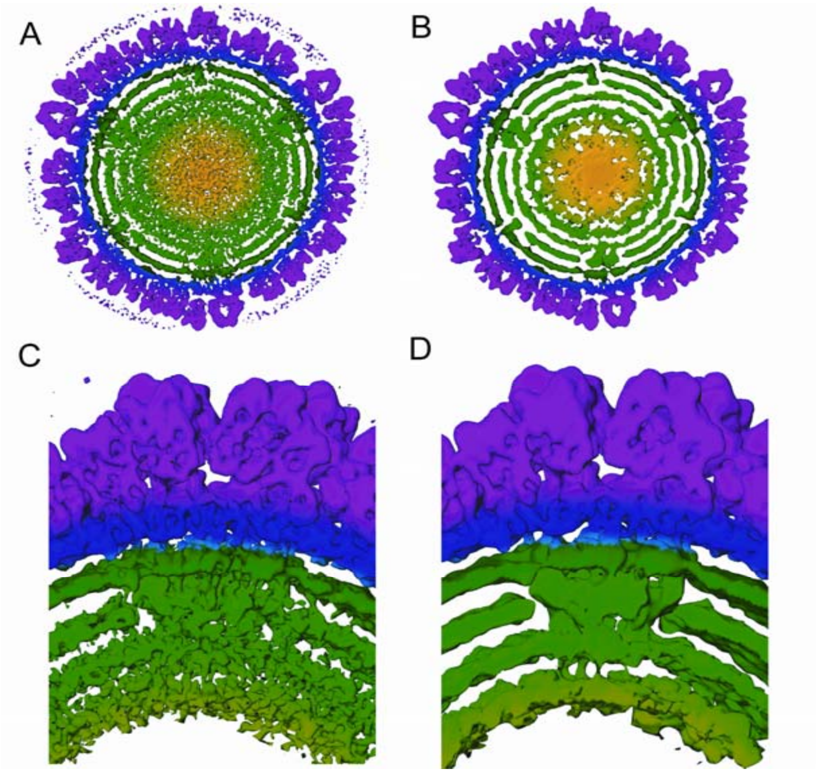
where σ_d and σ_r are parameters and *f*(.) is the image intensity value.

- 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.

- Anisotropic gradient vector diffusion

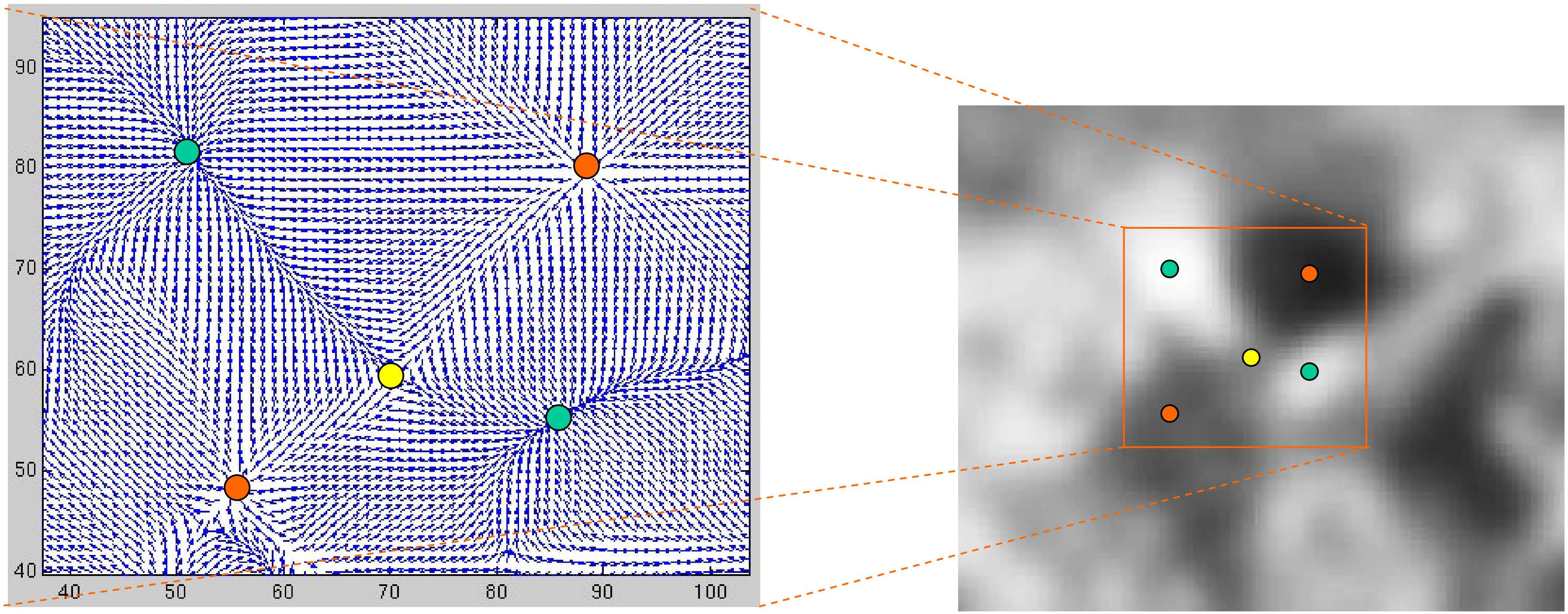


W. Jiang, M. Baker, Q. Wu, C. Bajaj, W. Chiu, Journal of Structural Biology, 144, 5,(2003),114-122

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

Z. Yu & C. Bajaj, Proc. Int'l Conf. Image Processing, 2002. pp. 1001-1004.
Z. Yu & C. Bajaj, Proc. Int'l Conf. Computer Vision and Pattern Recognition, 2004. 415-420.

Compute Critical Points Using AGVD



●: minimum ●: maximum ●: saddle

(0) (3) (1, 2)

Anisotropic Gradient Vector Diffusion (AGVD)

Isotropic Diffusion (Xu *et al*., 1998)

$$\begin{cases} \frac{\partial u}{\partial t} = \mu \nabla^2 u - (u - f_x)(f_x^2 + f_y^2) \\ \frac{\partial v}{\partial t} = \mu \nabla^2 v - (v - f_y)(f_x^2 + f_y^2) \end{cases}$$

Where:

(u(t), v(t)) stands for the evolving vector field;

μ is a constant;

f is the original image to be diffused;

(f_x, f_y) = (u(0), v(0)).

Anisotropic Diffusion (Yu & Bajaj ICPR'02)

$$\begin{cases} \frac{\partial u}{\partial t} = \mu \nabla (g(\alpha) \cdot \nabla u) - (u - f_x)(f_x^2 + f_y^2) \\ \frac{\partial v}{\partial t} = \mu \nabla (g(\alpha) \cdot \nabla v) - (v - f_y)(f_x^2 + f_y^2) \end{cases}$$

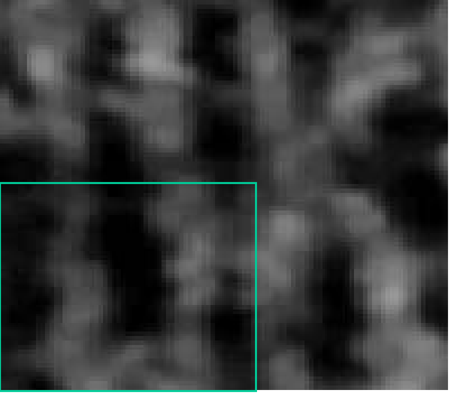
Where

(u(t), v(t)) stands for vector field;

μ is a constant; (f_x, f_y) = (u(0), v(0)).

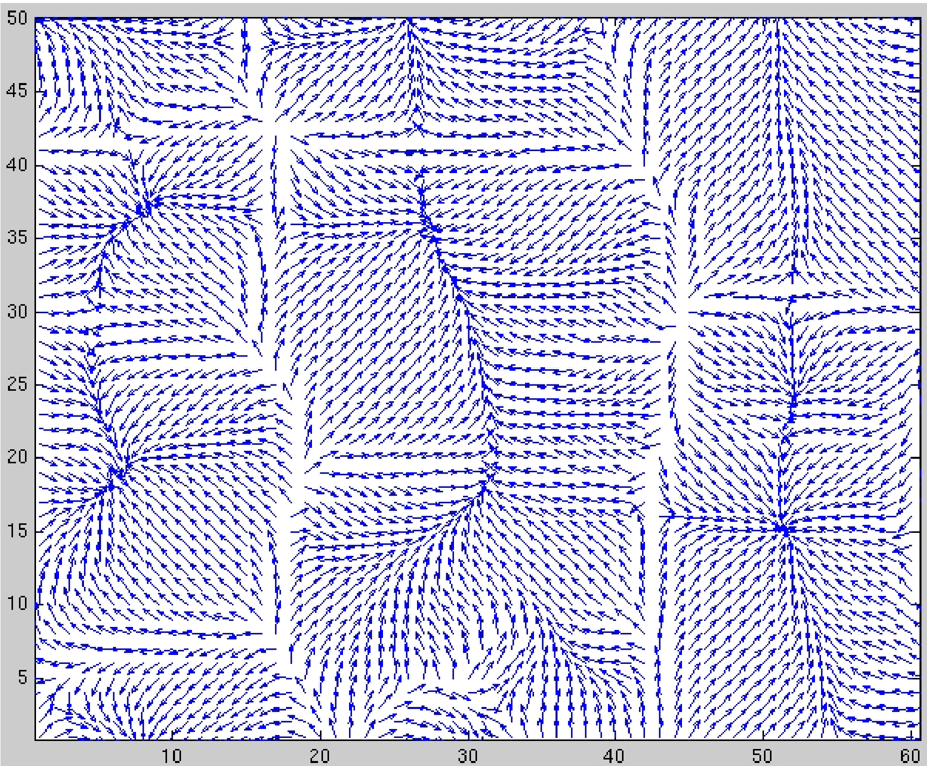
f is the original image to be diffused;

g(.) is the angle between two vectors

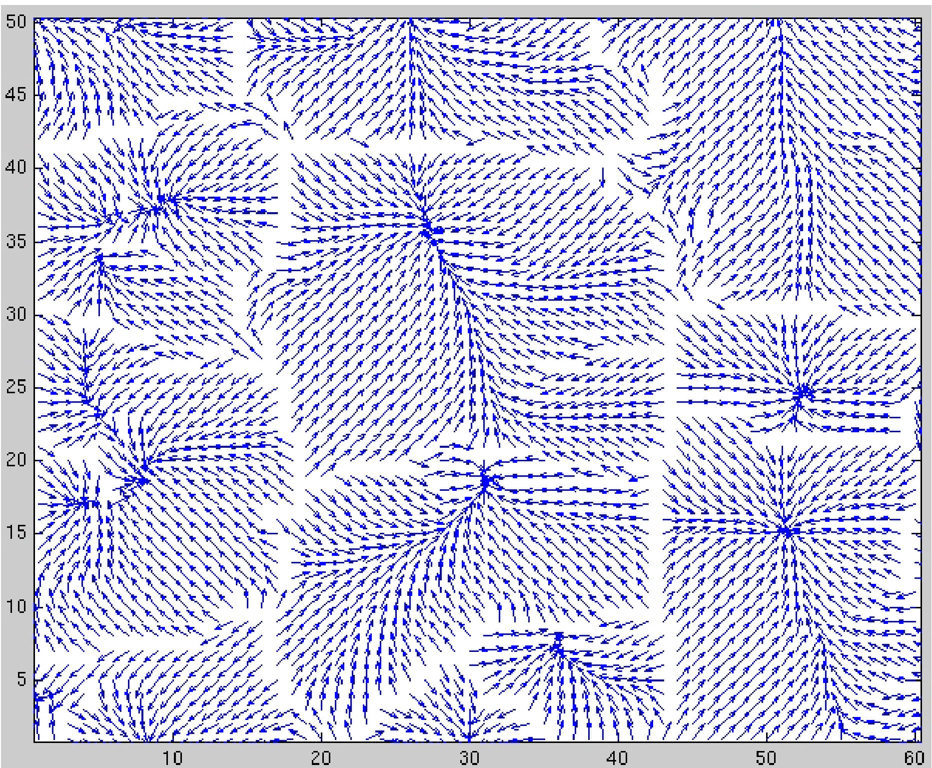


GVD v.s. AGVD

Isotropic diffusion

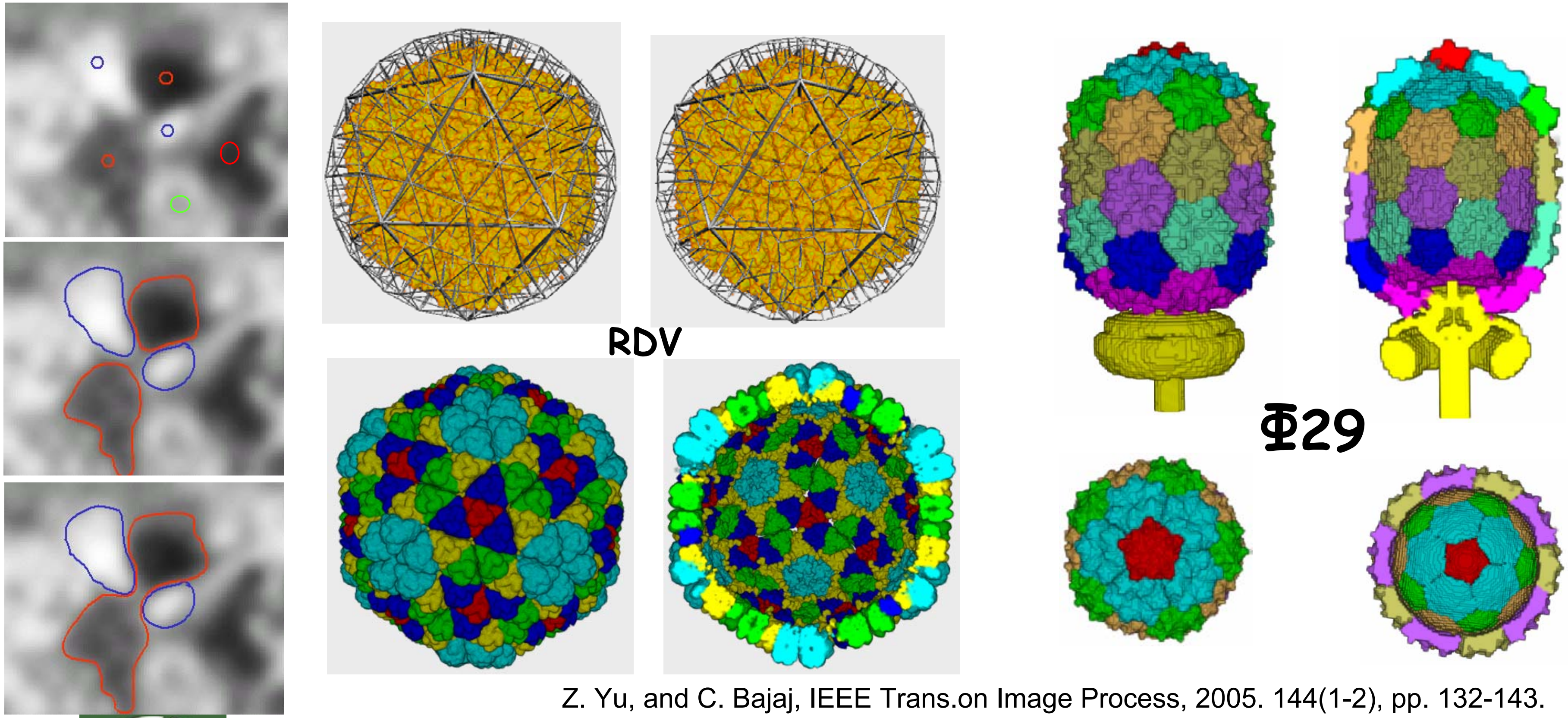


Anisotropic diffusion



Structure Elucidation 1(B)

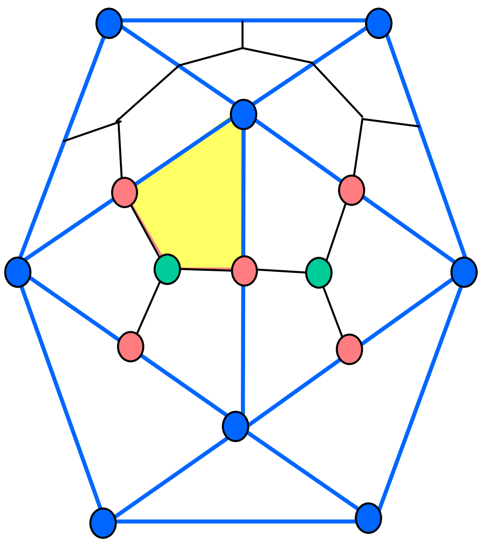
- Multi-seed Fast Marching Method
  - Classify map critical points as seeds based on local symmetry.
  - Each seed initializes a contour, with its group's membership.
  - Contours march simultaneously. Contours with same membership are merged, while contours with different membership stop each other.

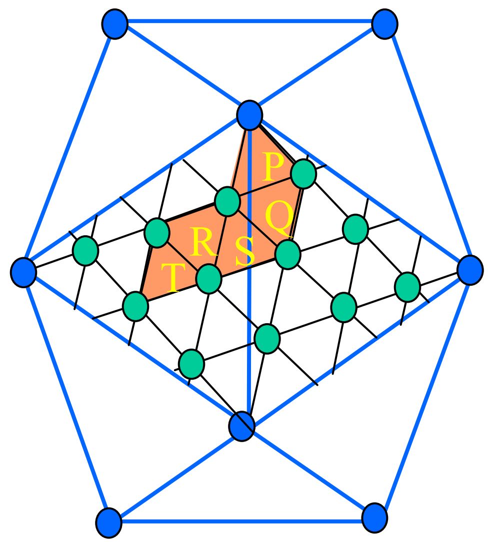


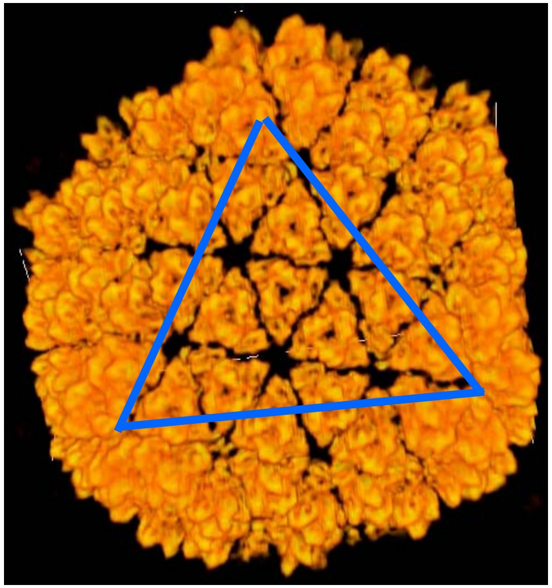
Z. Yu, and C. Bajaj, IEEE Trans.on Image Process, 2005. 144(1-2), pp. 132-143.

Global and Local Symmetry

- Automatic structure unit identification in a 3D Map







Two-fold vertices

Three-fold vertices

Five-fold vertices

Example: RDV

Symmetry Detection: Correlation Search

$$C(\theta,\varphi)=\sum_{\vec{r}\in V} f(\vec{r})\,f(R_{(\theta,\varphi,2\pi/5)}\cdot\vec{r})$$

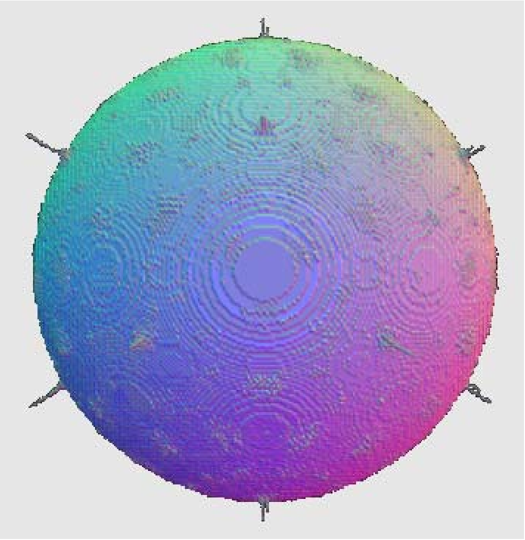
- Algorithm: *detect 5-fold rotation symmetry*
  - Compute the scoring function
    - For every angular bin B_j, compute θ_j,φ_j {

      For every critical point C_i {

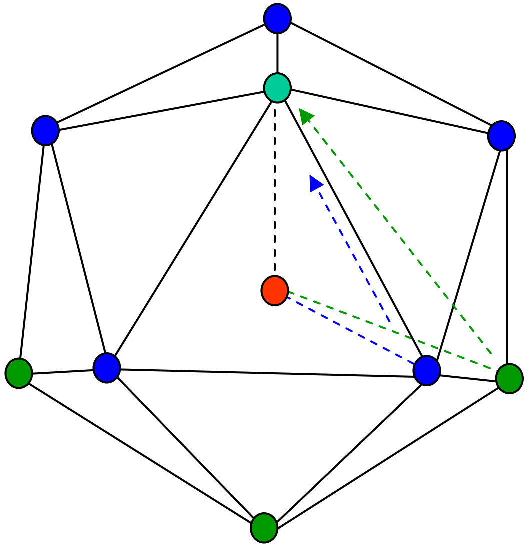
      $$\vec{r}_k(C_i,B_j)=R_{(\theta_j,\varphi_j,2k\pi/5)}\cdot C_i,\quad k=0,1,2,3,4$$

      $$Dev(C_i,B_j)=\frac{1}{5}\sum_{k=0}^{4}(f(\vec{r}_k)-\bar{f})\ \}$$

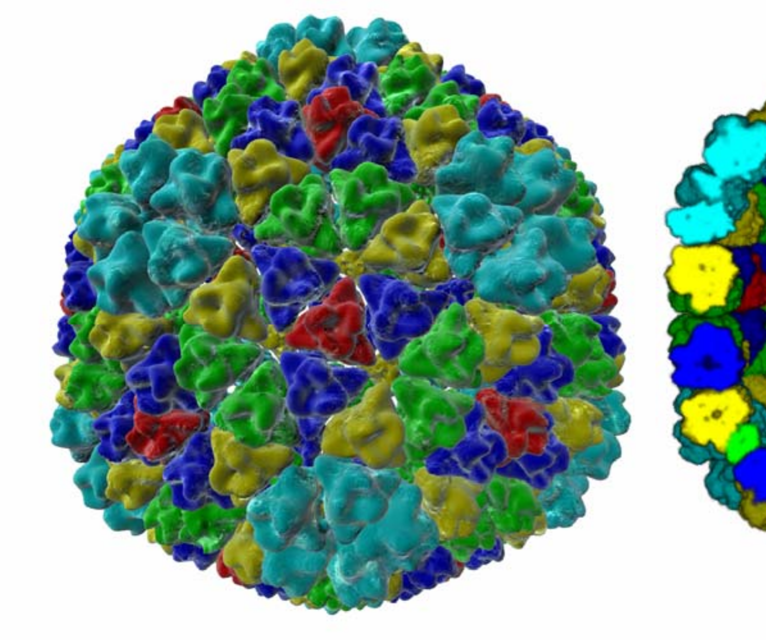
      $$SF(B_j)=\frac{1}{p}\sum_{i=0}^{p}Dev(C_i,B_j)\ \}$$
  - Locate the symmetry axes
    - The 12 peaks
  - Refine the symmetry axes
    - In order to locate a perfect icosahedron (rotate the axes by 0^0, 63.43^0, 116.57^0, 180^0)



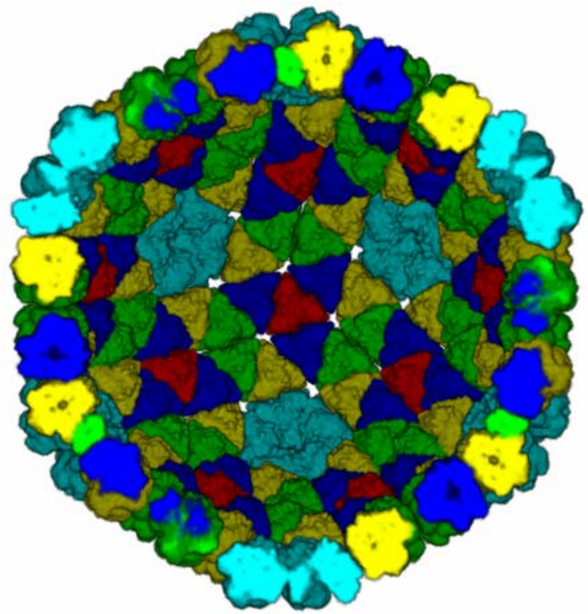
Inverted and normalized SF(*Bj*)



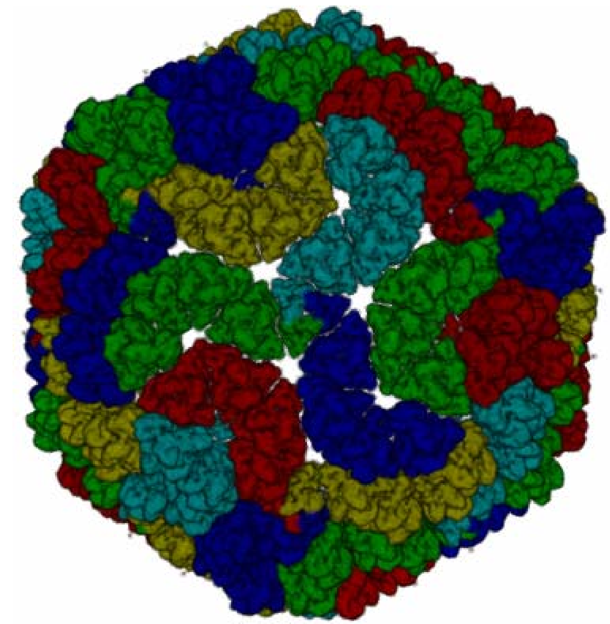
Structure Elucidation Results: RDV (Bakeoff)



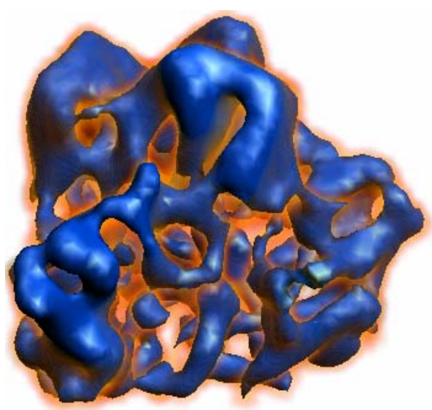
surface rendering (outside)



volume rendering (inside)

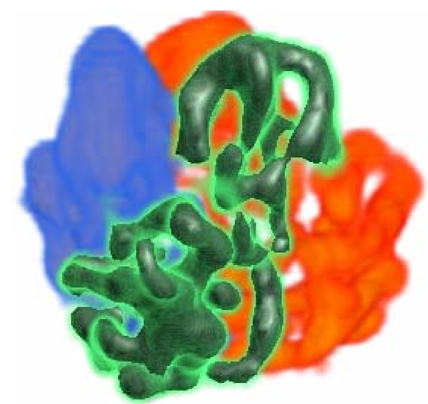


volume rendering (asymmetric unit)



averaged trimer (side)

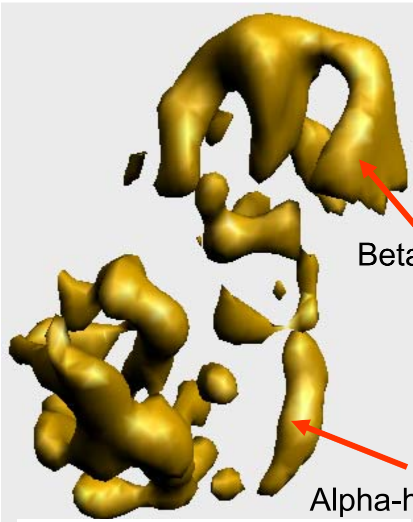
averaged trimer (bottom)

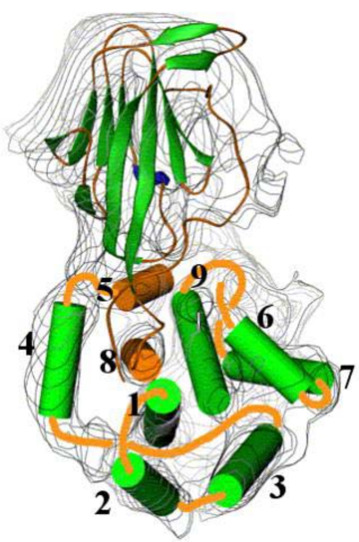


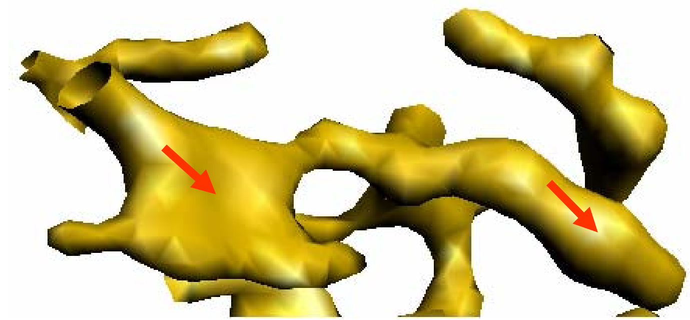
segmented monomers

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

Structure Elucidation 1(C): Secondary Structure Identification

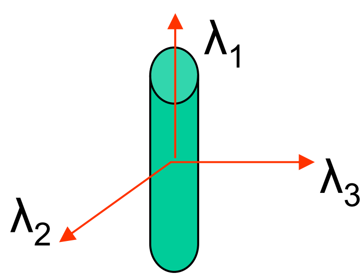


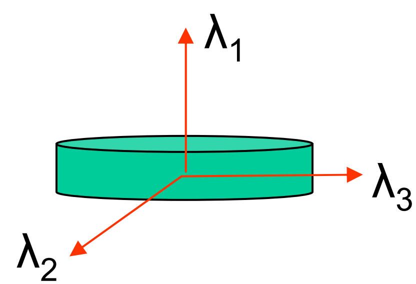




$$G_\sigma * \begin{pmatrix} I_x^2 & I_x I_y & I_x I_z \\ I_x I_y & I_y^2 & I_y I_z \\ I_x I_z & I_y I_z & I_z^2 \end{pmatrix}$$

The eigenvectors of the local structure tensor give the principal directions of the local features:





Line structure (alpha-helix)

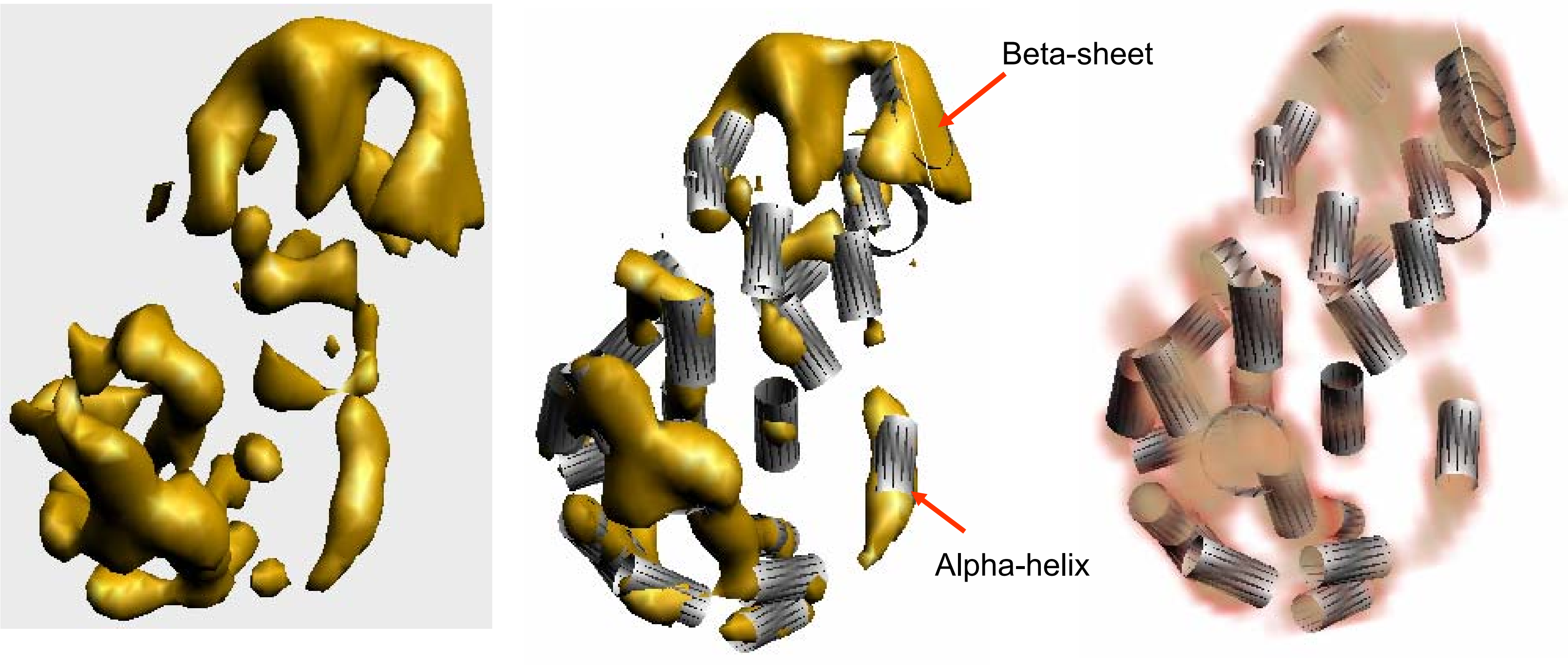
$$\lambda_2 \approx \lambda_3 >> \lambda_1 \approx 0$$

plane structure (beta-sheet)

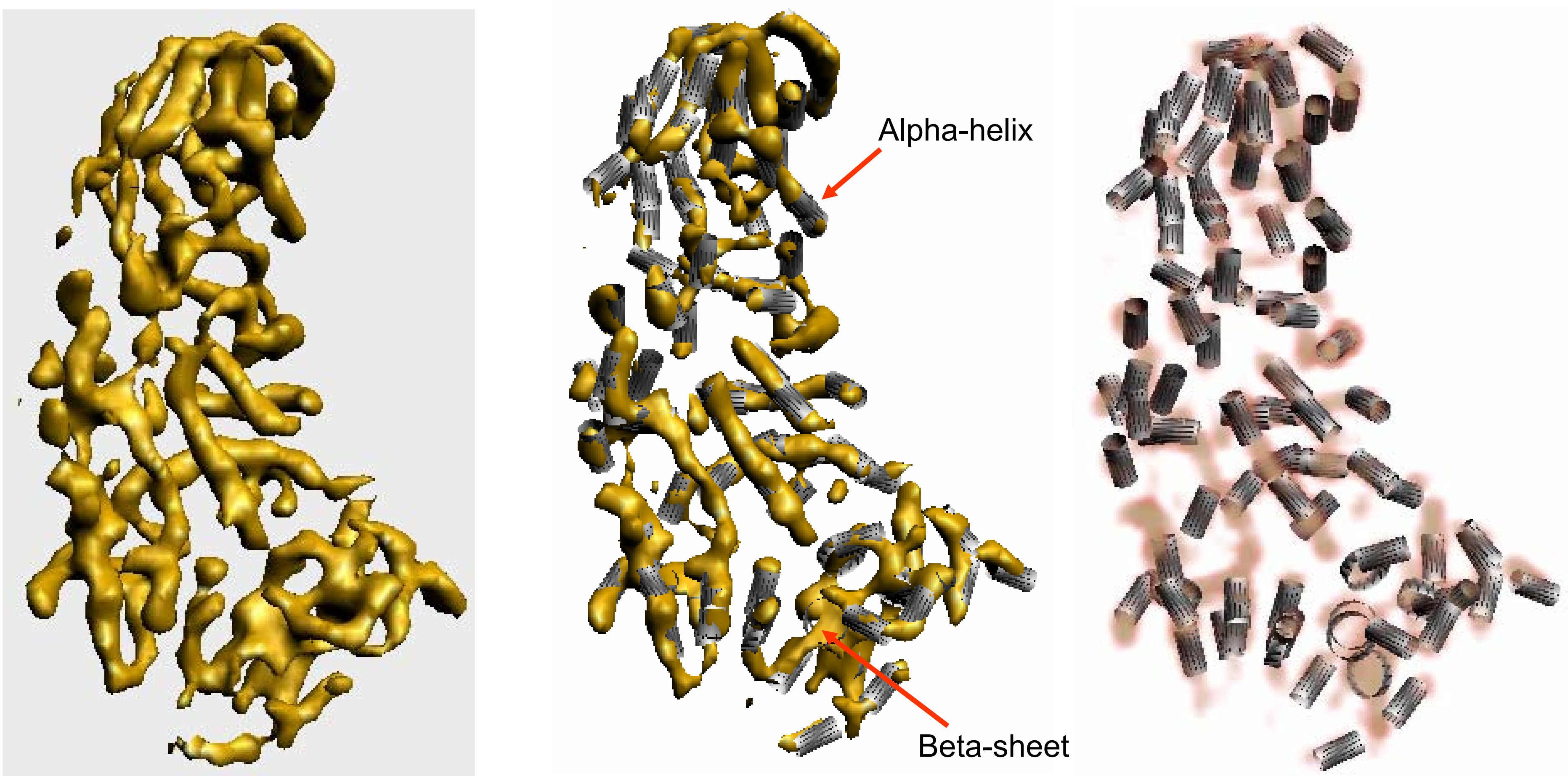
$$\lambda_1 >> \lambda_2 \approx \lambda_3 \approx 0$$

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

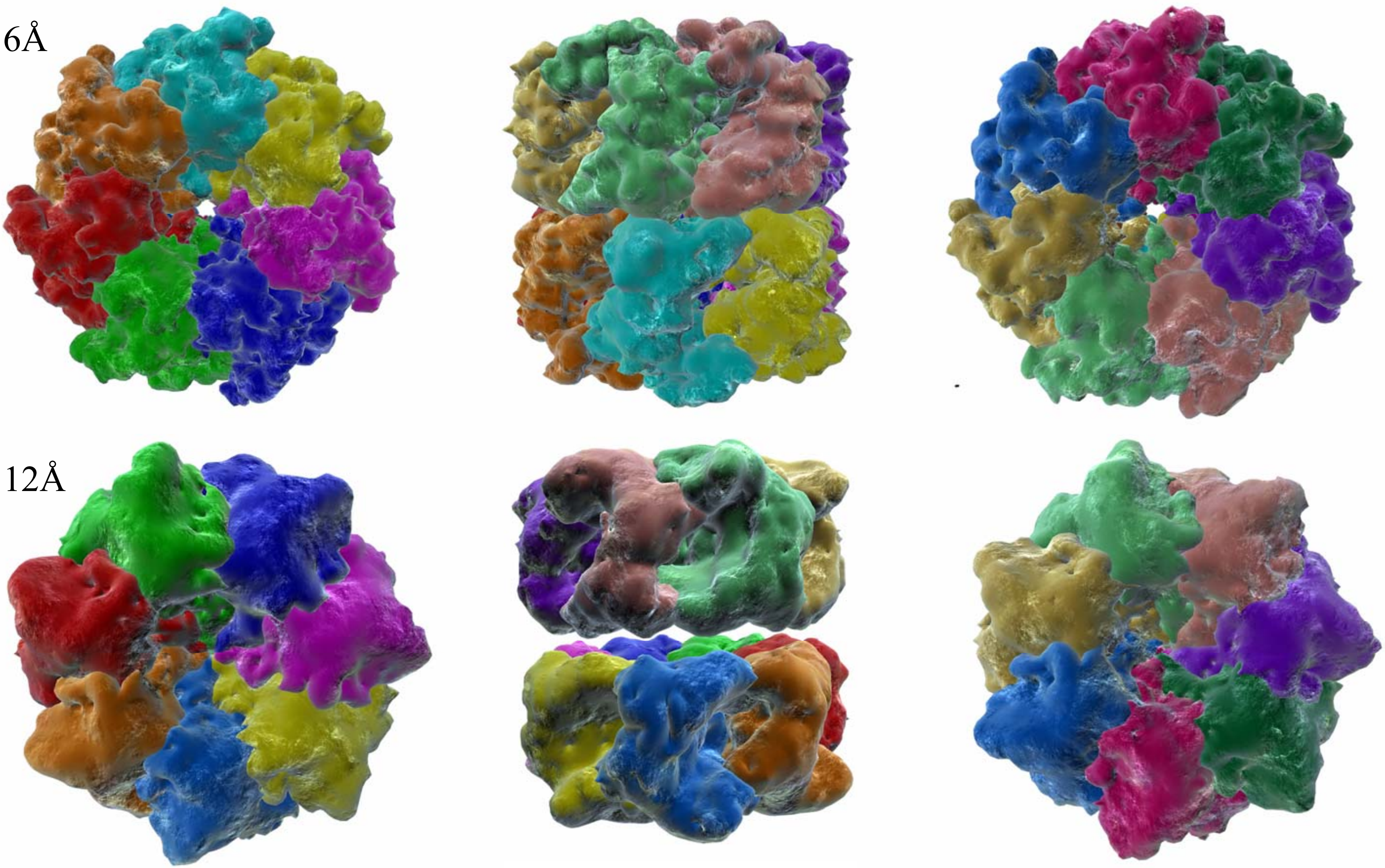
Monomeric Unit of Outer Capsid of RDV



Monomeric Unit of Inner Capsid of RDV

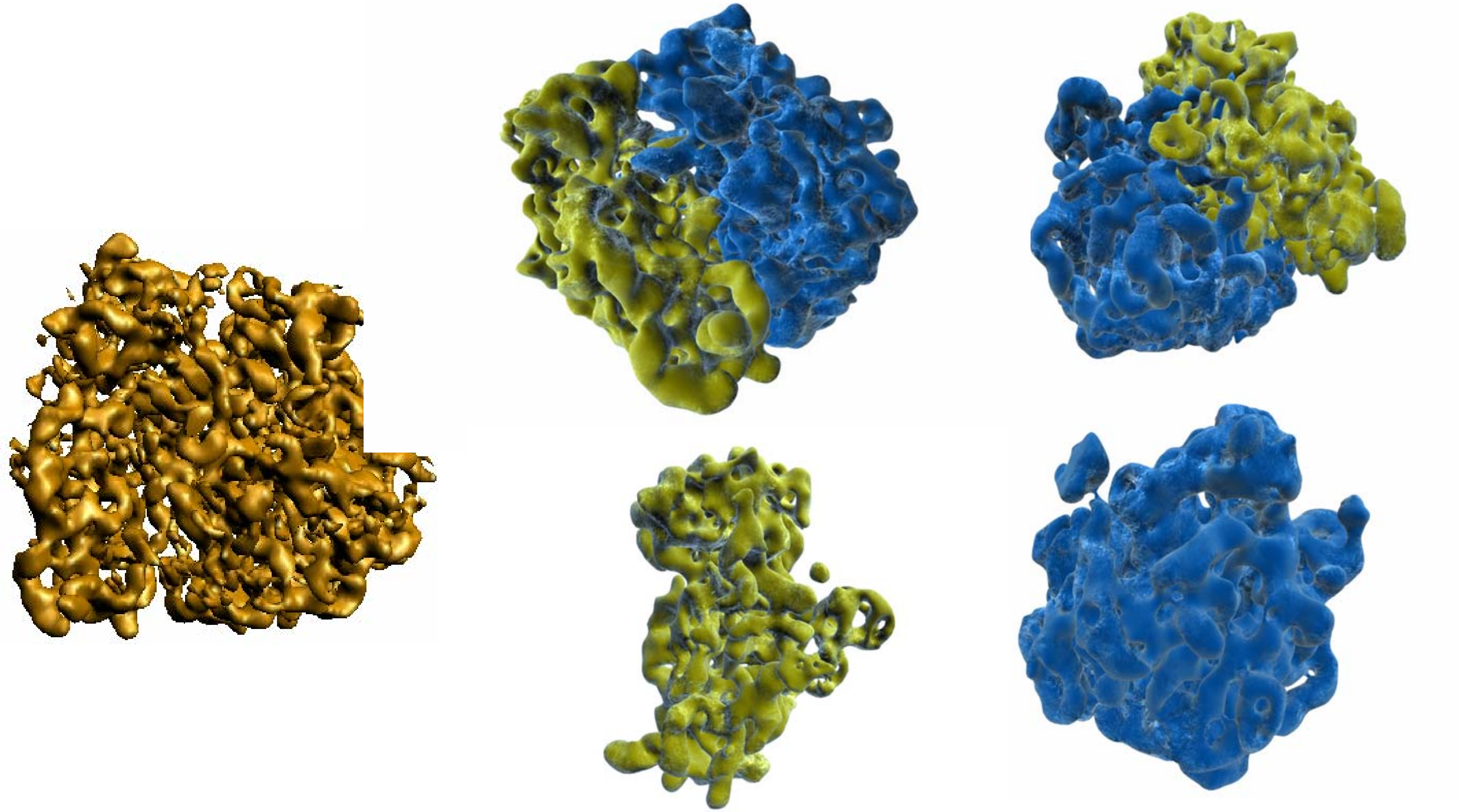


Structure Elucidation Results: GroEL (Bakeoff)



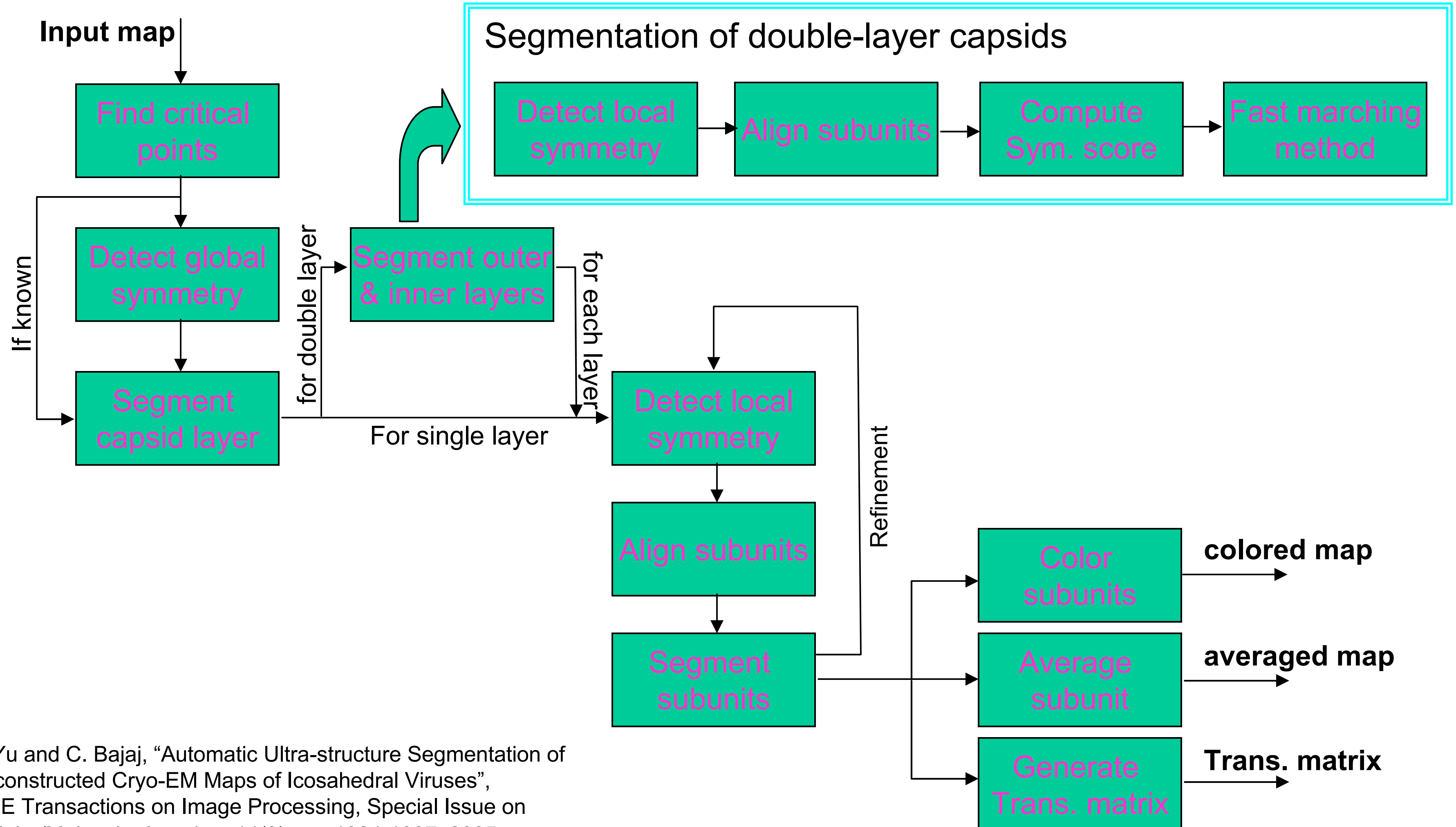
Data courtesy: Dr. Wah Chiu

Segmentation Results: Ribosome (Bakeoff)



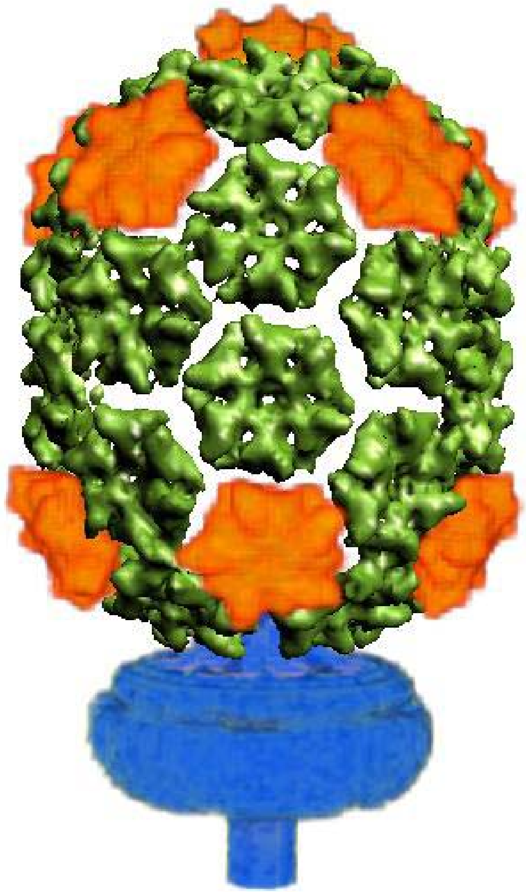
70S ribosome from E. coli complex. 70S-tRNAfMet-MF-tRNAPhe. Data courtesy: EBI & J.Frank

Structure Elucidation for Symmetric Capsid Viruses



Z. Yu and C. Bajaj, “Automatic Ultra-structure Segmentation of Reconstructed Cryo-EM Maps of Icosahedral Viruses”, IEEE Transactions on Image Processing, Special Issue on Cellular/Molecular Imaging, 14(9), pp. 1324-1337, 2005.

Subunit alignment (1): averaging

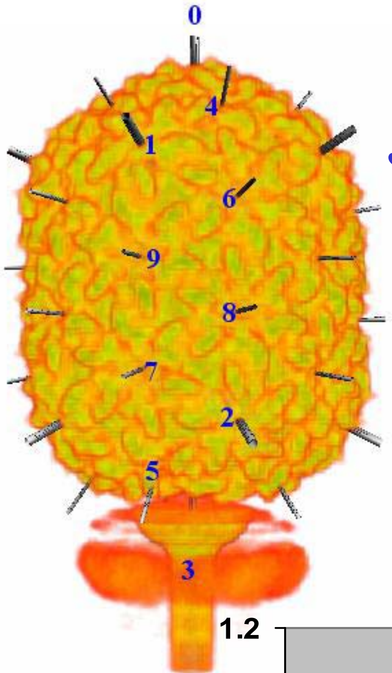


Φ29

The above two pictures (left: outer; right: inner) show the averaged capsid layer, calculated from one 5-fold subunit (orange) and one 6-fold subunit (green). The tail structure (blue) is augmented after the averaging.

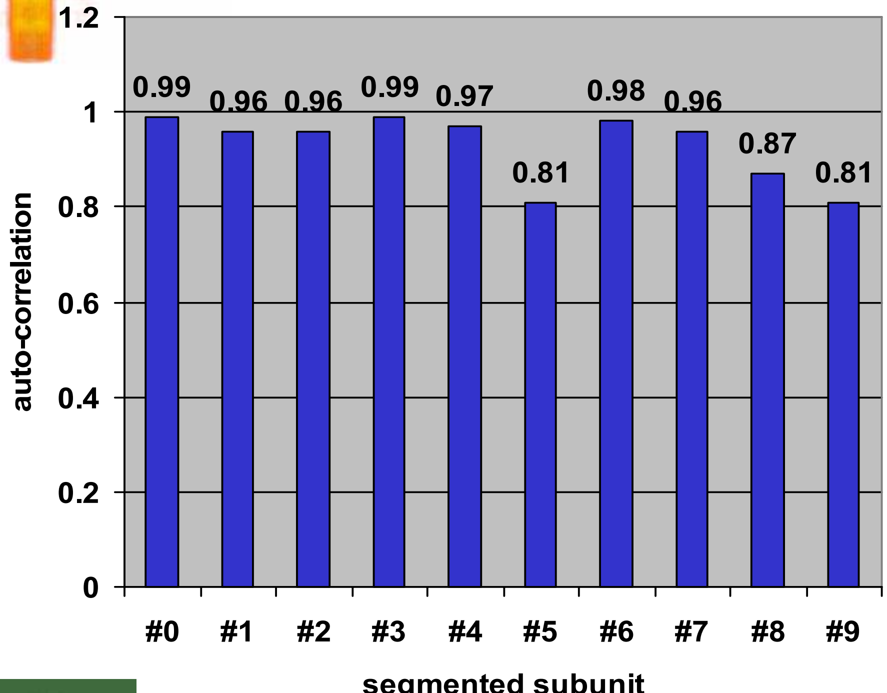
Data courtesy: Tim Baker

Structure Elucidation 1(C): Subunit Alignment



Φ29

- Symmetry score

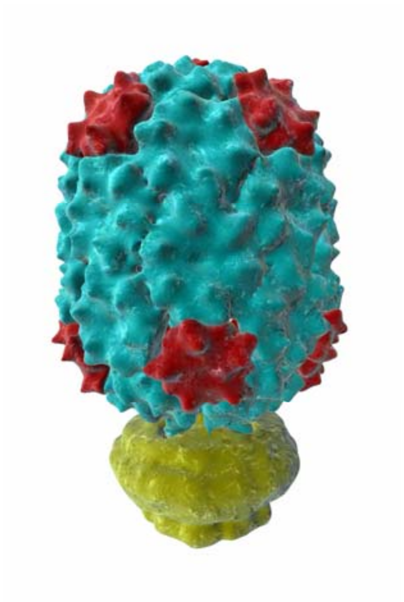


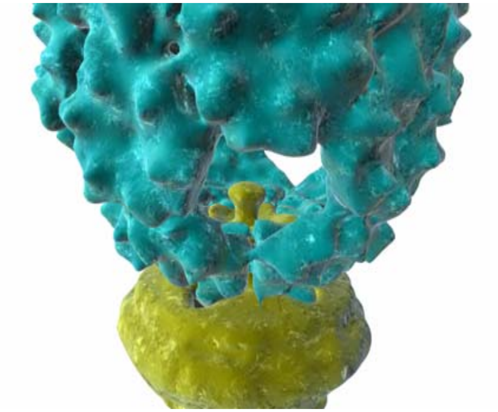
- Cross-correlation

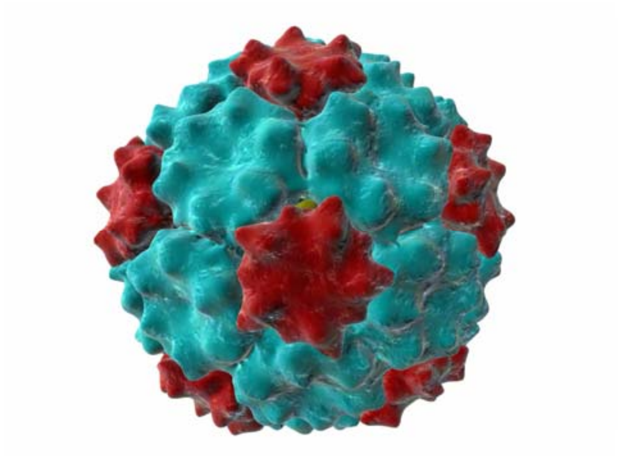
	#0	#1	#2	#3
#0	1	0.95	0.95	0.34
#1	0.95	1	0.96	0.31
#2	0.95	0.96	1	0.31
#3	0.35	0.31	0.32	1

	#4	#5	#6	#7	#8	#9
#4	1	0.79	0.95	0.94	0.87	0.88
#5	0.79	1	0.79	0.78	0.77	0.79
#6	0.95	0.79	1	0.96	0.88	0.88
#7	0.94	0.78	0.96	1	0.89	0.88
#8	0.87	0.77	0.88	0.89	1	0.94
#9	0.88	0.79	0.88	0.88	0.94	1

Structure Elucidation Results: **Φ29**



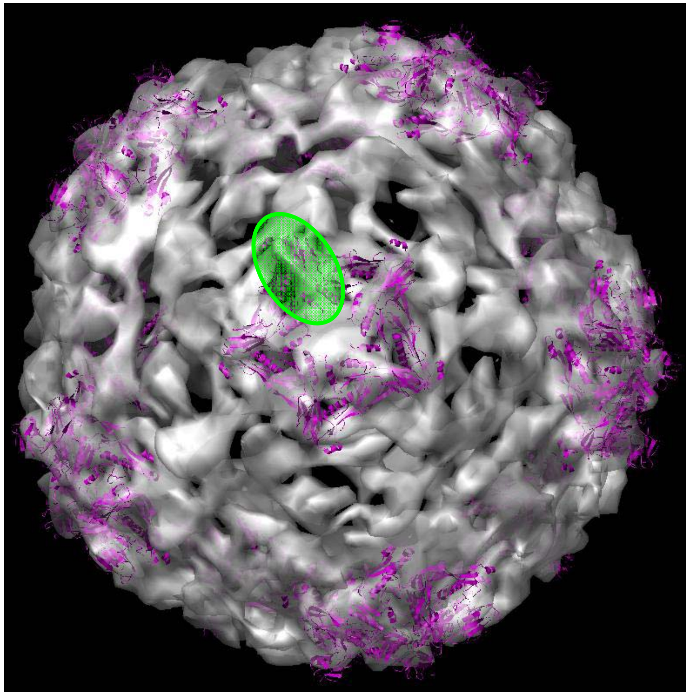


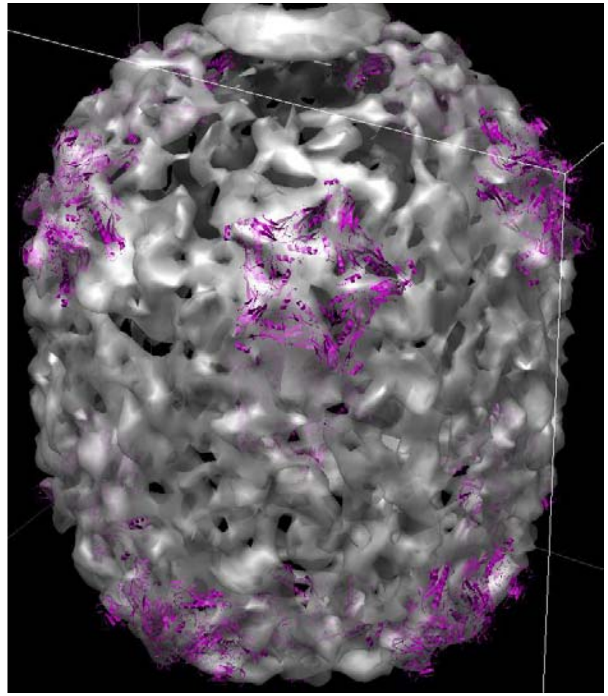


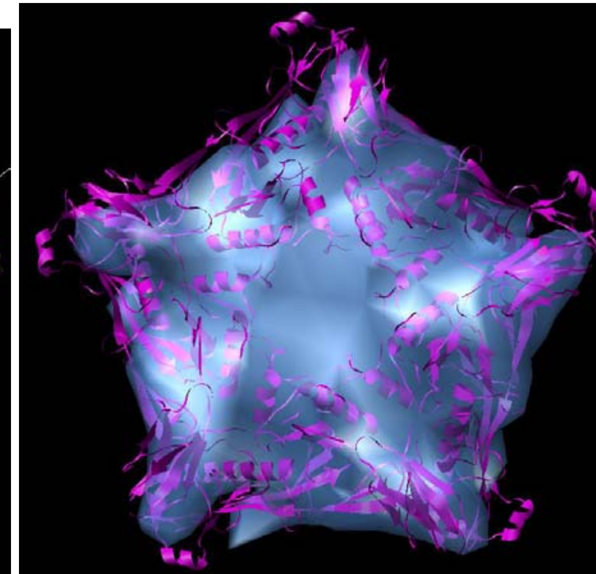
Center for Computational Visualization
Institute of Computational and Engineering Sciences
Department of Computer Sciences

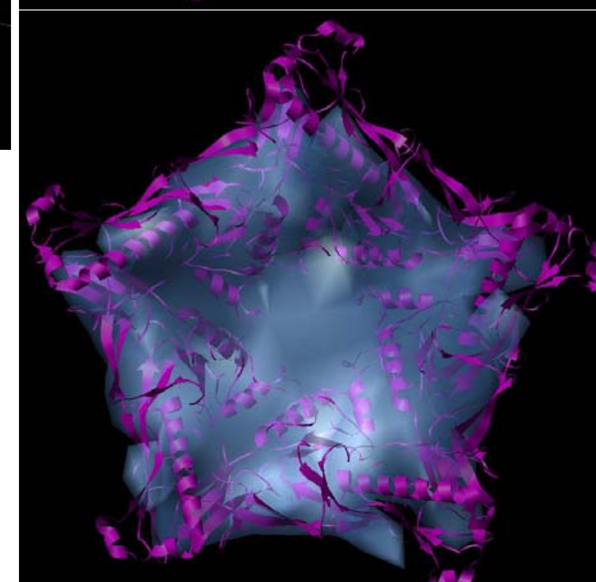
Data courtesy: Tim Baker

Subunit alignment (2): Fitting



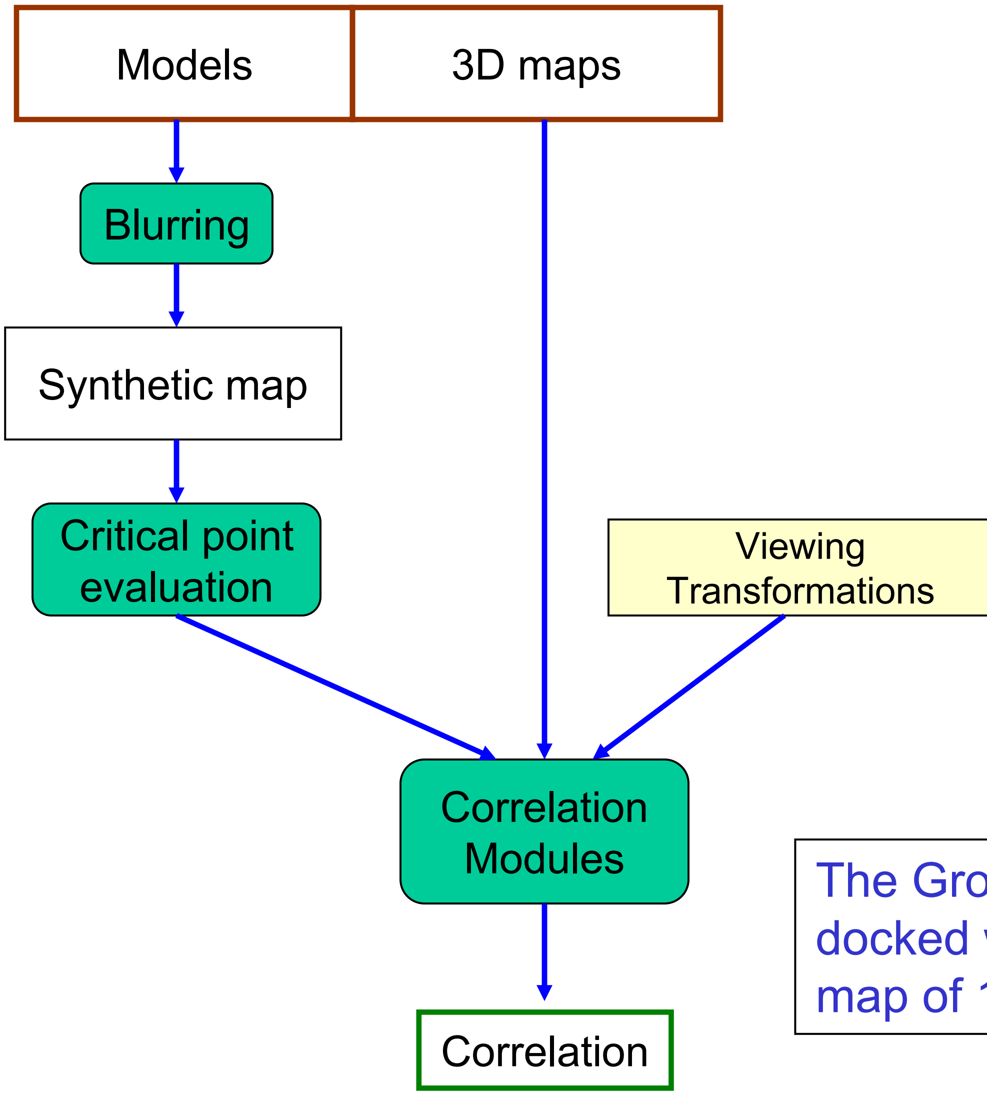


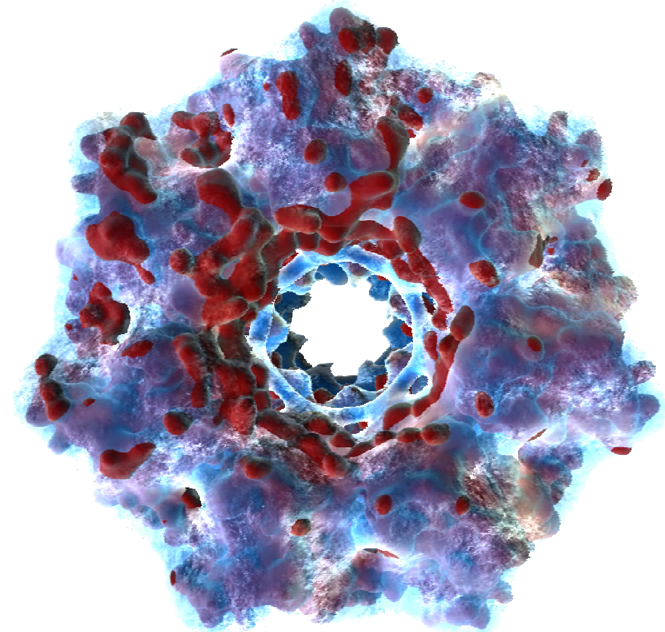


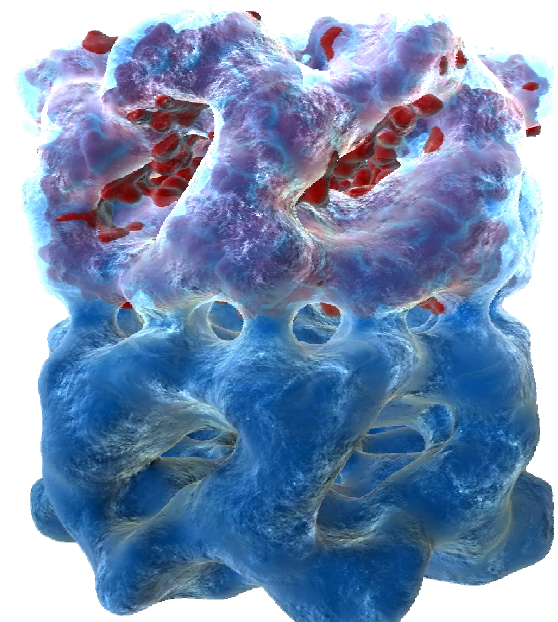


The PDB structure of one monomer is matched & fit into the cryo-EM map (as shaded in green in the left figure). Then all the quasi-symmetric 5-fold subunits are computationally fit with the PDB structure using the transform matrices obtained in subunit alignment. Similar procedure can be applied to all 6-fold subunits.

"Interactive" Fitting

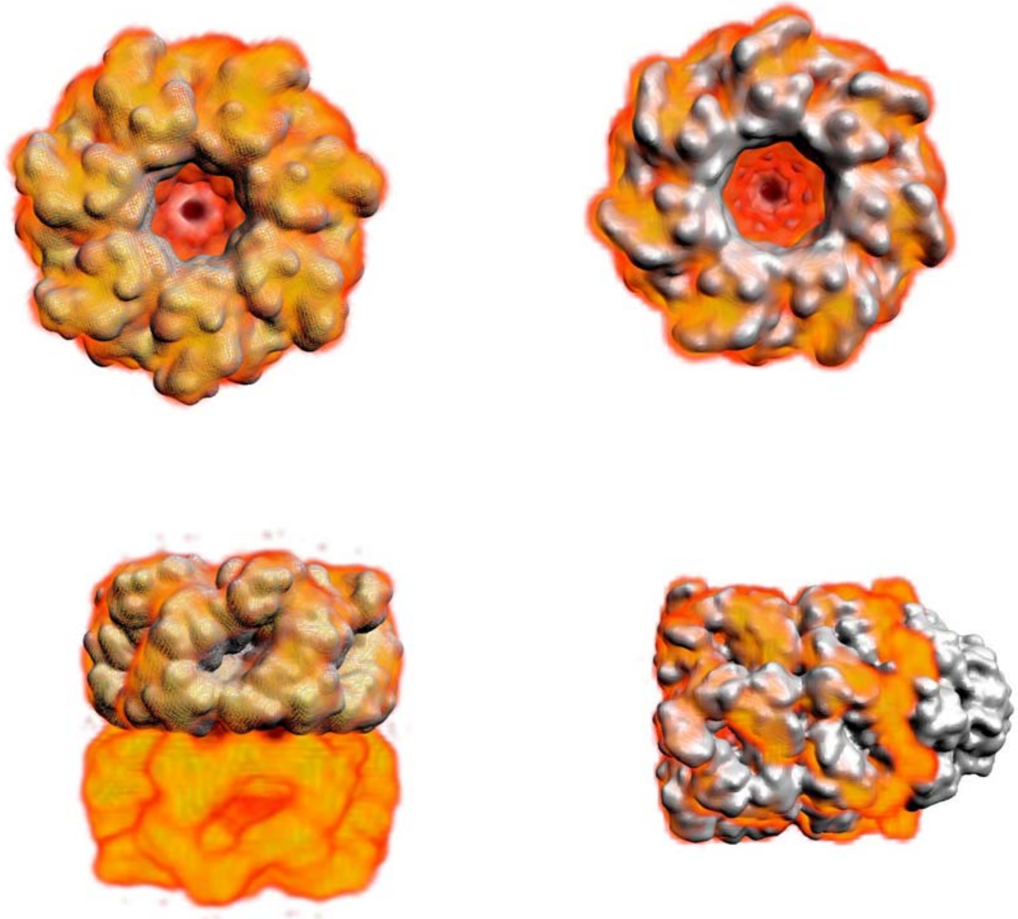






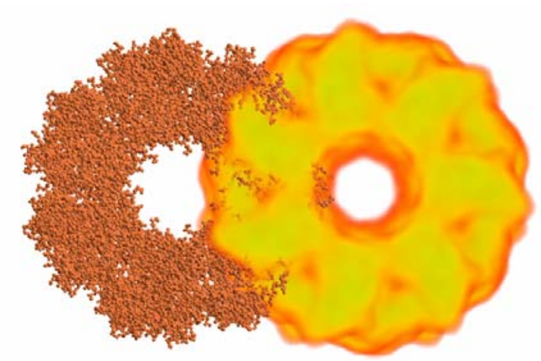
The GroEL map docked with a blur map of 1OEL.PDB

Gro-EL: X-ray structures docked in Cryo-EM

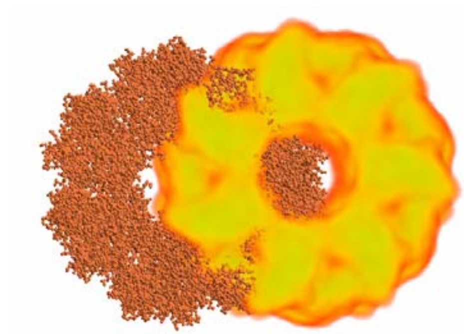


Interactive Correlation Analysis

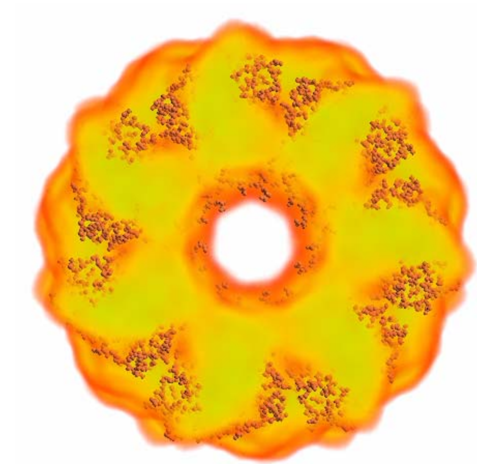
C = 0.2235



C = 0.269

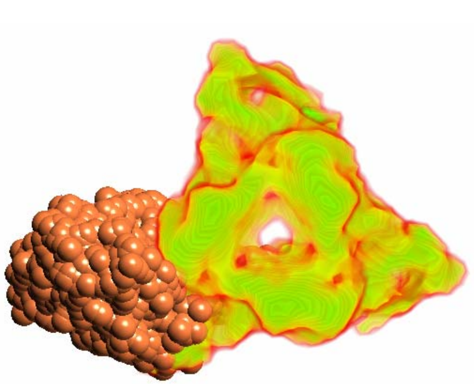


C = 0.593

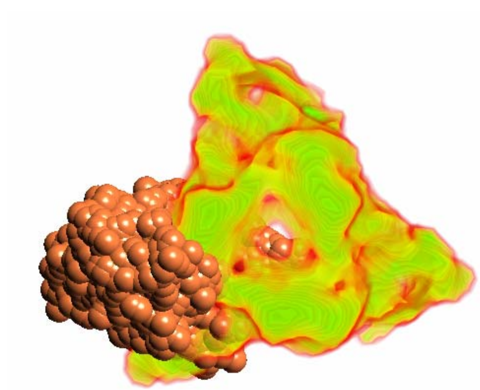


12A GroEL map and 1OEL.pdb

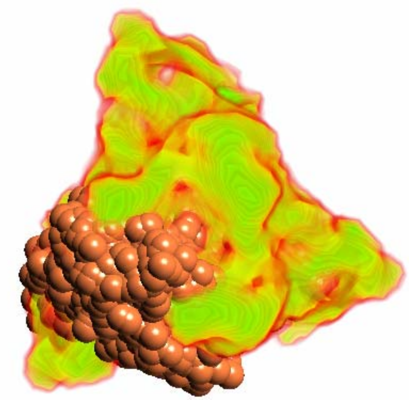
C = 0.208



C = 0.387



C = 0.542

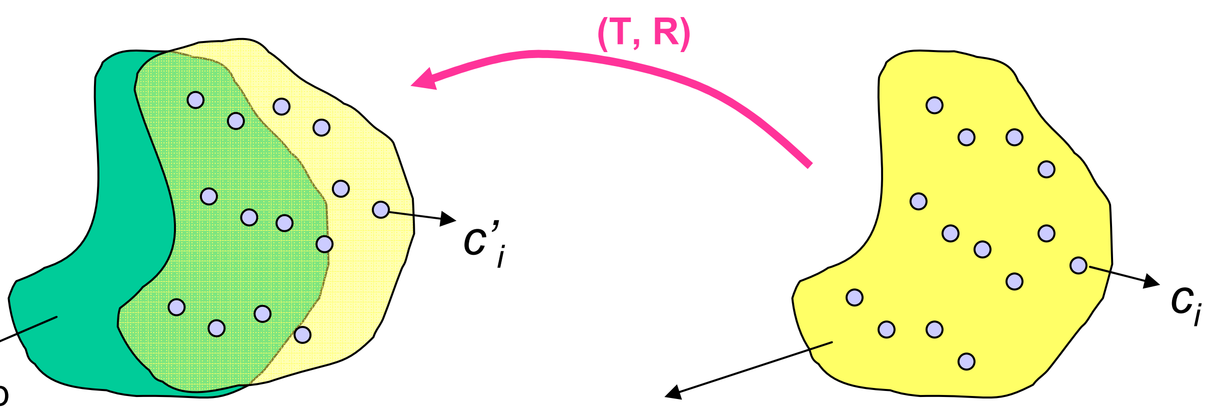


6.8A RDV map and 1UF2.pdb

Approximate Correlation Analysis

$$score = 1 - \frac{\sum_{i=1}^{N} |f(c_i) - g(c_i^{'})|}{\sum_{i=1}^{N} \max(f(c_i), g(c_i^{'}))}$$

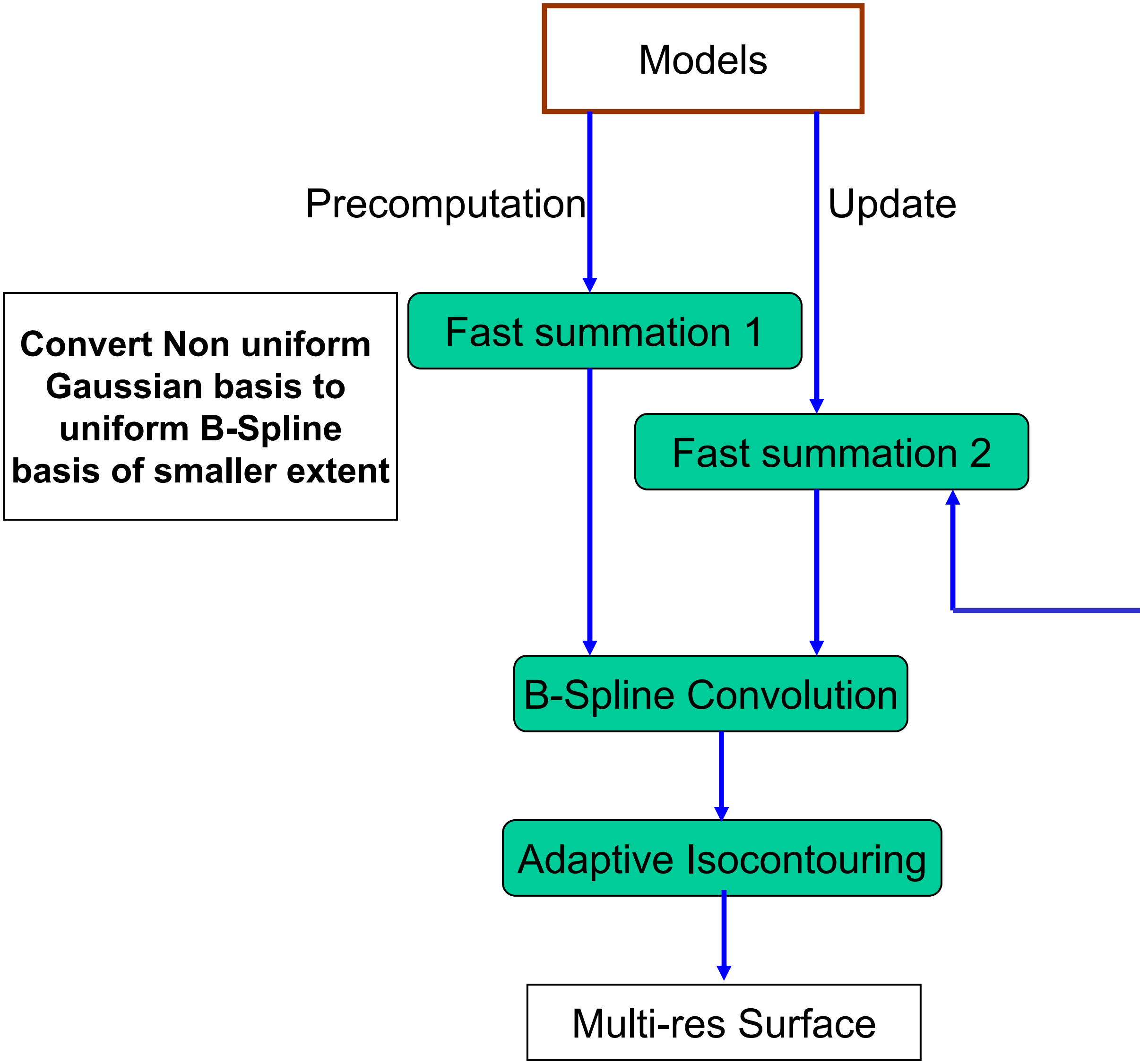
Where ***f*** is the normalized density function of the blurred crystal structure;
g is the normalized density function of the cryo-EM map;
c_i , i=1,2,…N, are the critical points of the blurred crystal structure;
c'_i , i=1,2,…N, are the transformations of the critical points.

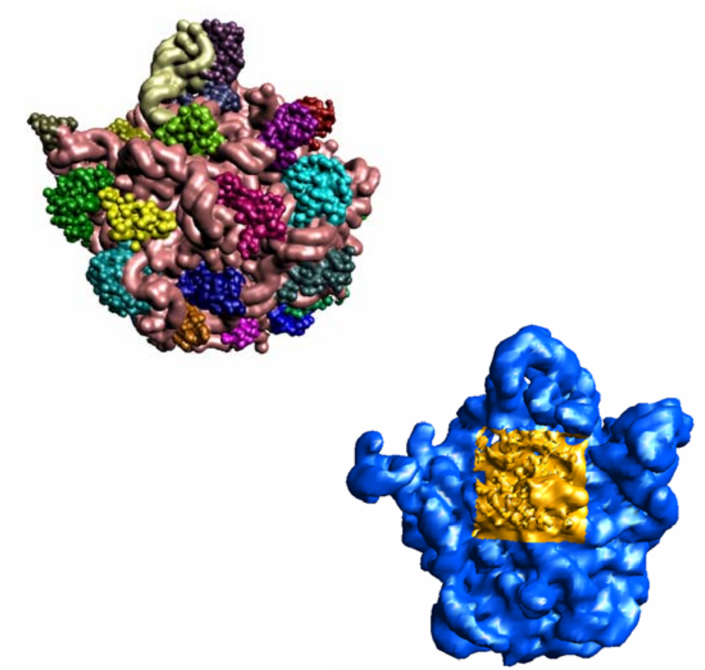


Cryo-EM map
(density function: *g*)

Blurred model
(density function: *f*)

Multi-resolution Molecular Surfaces

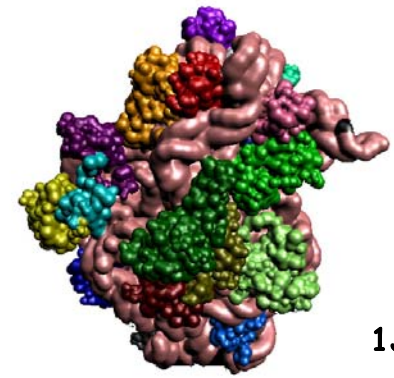




Blurring I

- For a molecule with M atoms, we can define a 3D electron density map as

$$f_{elec\_dens}(\vec{\mathbf{x}}) = \sum_{i=1}^{M} G_i(\vec{\mathbf{x}}) \qquad \mathbf{x} \in \mathbf{R}^3$$

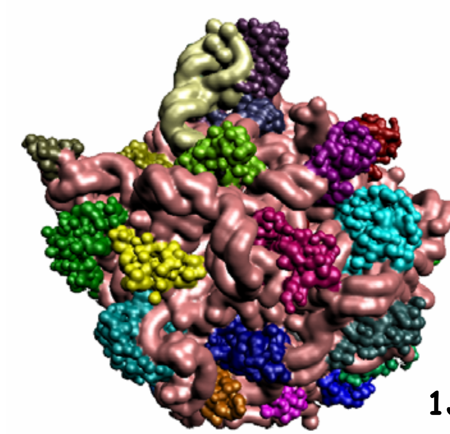


- For quadratic decay kernels, $A_i = e^d$:

$$f_{elec\_dens}(\mathbf{x}) = \sum_{i=1}^{M} A_i e^{-\frac{d}{r^2}\mathbf{x}^2} \delta(c_i)$$

- For linear decay kernels, $A_i = e^{dri}$:

$$f_{elec\_dens}(\mathbf{x}) = \sum_{i=1}^{M} A_i e^{-d|\mathbf{x}|} \delta(c_i)$$



Atomic Shape Parameters

- Isotropic Quadratic Kernel

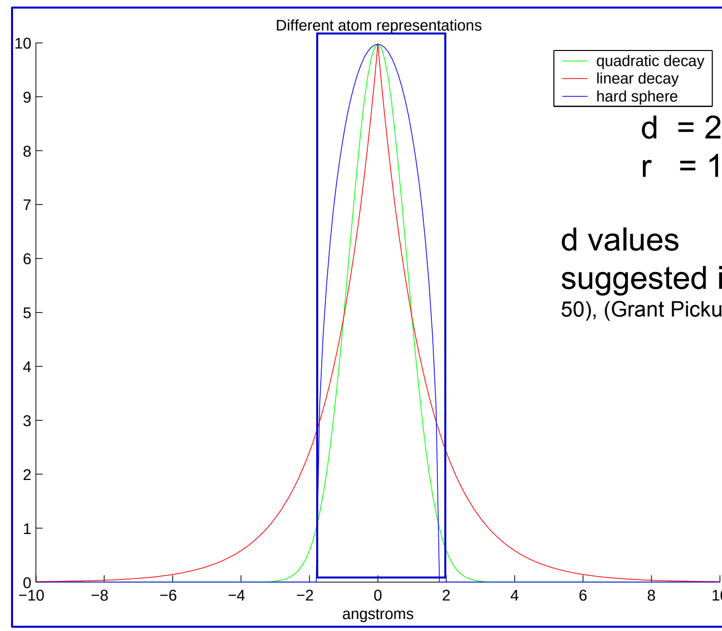
$$G_i(\mathbf{x}) = e^{-\frac{d}{r_i^2}\left((\mathbf{x}-\mathbf{x_{ci}})^2 - r_i^2\right)}$$

- Isotropic Linear Kernel

$$G_i(\mathbf{x}) = e^{-d\left(|\mathbf{x}-\mathbf{x}_{ci}| - r_i\right)}$$

  - where
    - The decay d controls the shape of the Gaussian function.
    - The van der Waals radius is r_i
    - The center of the atom is $\mathbf{x}_c$.

- Anisotropic Kernels



d = 2.3
r = 1.8A

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

Blurring II

- For quadratic decay kernels, $A_i=e^d$:

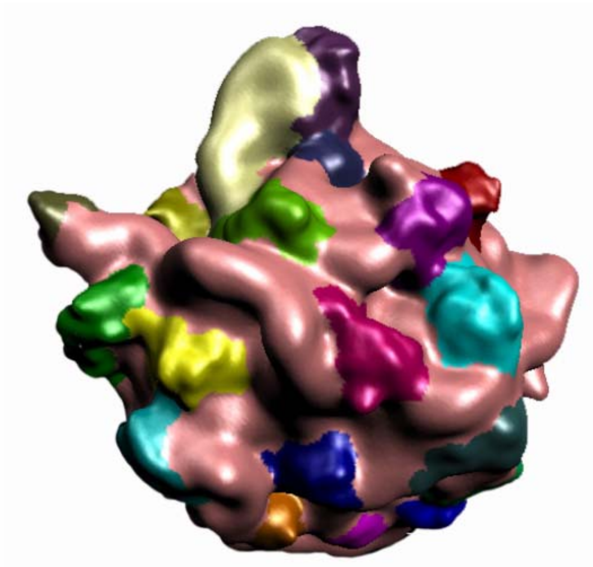
$$f_{elec\_dens}(\mathbf{x}) = \sum_{i=1}^{M} A_i e^{-\frac{d}{r^2}\mathbf{x}^2}\delta(c_i)$$

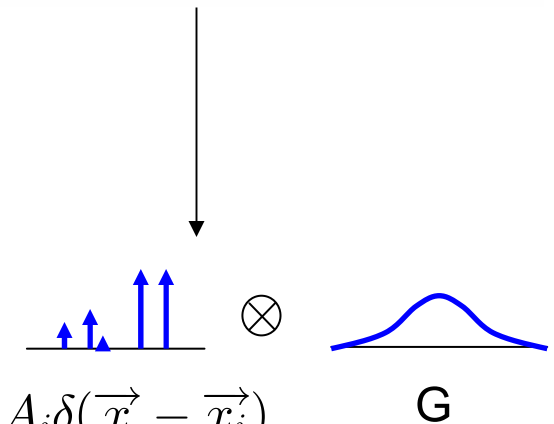
- For linear decay kernels, $A_i=e^{dr_i}$:

$$f_{elec\_dens}(\mathbf{x}) = \sum_{i=1}^{M} A_i e^{-d|\mathbf{x}|}\delta(c_i)$$

- For above kernels G:

$$f_{elec\_dens}(\mathbf{x}) = G \otimes \sum_{i=1}^{M} A_i \delta(c_i)$$



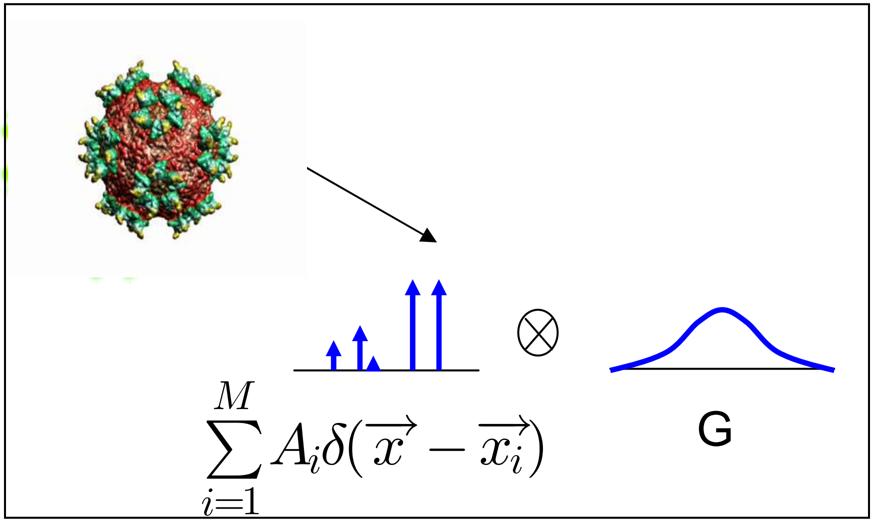


$$\sum_{i=1}^{M} A_i \delta(\overrightarrow{x} - \overrightarrow{x_i})$$

FFT Solutions & Approximations

- Use the convolution theorem.
  - Convolution in spatial domain = multiplication in frequency domain.
- The 3D electron density map is a convolution of a Kernel with delta functions at the atom centers
  - $f_{elec\_dens} = FFT^{-1}(FFT(\text{Kernel}) \times FFT(\text{atom centers}))$

- **Accuracy with Speed Tradeoff:**
  - The kernels are smooth functions. Hence only the first few frequencies need to be computed to obtain the summation. Also, convolution is a smoothing operator.
  - We can use an approximate FFT algorithm to calculate the low frequencies of atom centers to a high accuracy at a lower cost than a full FFT.



Evaluation of 3D Electron Density Map

Evaluation of the Electron Density of an M atom molecule at multiple points N

(typically N ~ 100* M)

[Naiive Algorithm]

For each of the output points N, calculate the summation due to each of the M atomic kernels

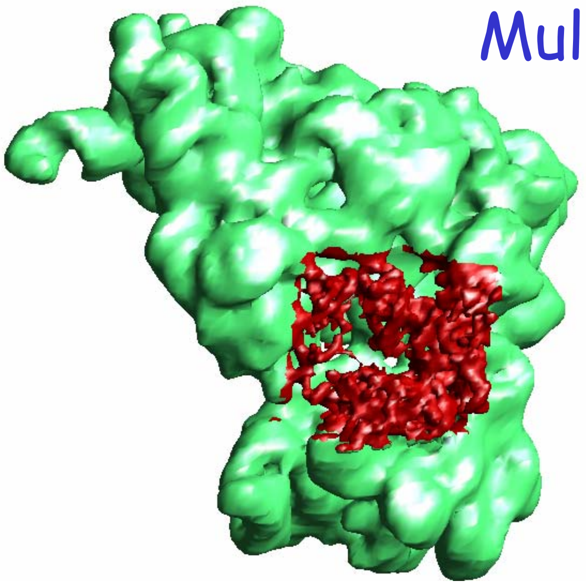
- **Cost**: O(NM) → very high!

Fast Approximate FT Solution

– f_{elec_dens} = *FFT*^*-1* (*FFT*(Kernel) x *FFT*(atom centers))

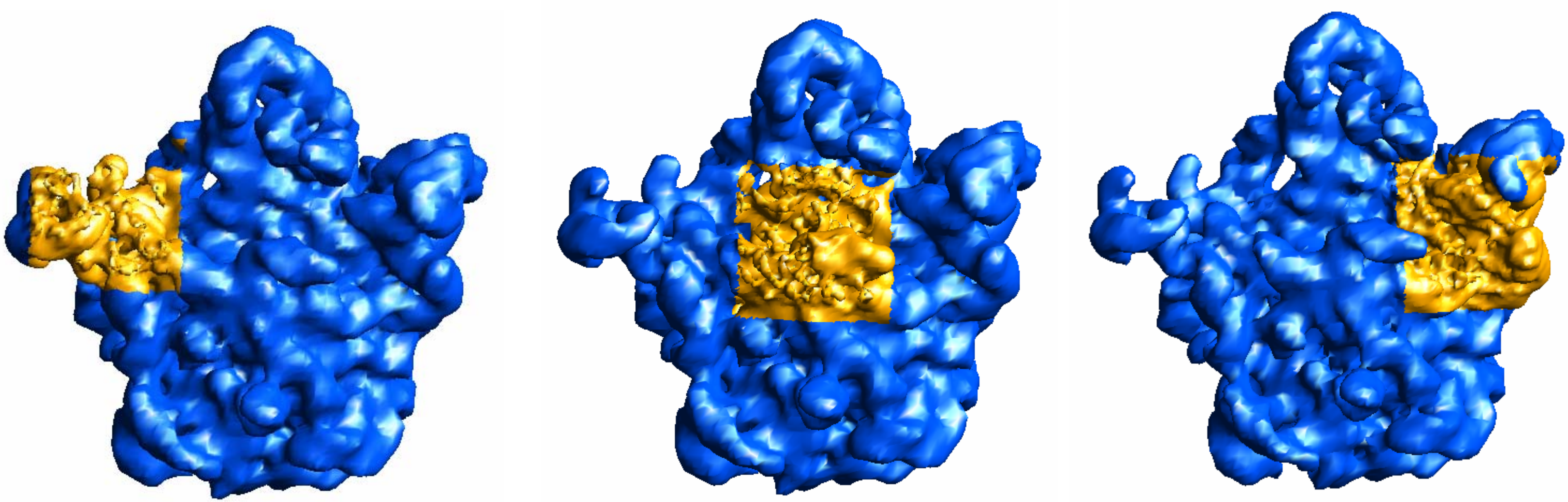
- Space complexity:
  - O(M+N)
- Time complexity:
  - Irregularly spaced output evaluation points
    - O(MlogM + N) with larger constants
  - Regularly spaced output evaluation points
    - O(MlogM + N log N) with small constants

Multi resolution Molecular Surfaces (Bakeoff)



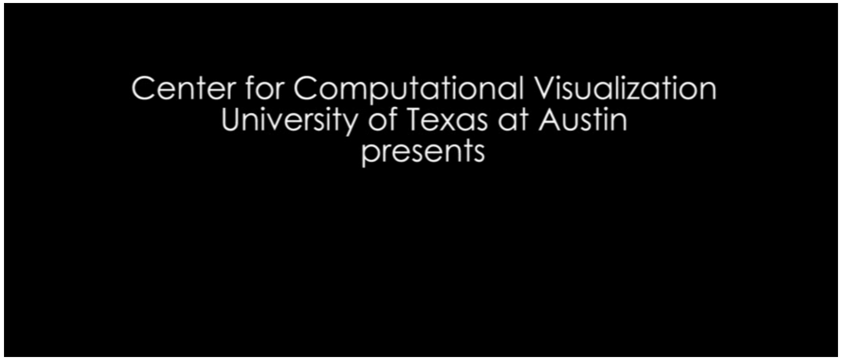
1PNX.PDB

1PNY.PDB



Interrogative Scalable Visualization

- **Techniques**
  - Surface/2D Textures/3D Texture based rendering exploit hardware acceleration
  - Multi-dim Transfer Functions for Regions of Interest
  - Multi-resolution processing
  - Compressed data processing
  - Parallelism (back-end PC cluster)



Communication & Education: Beyond silent movies!

C.Bajaj, I. Ihm, S. Park, *ACM Trans. on Graphics*, 20, 1, 10-28, 2001

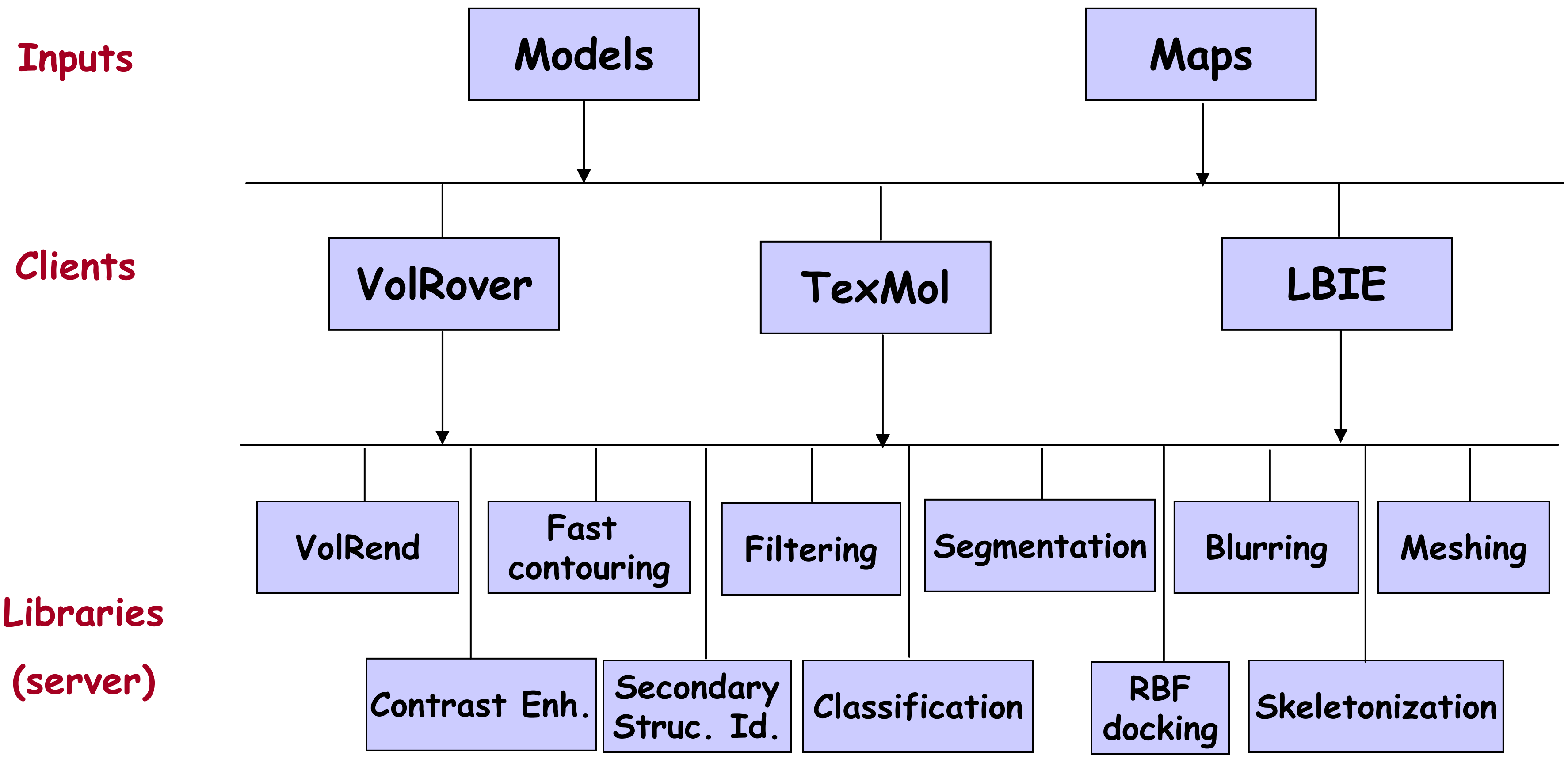
C Bajaj, P Djeu, V Siddavanahalli, A Thane, *IEEE visualization*, 2003. 243-250.

C. Bajaj, J. Castrillon-Candas, S. Vinay, A. Xu, *Structure*,13,3,2005,463-471

X. Zhang, C. Bajaj, *IEEE Symp. On Parallel, Large Data Visualization*, 2001, 51-58

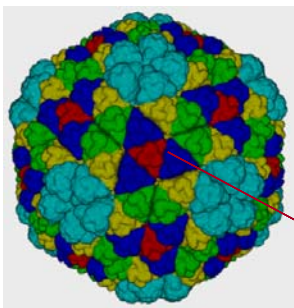
Molecular Visualization and Processing (MVP) Environment

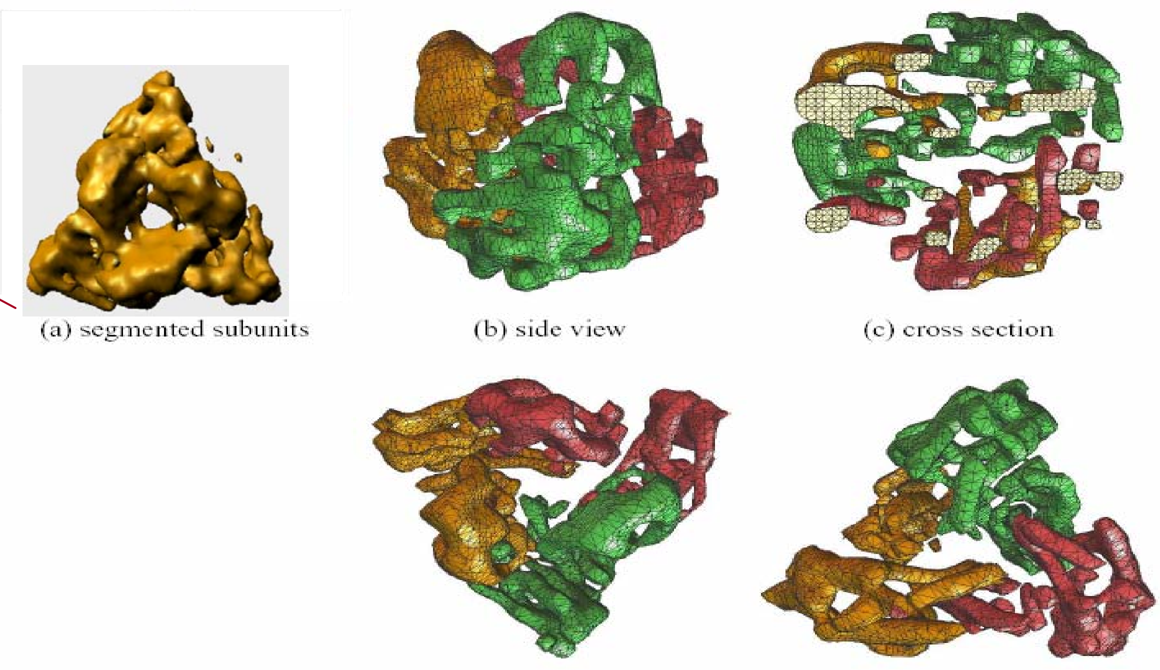
Open Source/Public Domain under LGPL
(http://ccvweb.csres.utexas.edu/software/)



Modeling 2(A): Finite Element Meshing

- LBIE : also supports Integral/Derivative Property computations (areas, volumes, curvatures)





(a) segmented subunits (b) side view (c) cross section

(d) top view (e) bottom view

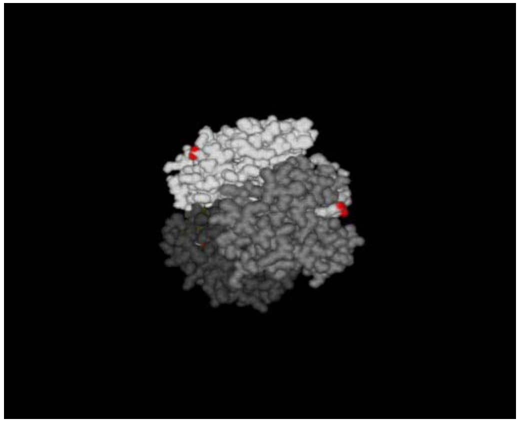
•Y. Zhang, C. Bajaj, B. Sohn, *CMAME*, Spec. Issue on Unstruc. Mesh. Generation., 2004

•Y.Zhang, C.Bajaj, G. Xu, 14th Intl Meshing Roundtable, San Diego2005.

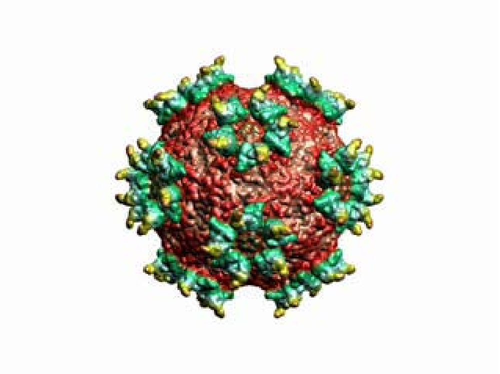
•Y. Zhang, C. Bajaj, *CMAME*,2005 (in press)

Summary: Algorithms & Tools

- **Structure elucidation**: filtering, contrast enhancement, segmentation, skeletonization, subunit identification
- **Structure Modeling**: finite element meshing, spline representations(A-spline,RBF representations) for structural fitting & complementary docking
- **Visualization**: multi-dimensional transfer functions, surface and volume texture rendering, wavelet compression, hierarchical representations, cluster based parallelism



VolRover



TexMol

C. Bajaj, Chap in Modeling Biology, MIT Press 2005

C. Bajaj, Z. Yu, Chap in Handbook of Comp. Mol. Bio, Chapman&Hall/CRC Press, (2005)

The CVC Team and Collaborators

- *Personnel*
  - **Inderjit Dhillon (Assoc. Director)**
  - **Albert Chen (CS, Ph.D)**
  - **Katherine Clarridge (MBE,M.S.)**
  - **KL Chandrasekhar (ME, Ph.D.)**
  - **Tamal Dey (OSU) ****
  - **Samrat Goswami (PostDoc, CS)**
  - **Rick Hankins (Res. Scientist)**
  - **Insung Ihm (SNU, S. Korea)****
  - **Sangmin Park (CS, Ph.D.)**
  - **Bong-June Kwon (CS, M.S.)**
  - **Bong-Soo Sohn (Stanford U)****
  - **Jason Sun (Res. Scientist)**
  - **John Wiggins (Res. Scientist)**
  - **Vinay Siddahanavalli (CS, Ph.D.)**
  - **Guoliang Xu (AS,China)****
  - **Zeyun Yu (CS, Ph.D)**
  - **Xiaoyu Zhang (CSU)****
  - **Jessica Zhang (PostDoc, ICES)**
  - **Wenqi Zhao (ICES, Ph.D.)**

- *Senior Collaborators*
  - Manfred Auer (LBL)
  - Nathan Baker (Wash. U.)
  - Helen Baker, Cathy Lawson (Rutgers U)
  - Tim Baker (UCSD)
  - Tom Bartol (SALK)
  - Luis Caffarelli (ICES)
  - Wah Chiu, Matt. Baker (Baylor)
  - Leszek Demkowicz (ICES)
  - Gregory Gladish, J. Hazle (MD Anderson)
  - Tom Hughes (ICES)
  - Andy McCammon (USCD)
  - Tinsley Oden (ICES)
  - Art Olson, M. Sanner, D. Goodsell, Charlie Brooks, V. Reddy (Scripps)
  - Peter Rossky (ICES)
  - Andre Sali (UCSF)

- ***Funding***
  - **NIH: P20(planning), R01**
  - **NSF: ITR , DDDAS**
  - **Whitaker Foundation**