Quantitative Visualization of Static and Dynamic Biological Complexes

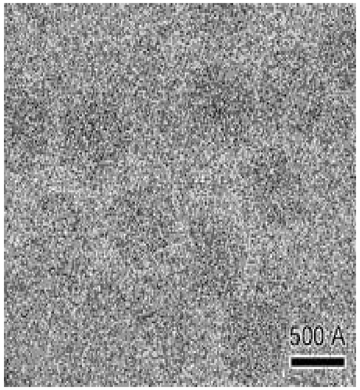
Chandrajit Bajaj

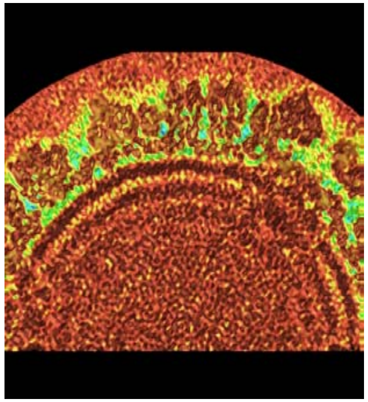
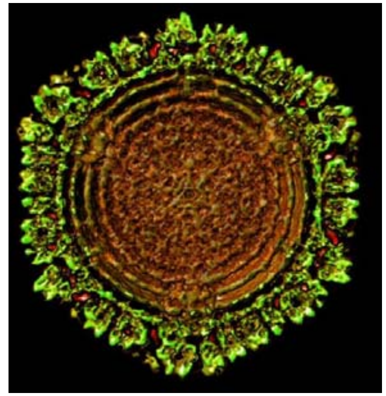
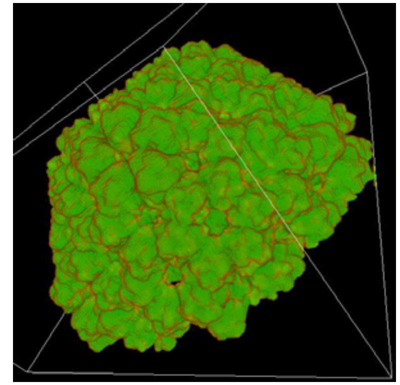
http://www.ices.utexas.edu/CCV

Outline

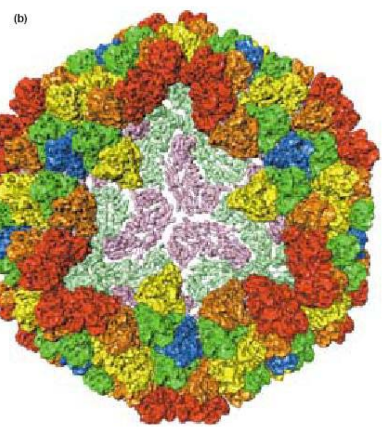
- Domains
  - Cryo-EM Maps
  - Tomographic
  - PDB structures (shape), Properties (electrostatics, hydrophobicity)
- Techniques
  - Image Processing (Scalar/Vector Filtering, Contrast Enhancement, Skeletonization, InPainting)
  - Finite Element Meshing (Linear, Higher-Order)
  - Analysis (Area, Volumes, Combinatorics, Topological)
  - Compression (Hierarchical, Progressive)
  - Visualization (Surface+Volume Rendering, Texture Rendering)

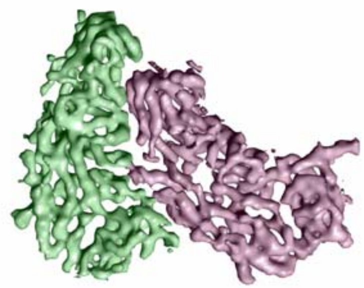
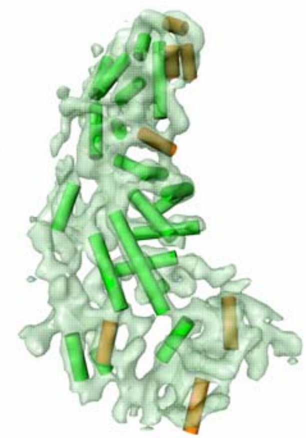
Imaging to Structure to Modeling to Visualization




Cryo-EM →Anistropic and Vector Diffusion Filtering → Structure Segmentation → Sub-Atomic Modeling → Functional Analysis →Visualization

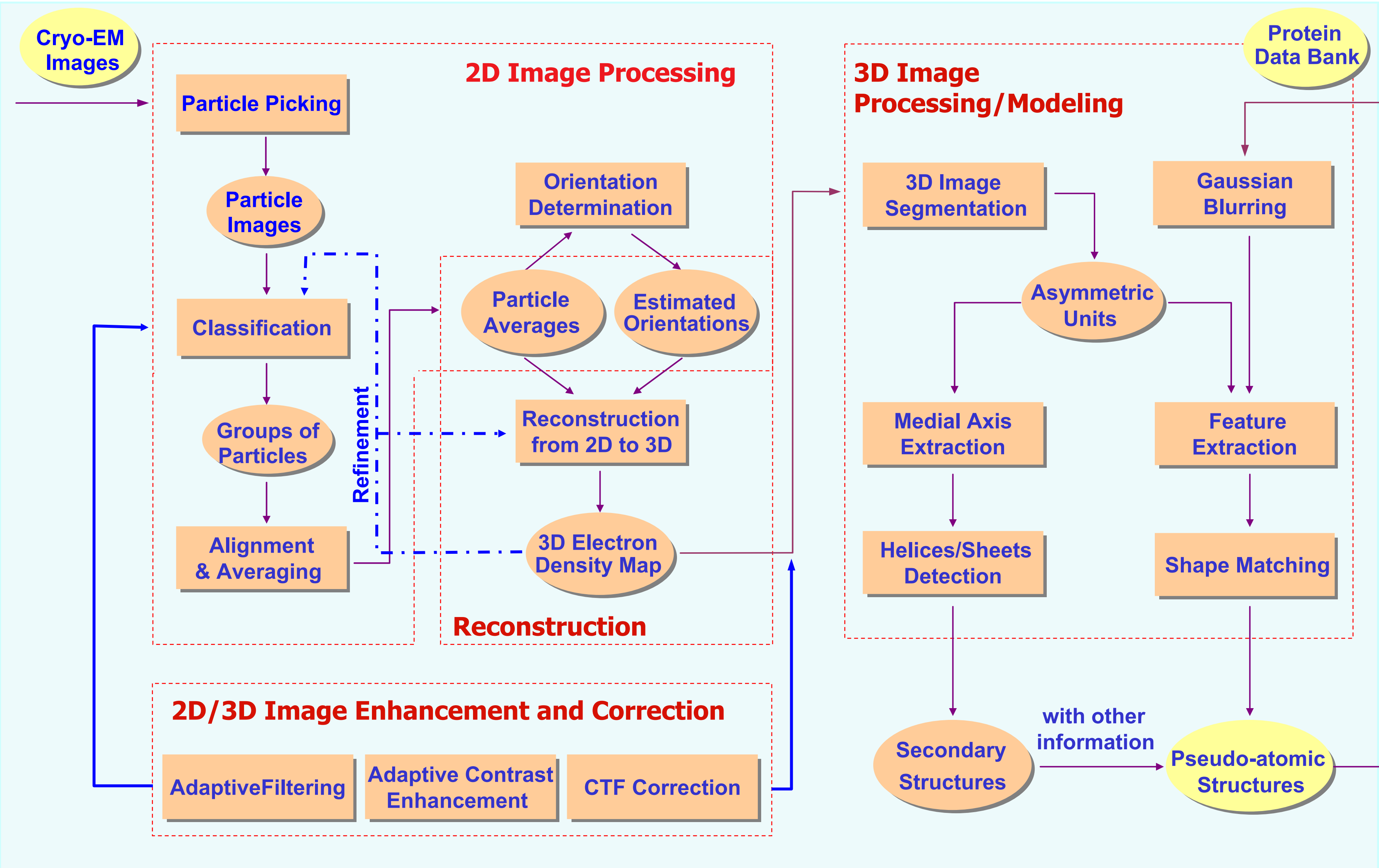


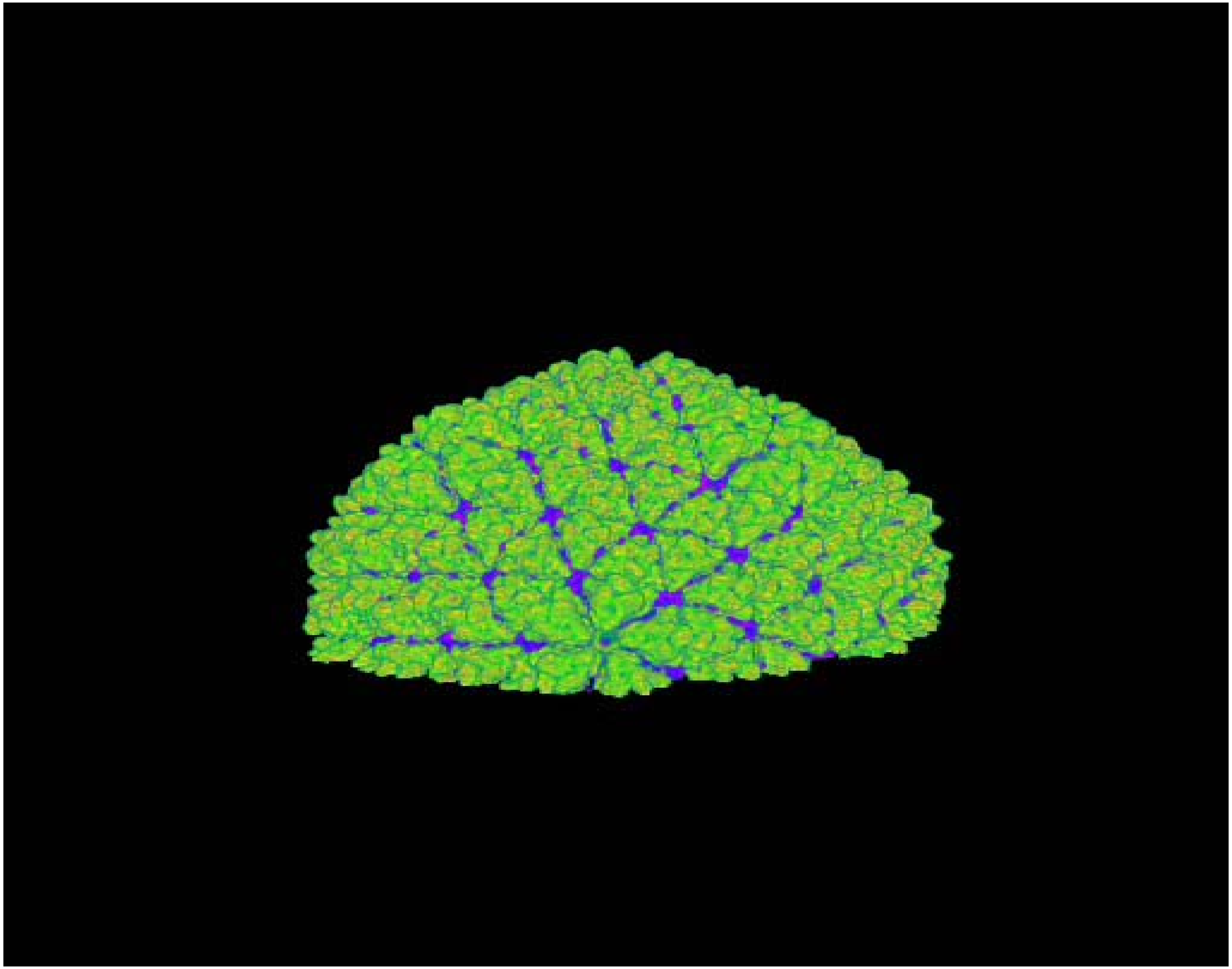
(Collaborators: Wah Chiu,NCMI, Baylor College of Medicine, A. Sali, UCSF)

**Sponsored by NSF-ITR

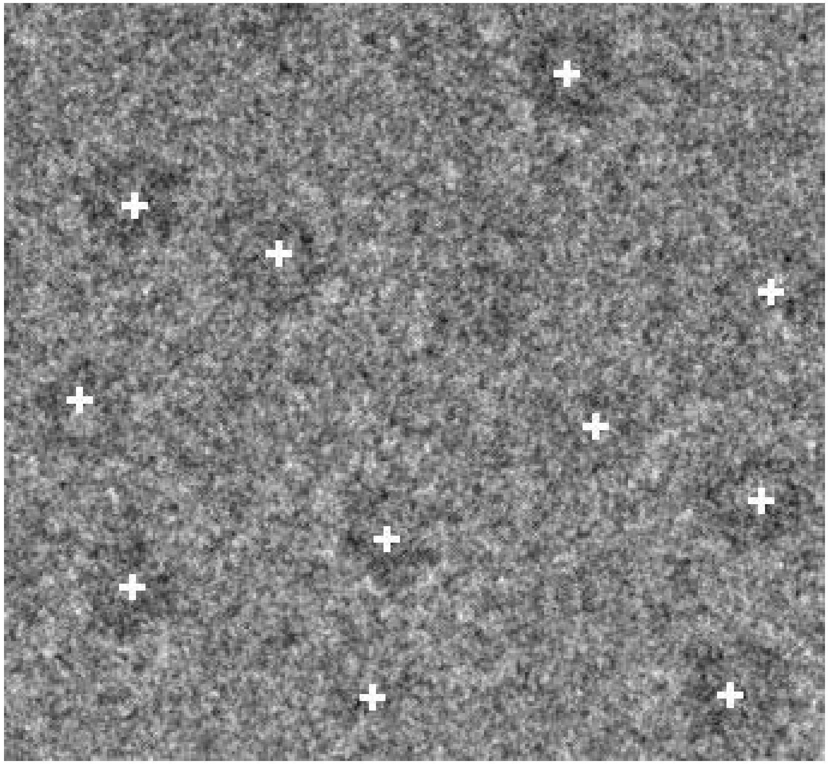
Computational Pipeline



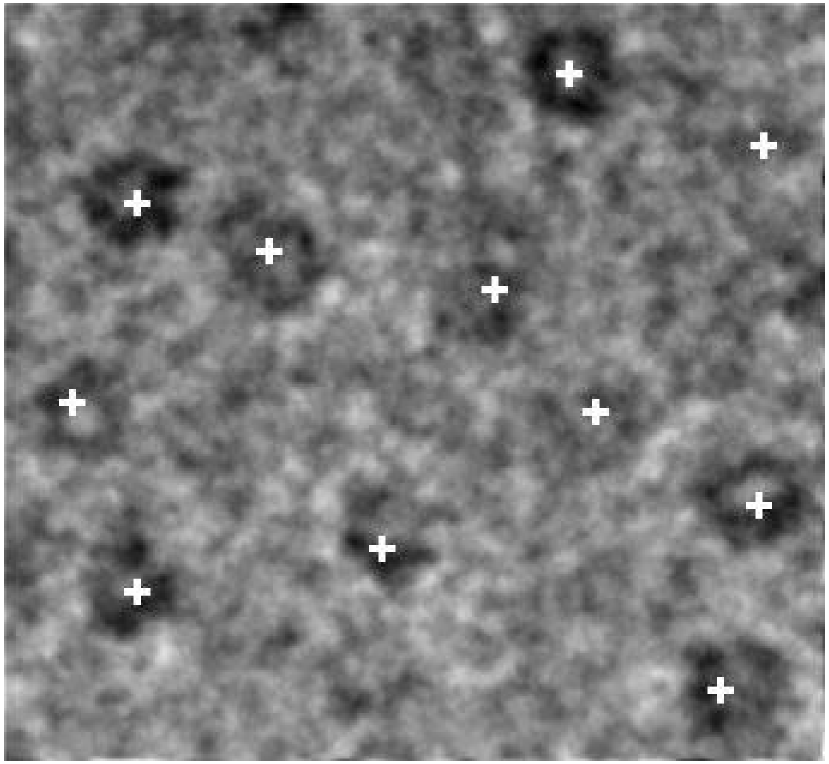
Rice Dwarf Virus



Would Image Filtering Help Structure Determination ?

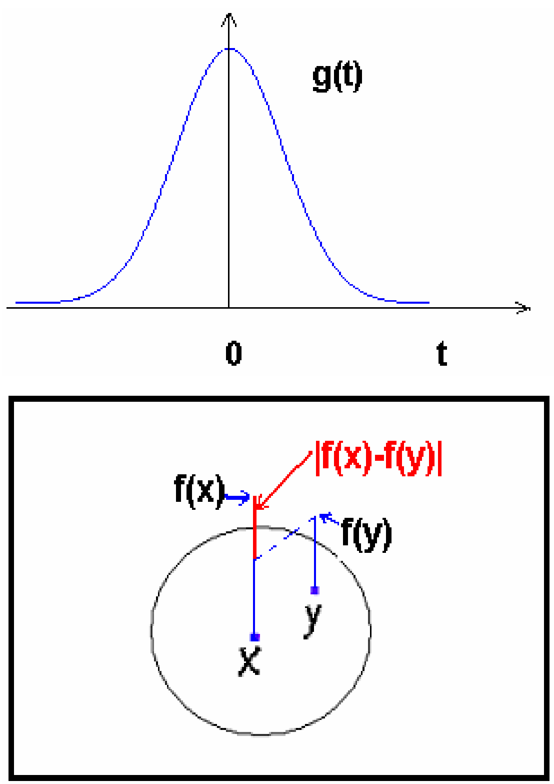


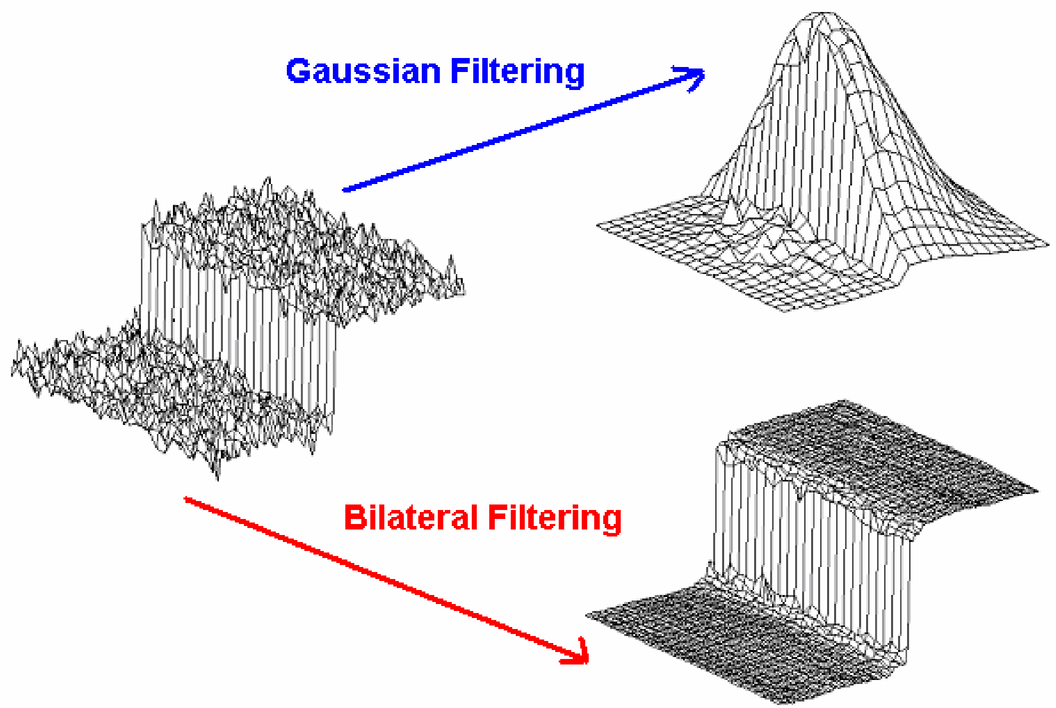
Original image



After anisotropic diffusion

Image Filtering: Gaussian vs Bilateral



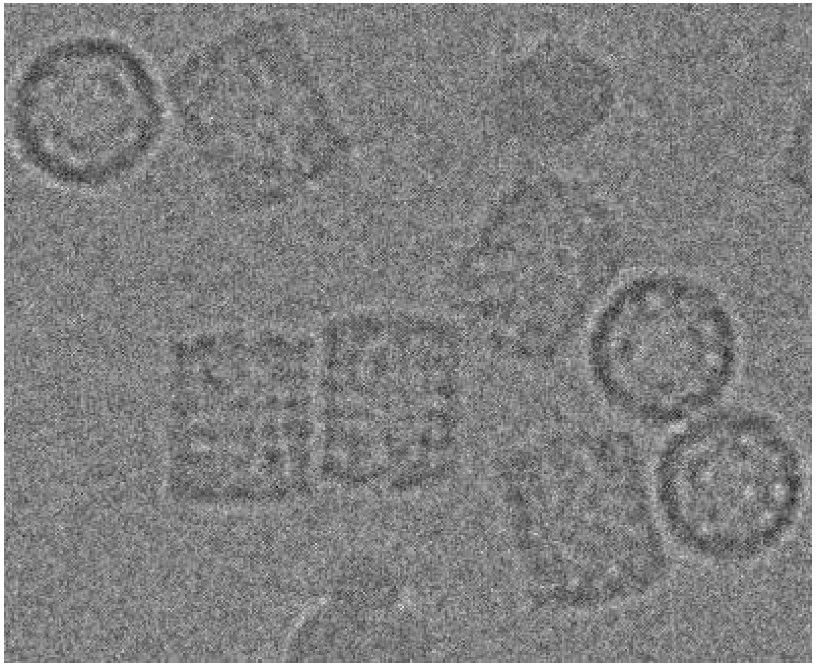


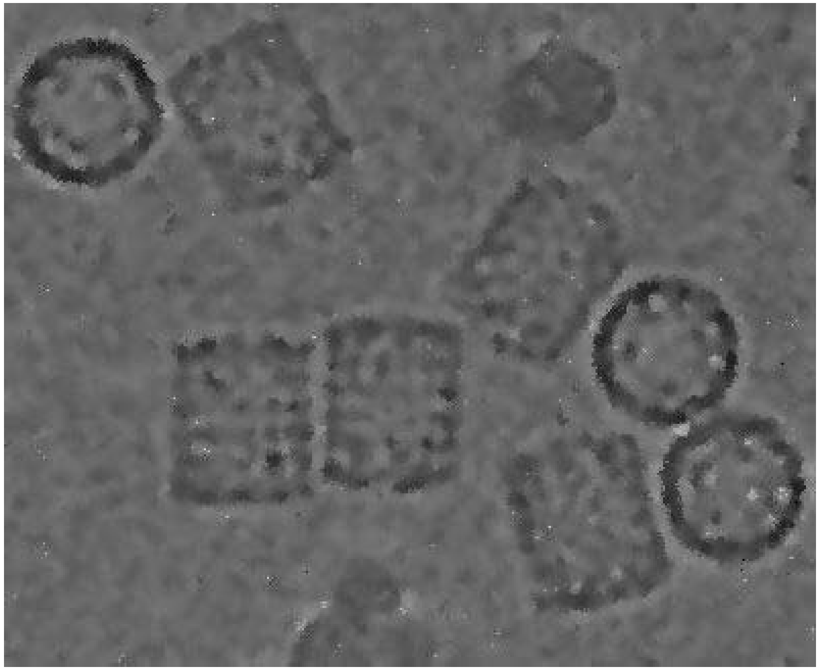
Bilateral Filtering

- Weighting Function

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

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



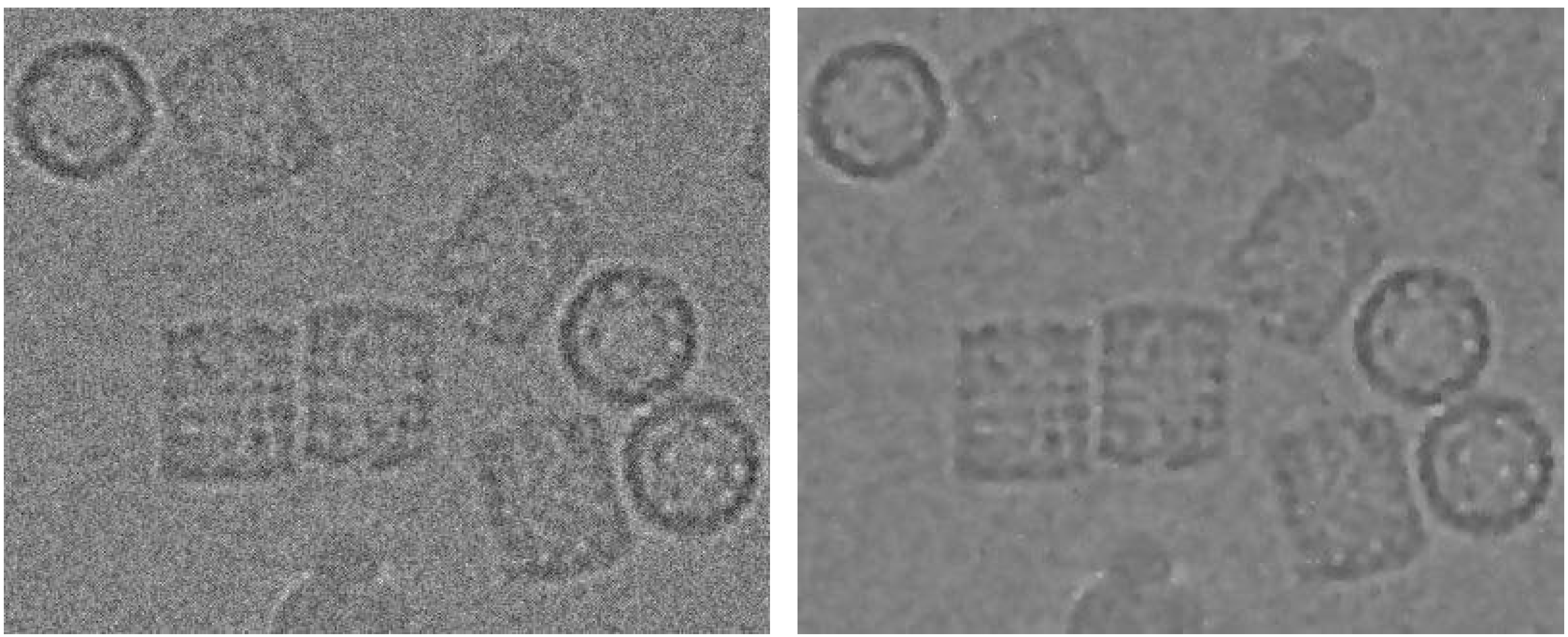


Non-Linear Filtering (using PDEs)

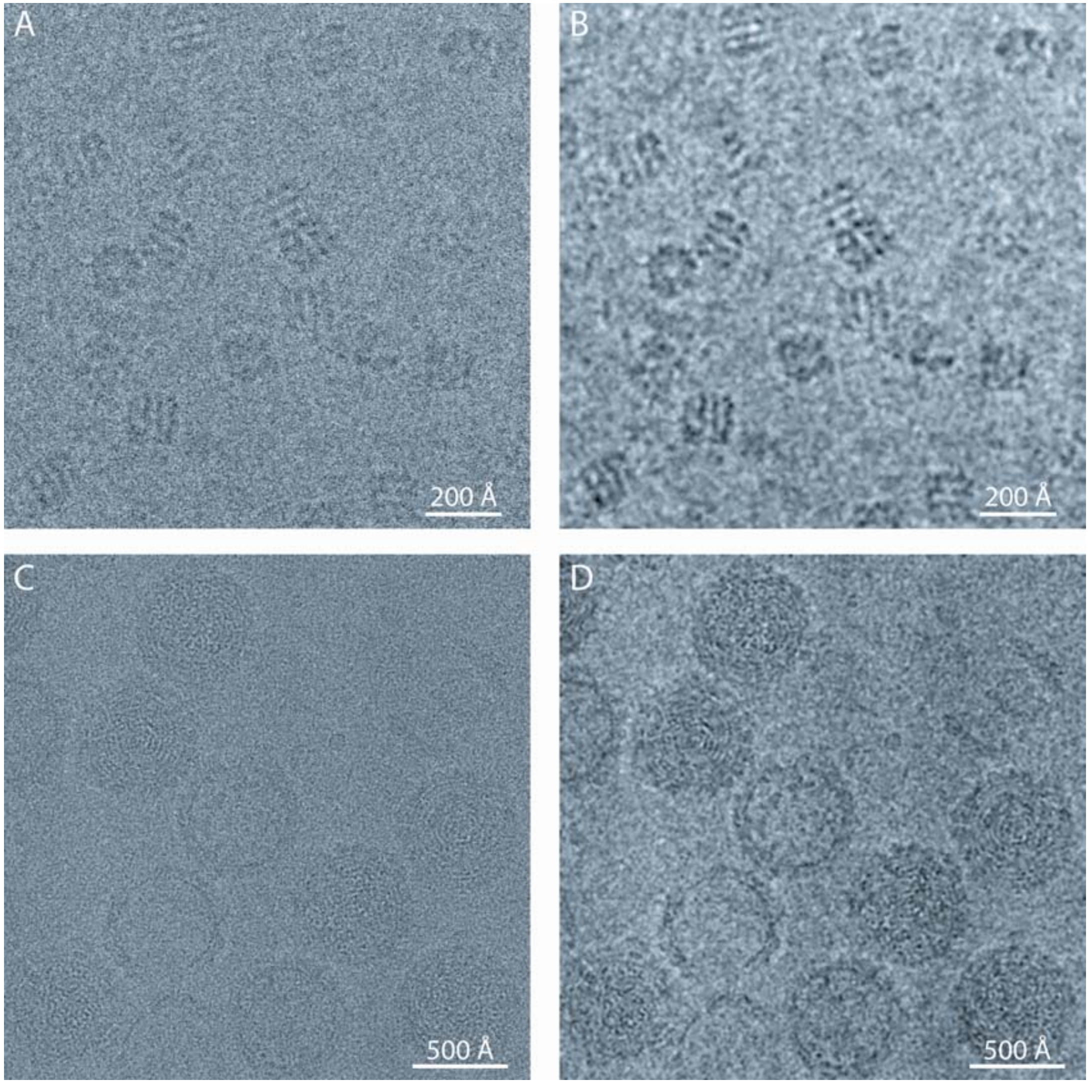
- Diffusion Equation ==Weighted Gaussian

$$\partial_t \phi - \mathrm{div}(g(|\nabla \phi|)\nabla \phi) = 0$$

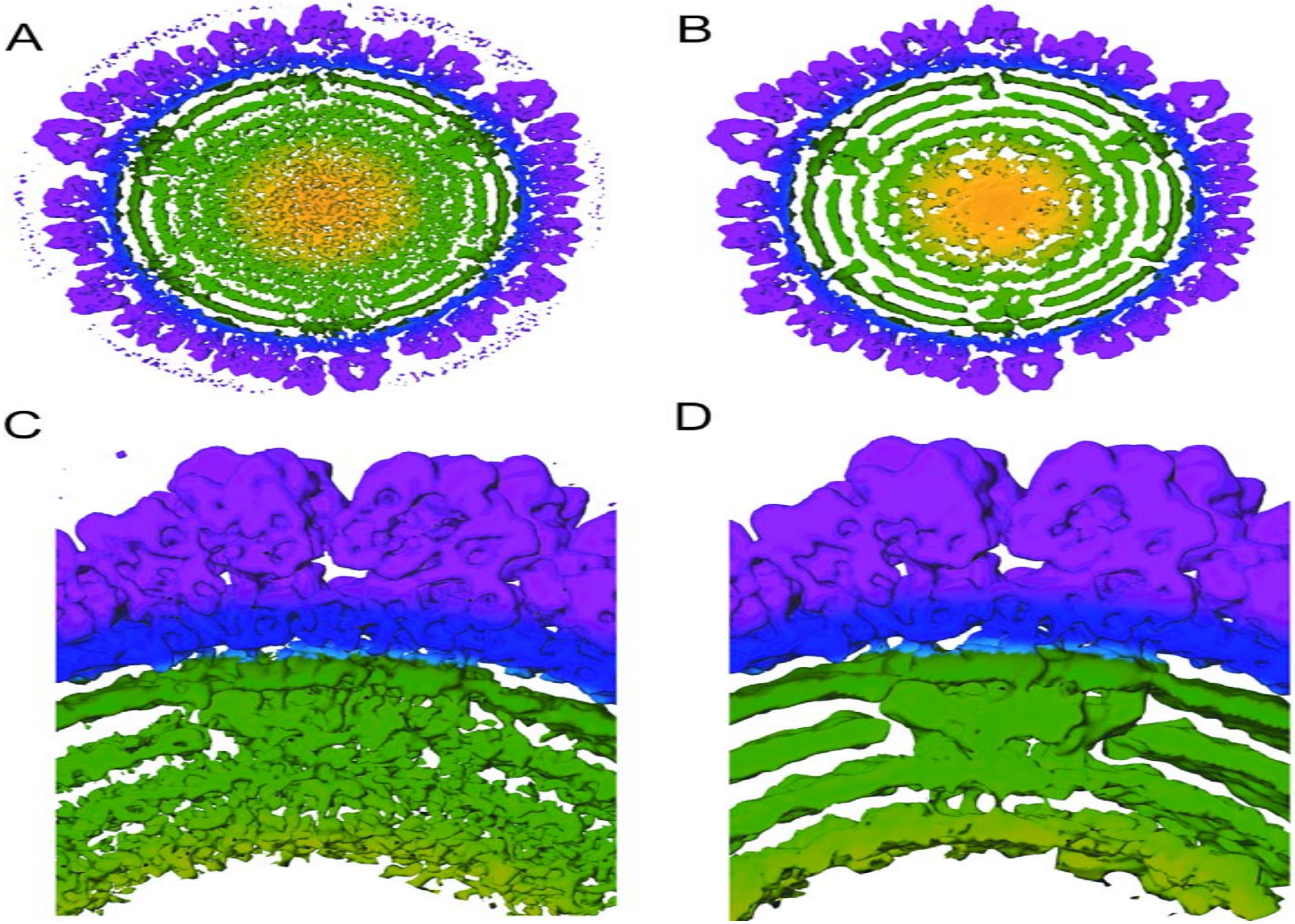
where $g(.)$ is a decreasing scalar function, e.g., $g(|\nabla \phi|) = \dfrac{1}{1 + |\nabla \phi|^2 / \lambda^2}$



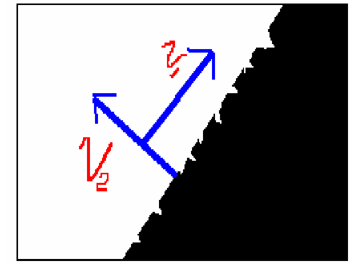
Bilateral Filtering
(Wen Jiang et,al., JSB, 2003)



Bilateral Filtering on RDV Map



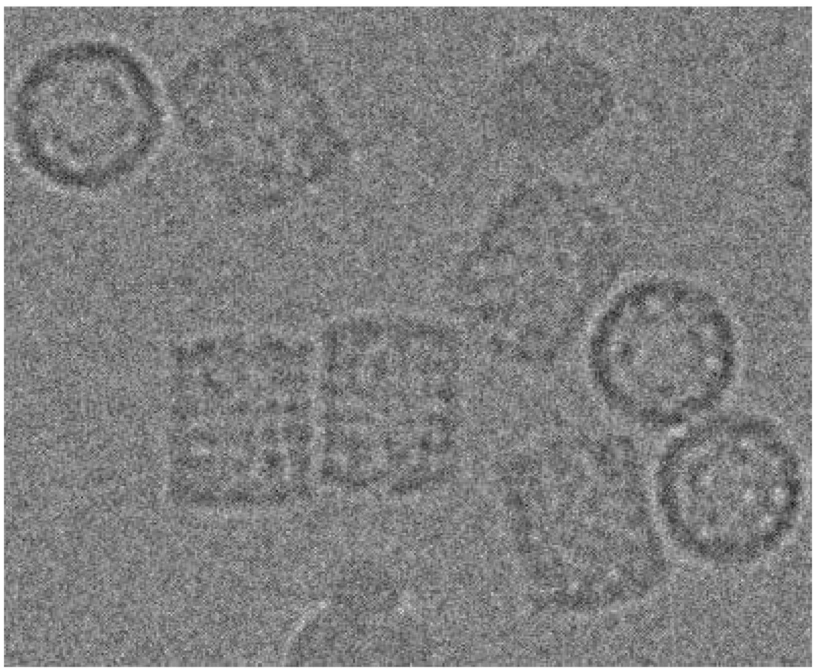
Anisotropic Diffusion (AD) Filtering

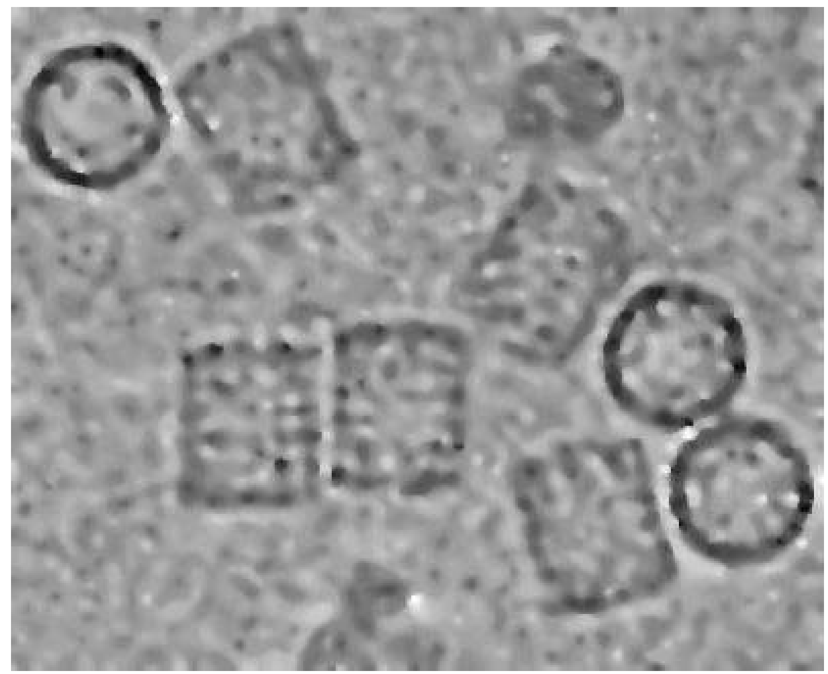


- Diffusion Equation

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

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





Finite Element Method for Anistropic Diffusion (Bajaj, Xu 2002,TOG)

Model

$$\partial_t x(t) - div(a(x)\nabla_{M(t)} x(t)) = 0$$

$a(x)$ is symmetric, positive definite matrix

Variational form

$$\left(\partial_t x(t), \theta\right)_{M(t)} + \left(\nabla_{M(t)} x(t), \nabla_{M(t)} \theta\right)_{TM(t)} = 0,$$

$\forall \theta \in C^{\infty}(M(t))$ where

$$(f, g)_M = \int_M fg dx, \qquad (\phi, \psi)_{TM} = \int_M \phi^T \psi dx$$

- How to represent $M(t)$?
- How to choose θ ?

Solution of the linear system

$$(M^n + \tau L^n)C((n+1)\tau) = M^n C(n\tau)$$

$$M^n = \left((\phi_i, \phi_j)_{M(n\tau)}\right)_{i,j=1}^m \qquad C(t) = [c_1(t), \cdots, c_m(t)]$$

$$L^n = \left((\nabla_{M(n\tau)}\phi_i, \nabla_{M(n\tau)}\phi_j)_{TM(n\tau)}\right)_{i,j=1}^m$$

- M^n and L^n are sparse.
- M^n is symmetric and positive definite.
- L^n is symmetric and nonnegative definite.
- $M^n + \tau L^n$ is symmetric and positive definite.

The system is solved by a conjugate gradient method.

Choice of Anisotropic Diffusion Tensor

Let $v^{(1)}(x), v^{(2)}(x),$ be the principal curvature directions of $M(t)$

at point $x(t)$ Let $N(x)$ be the normal at that point.

Then any vector $z = \alpha v^{(1)}(x) + \beta v^{(2)}(x) + \delta N(x)$

And define a, such that

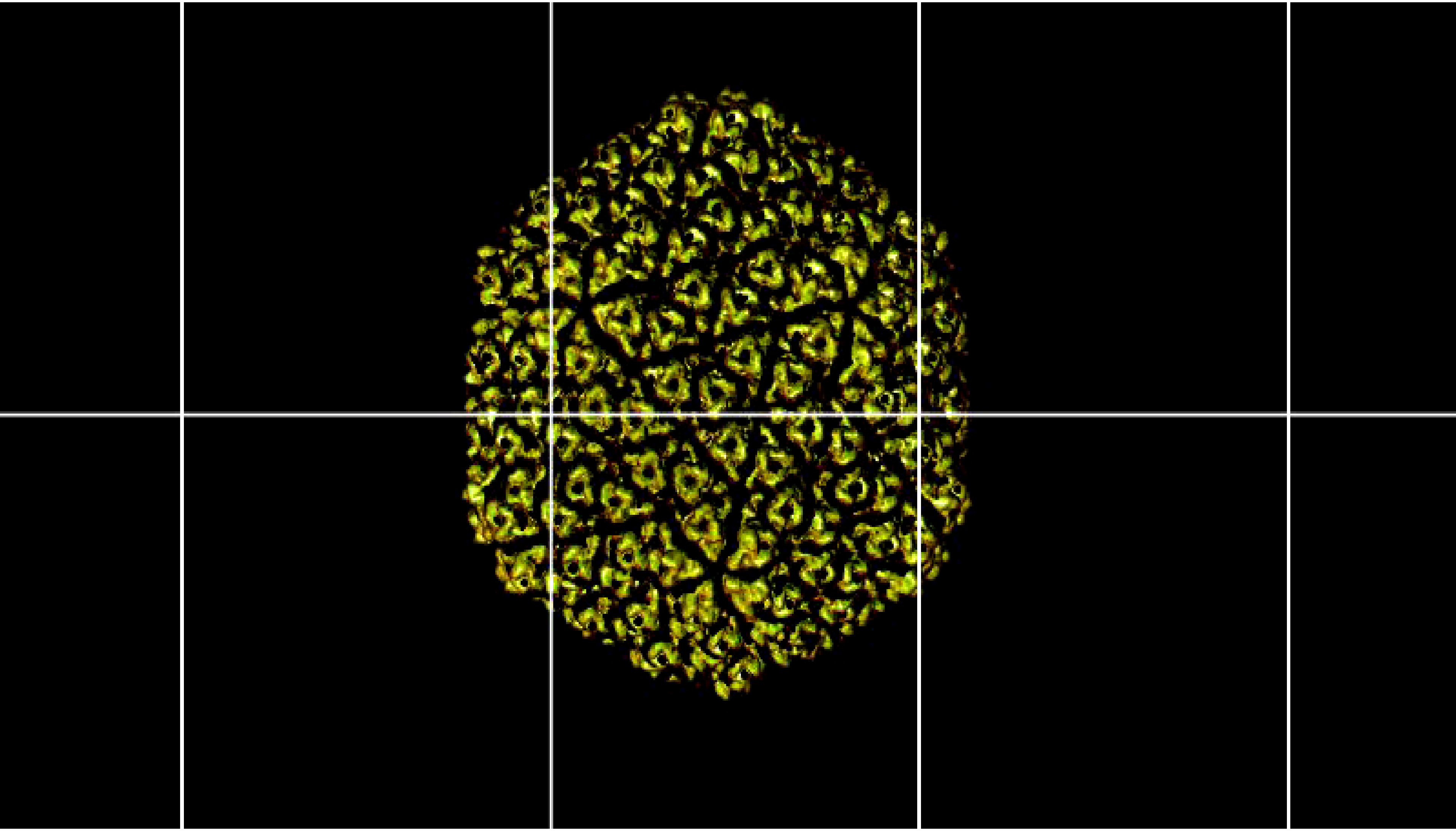
$$az = g(k_1)\alpha v^{(1)}(x) + g(k_2)\beta v^{(2)}(x) + \delta N(x)$$

where

$$g(s) = \begin{cases} 1, & s \le \lambda \\ 2(1+\frac{s^2}{\lambda^2})^{-1}, & s > \lambda \end{cases}$$

$\lambda > 0$ is a given constant.

Rice Dwarf Virus



Anisotropic Gradient Vector Diffusion

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

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

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)

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

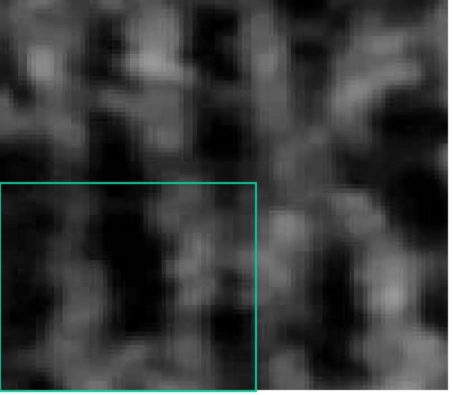
Where

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

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

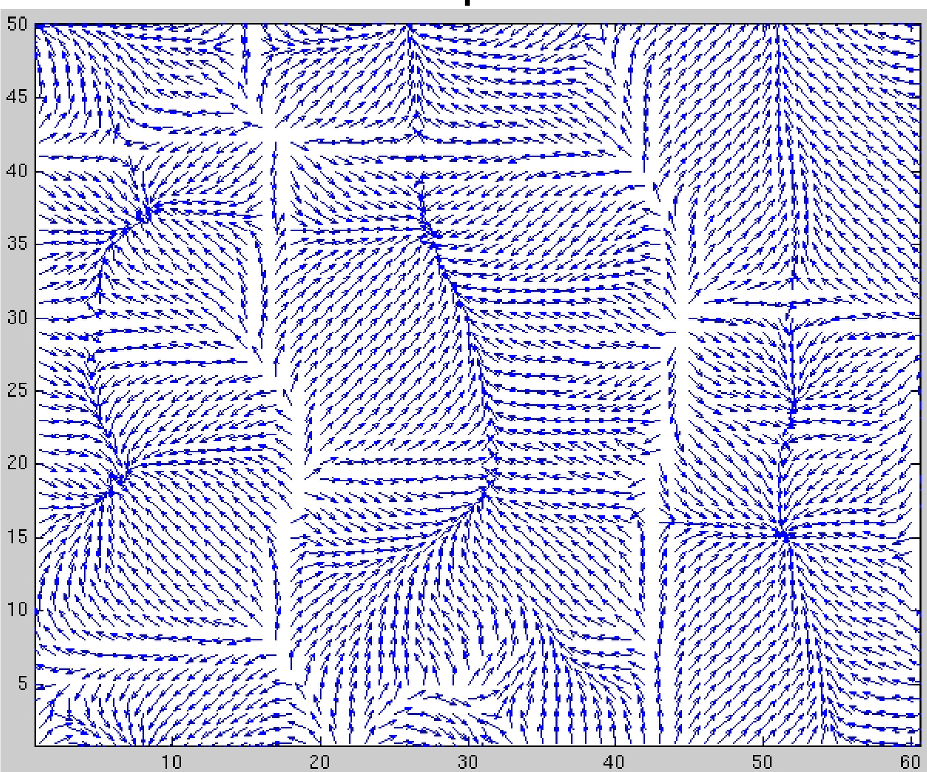
f is the original image to be diffused;

g(.) is the angle between two vectors

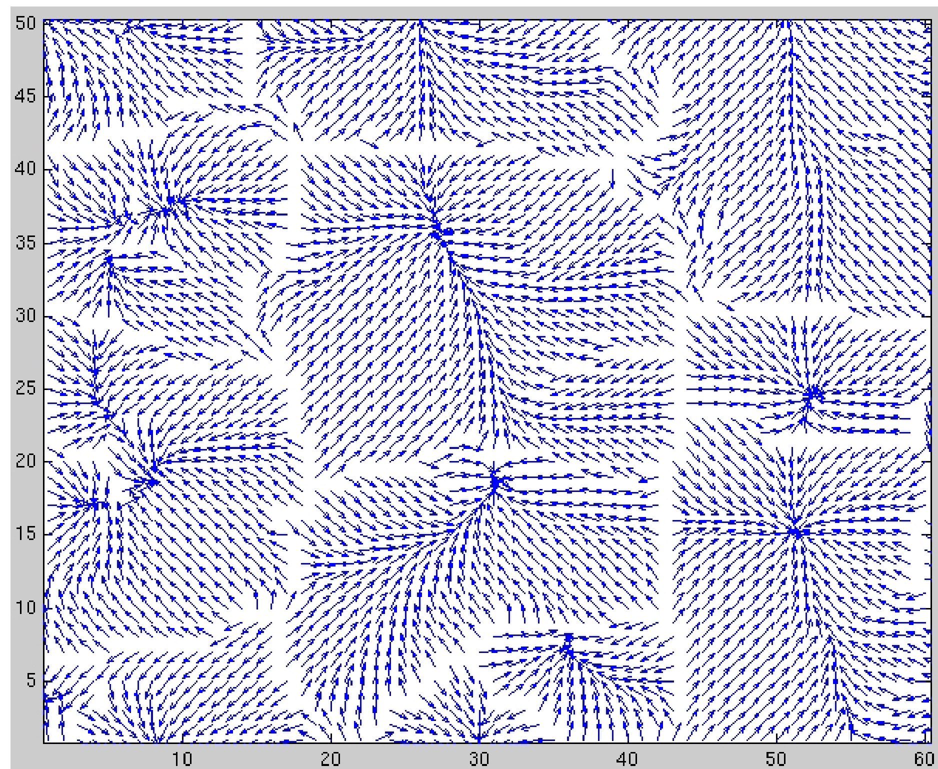


GVD v.s. AGVD

Isotropic diffusion



Anisotropic diffusion

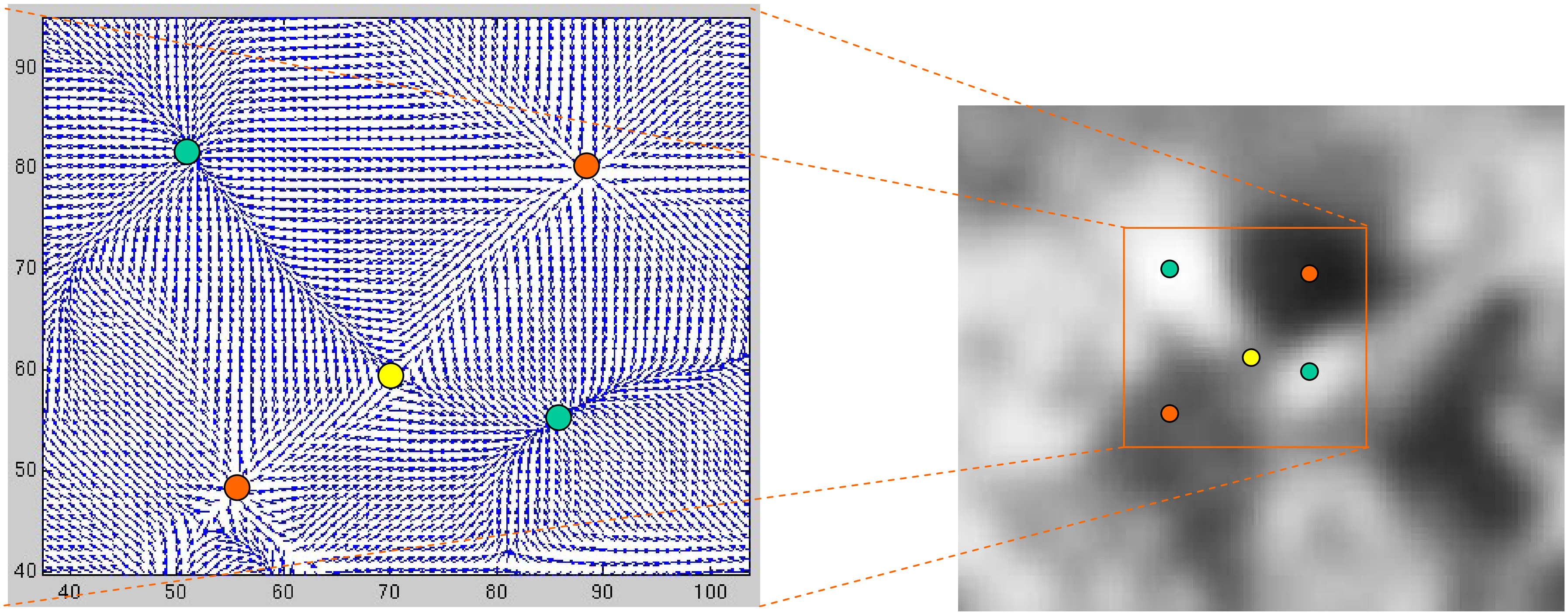


How AGVD Helps Image Segmentation ?

- Fast Marching Method
  - Initial seed points
  - Stopping criterion

- Use AGVD to locate seed points
  - Compute min/max critical points
    (discard saddle critical points)
  - All such critical points are used as seeds
  - Advantages: automatic, close to centers of homogenous regions, robust to noise due to vector diffusion.

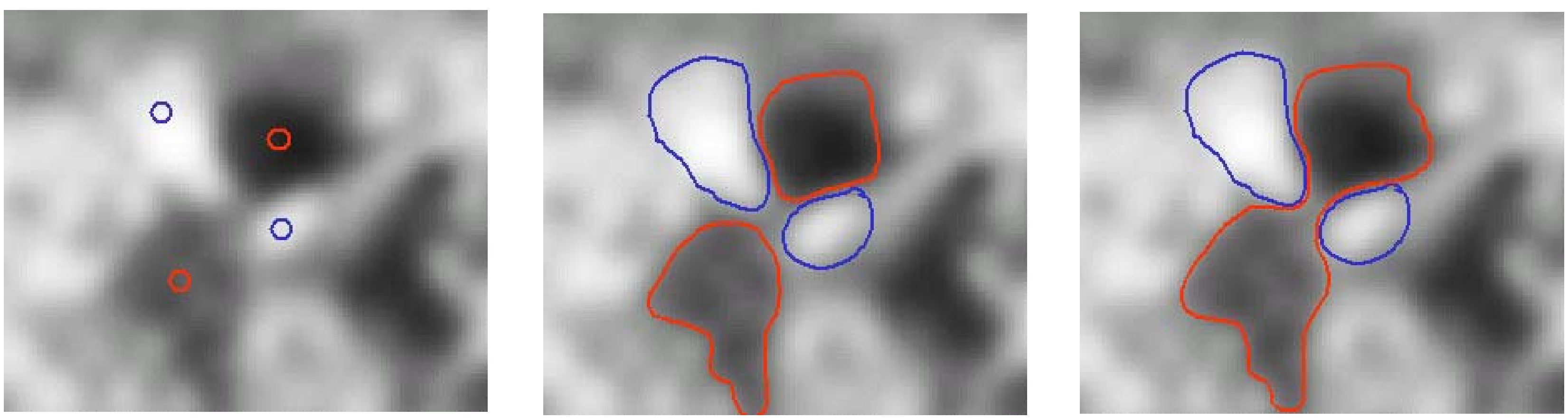
Compute Critical Points Using AGVD



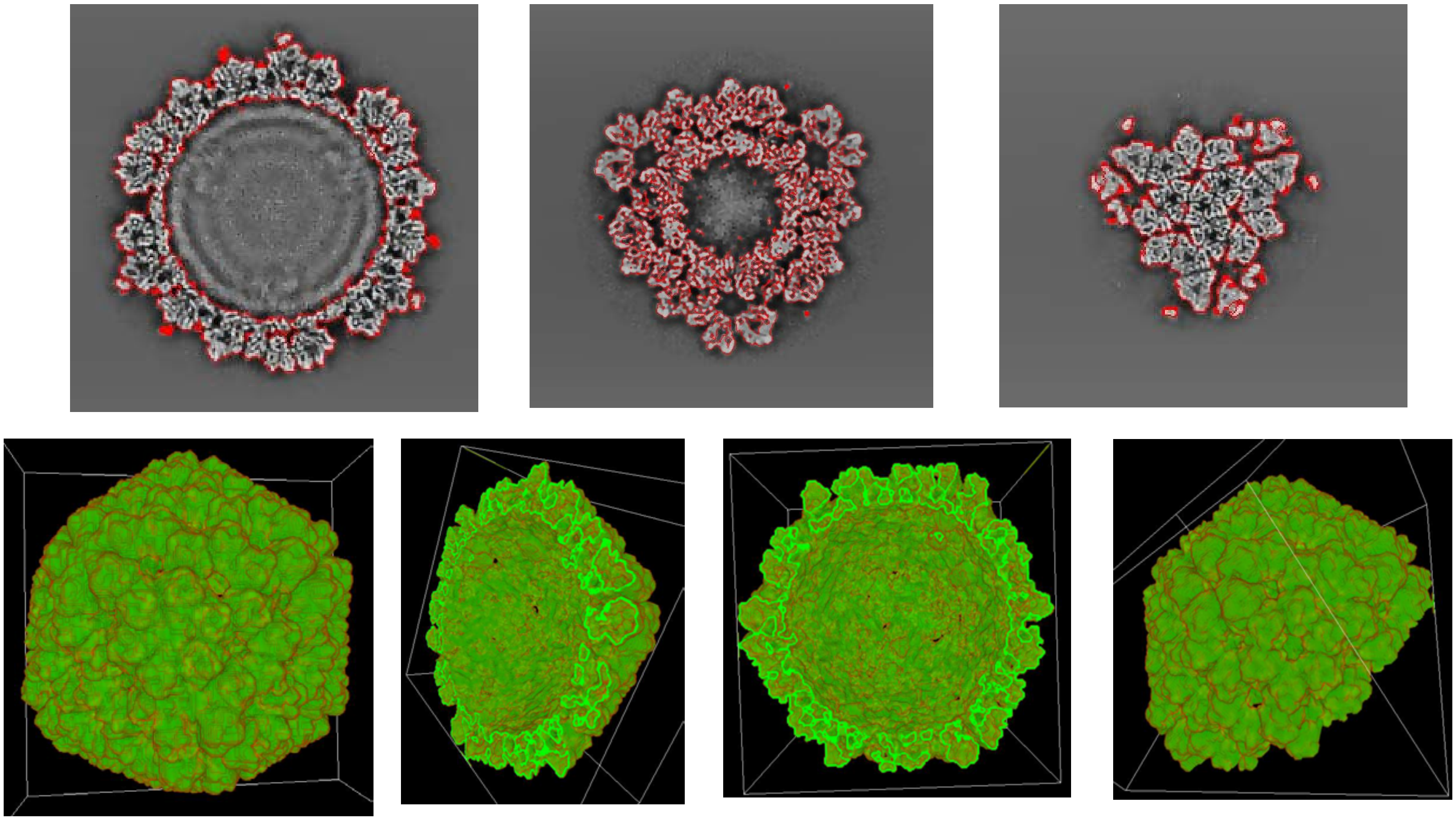
: minimum   : maximum   : saddle

Stopping Criteria Using Multiple-Contour

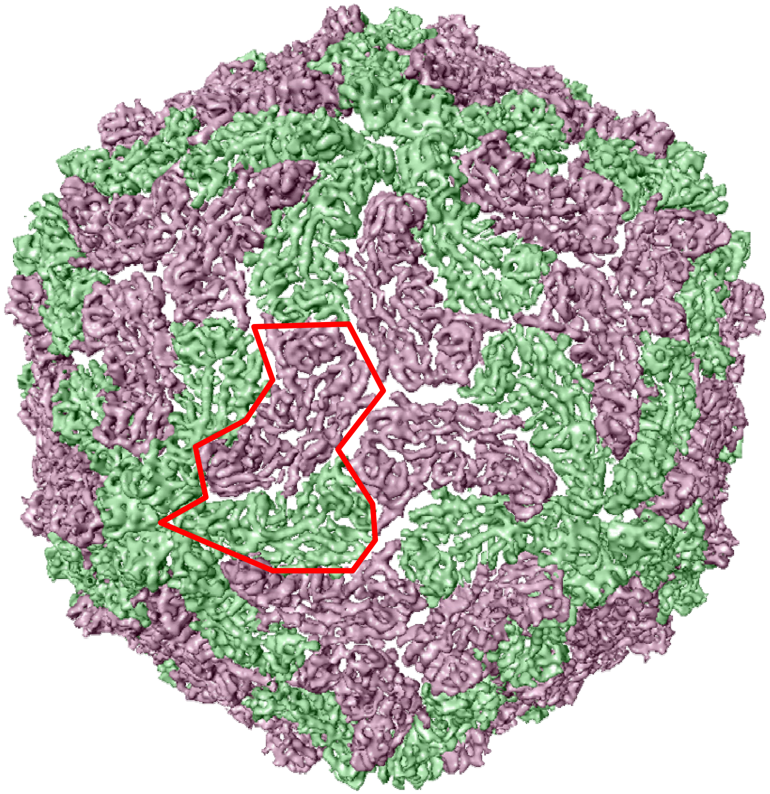
- Multiple-Contour
  - Group the critical points (for example, two groups as follows:
    max. critical points ⟹ feature & min. critical points ⟹ background)
  - Each seed initializes one contour, coupled with its group's I.D.
  - Contours march simultaneously. Contours with same I.D. are merged while contours with different I.D. stop on their common boundaries



Boundary Segmentation after Filtering

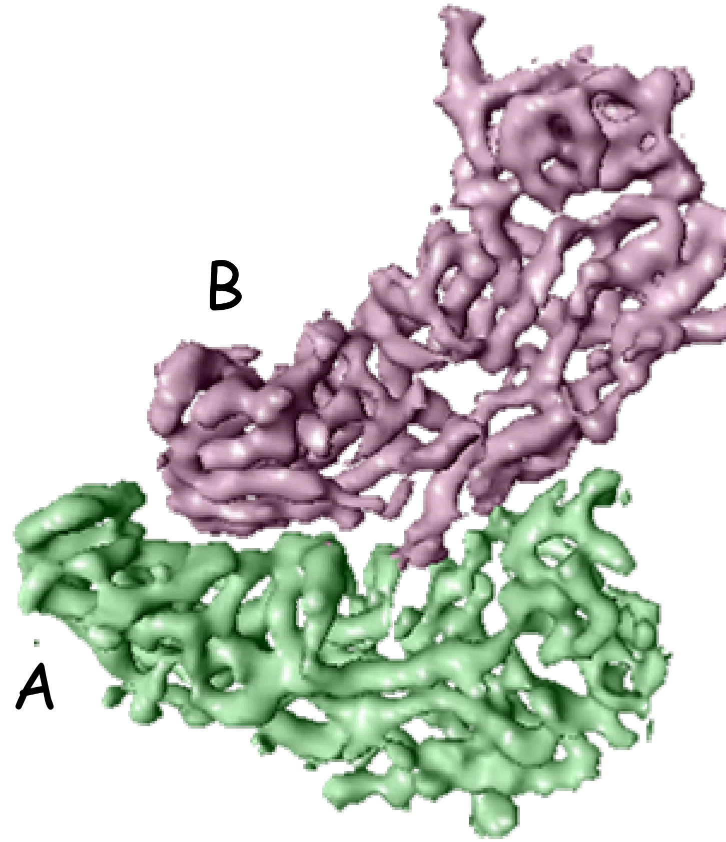


Boundary Segmentation of Inner Shell



Inner shell (T=1)

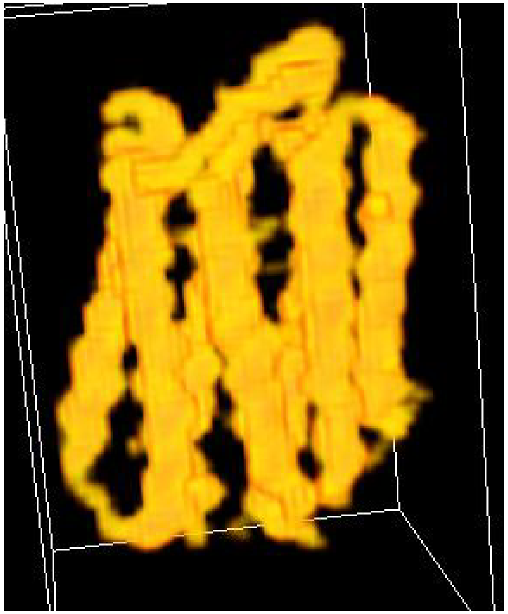
540 Å in diameter
P3 (114kDa) 29% of total protein
2 isoforms (A/B)



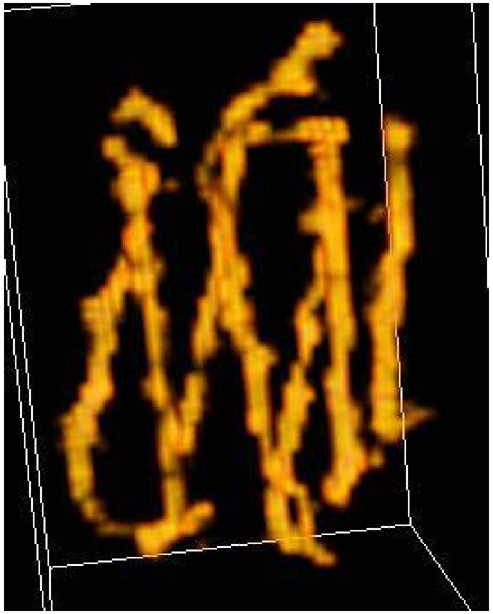
P3 dimer

Volumetric Skeletonization/ InPainting

- Pre-Processing for Docking Structures (Match & Fit)



Original Filtered Map

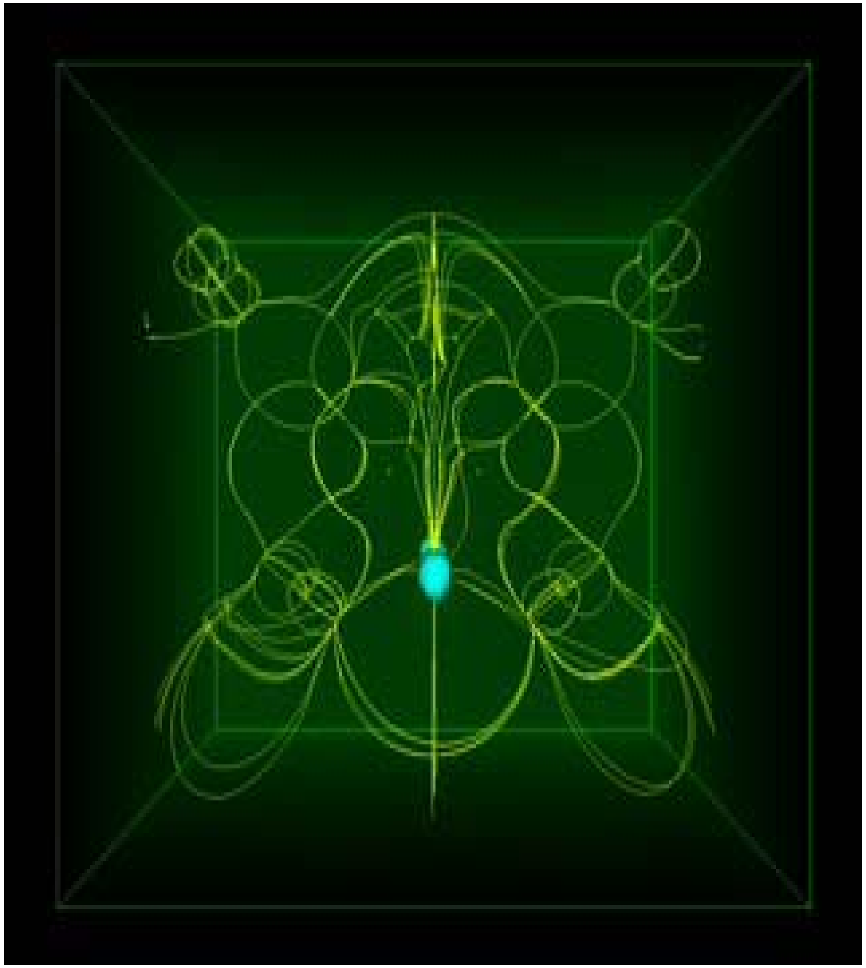


Skeleton Map

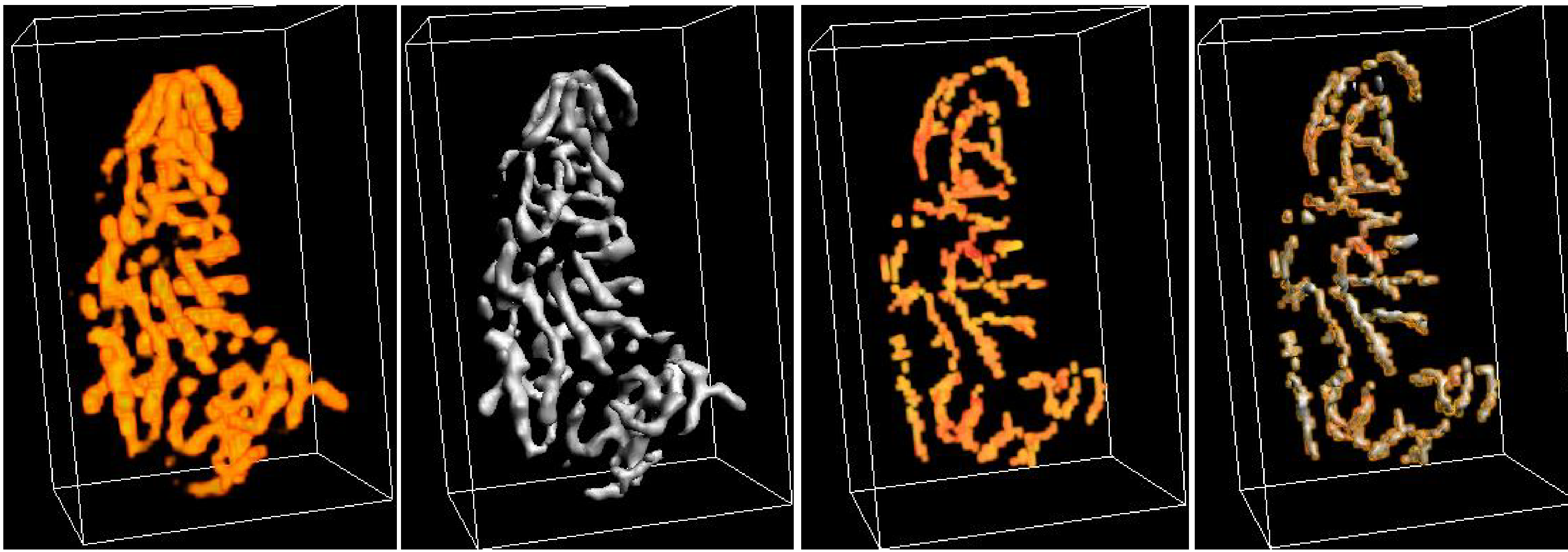
How GVD Helps Image Skeletonization ?

- Use GVD to locate critical points
  - Include minimum/maximum/saddle critical points
- Start from saddle points; trace integral lines along the diffused gradient vector field → Morse graph
- Prune the Morse graph for more meaningful skeletons
- Advantages:
  - Robust to noise due to vector diffusion.
  - Critical points are on the "skeletons" of features even for "flat" regions.

3D Morse Complex

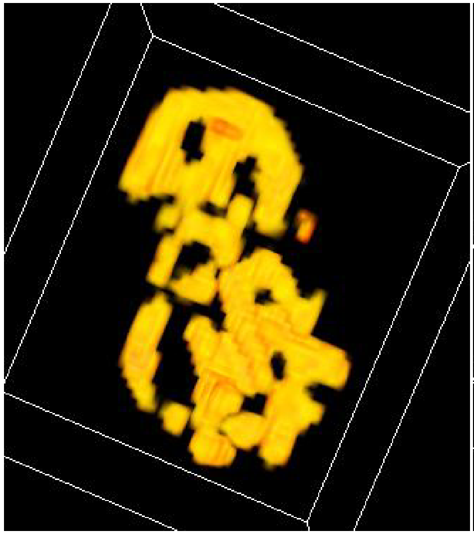


RDV: P3 monomer

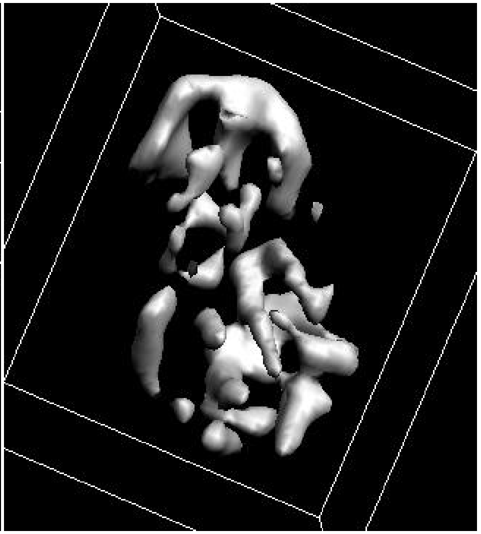


Volume-rendering | Isosurface | Skeleton | Skeleton with InPainting

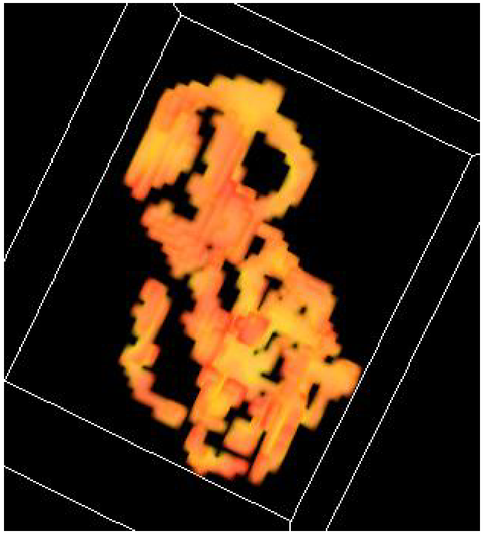
RDV: P8 monomer



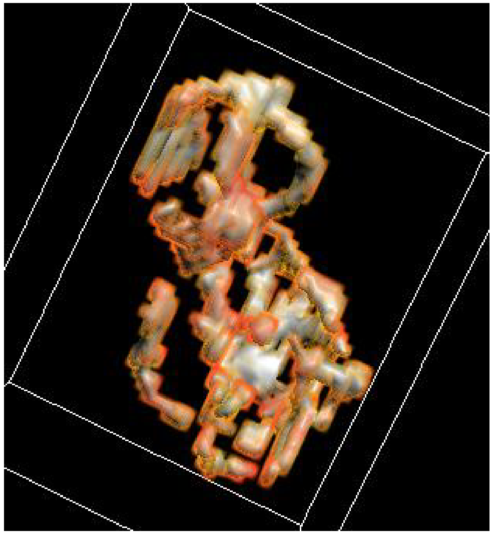
Volume-rendering



Isosurface

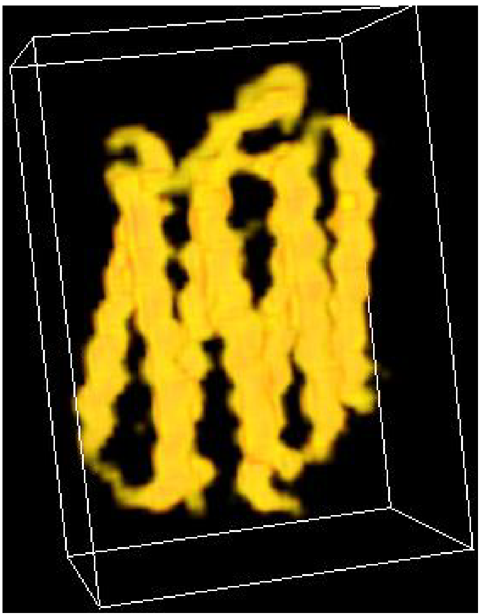


Skeleton

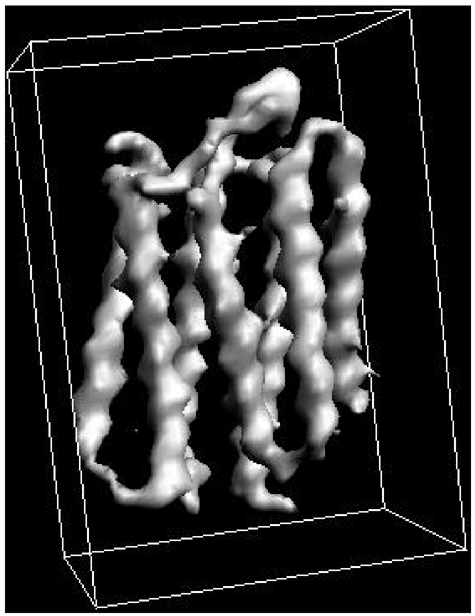


Skeleton with another isosurface

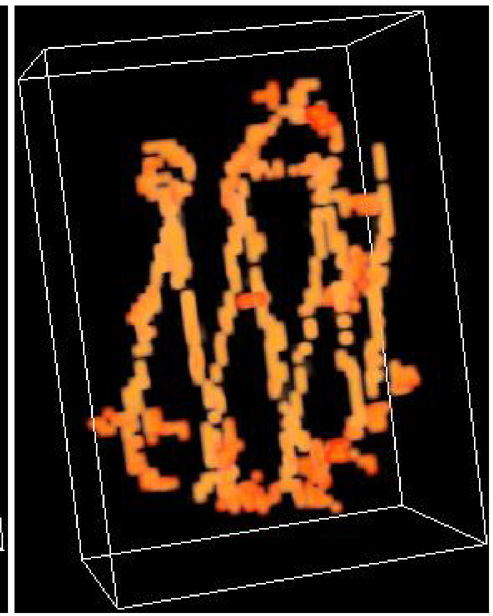
BacteriorRhodopsin/Lipid Complex (PDB: id=1c3w)



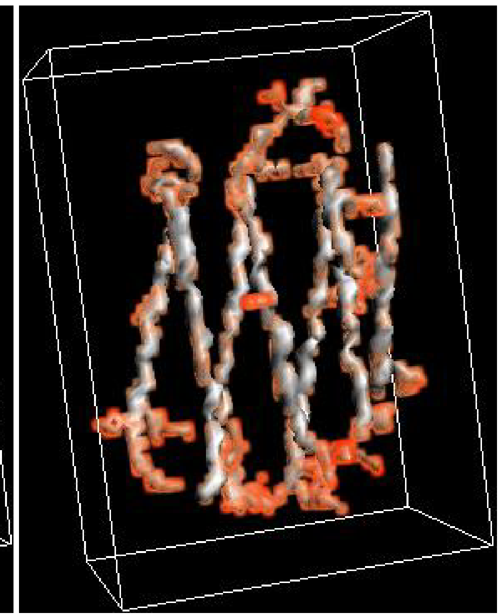
Volume-rendering



Isosurface

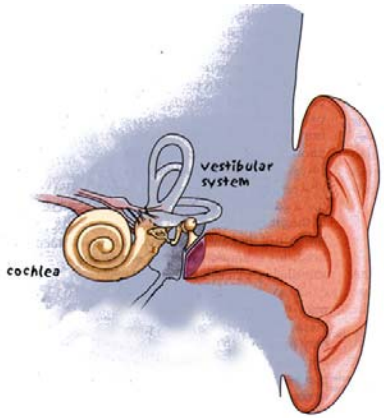


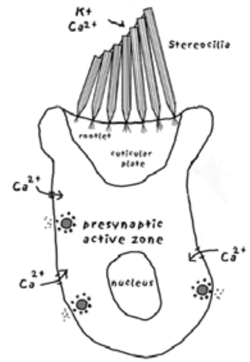
Skeleton



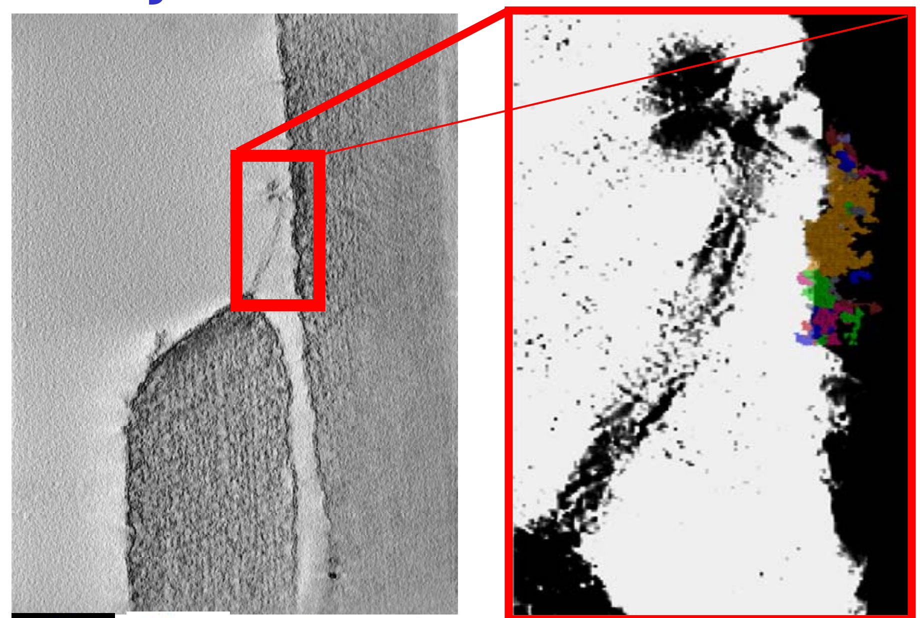
Skeleton with InPainting

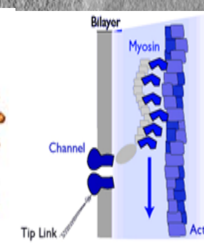
Tomographic Imaging to Structure to Analysis & Visualization of Hearing Machinery

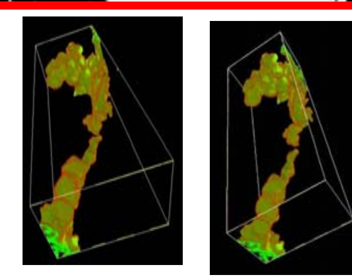




Tomographic Molecular Imaging →Anistropic Diffusion Filtering → Classification, Segmentation,Skeletonization of 3D Density Maps → Quantitative Structure Analysis →Visualization

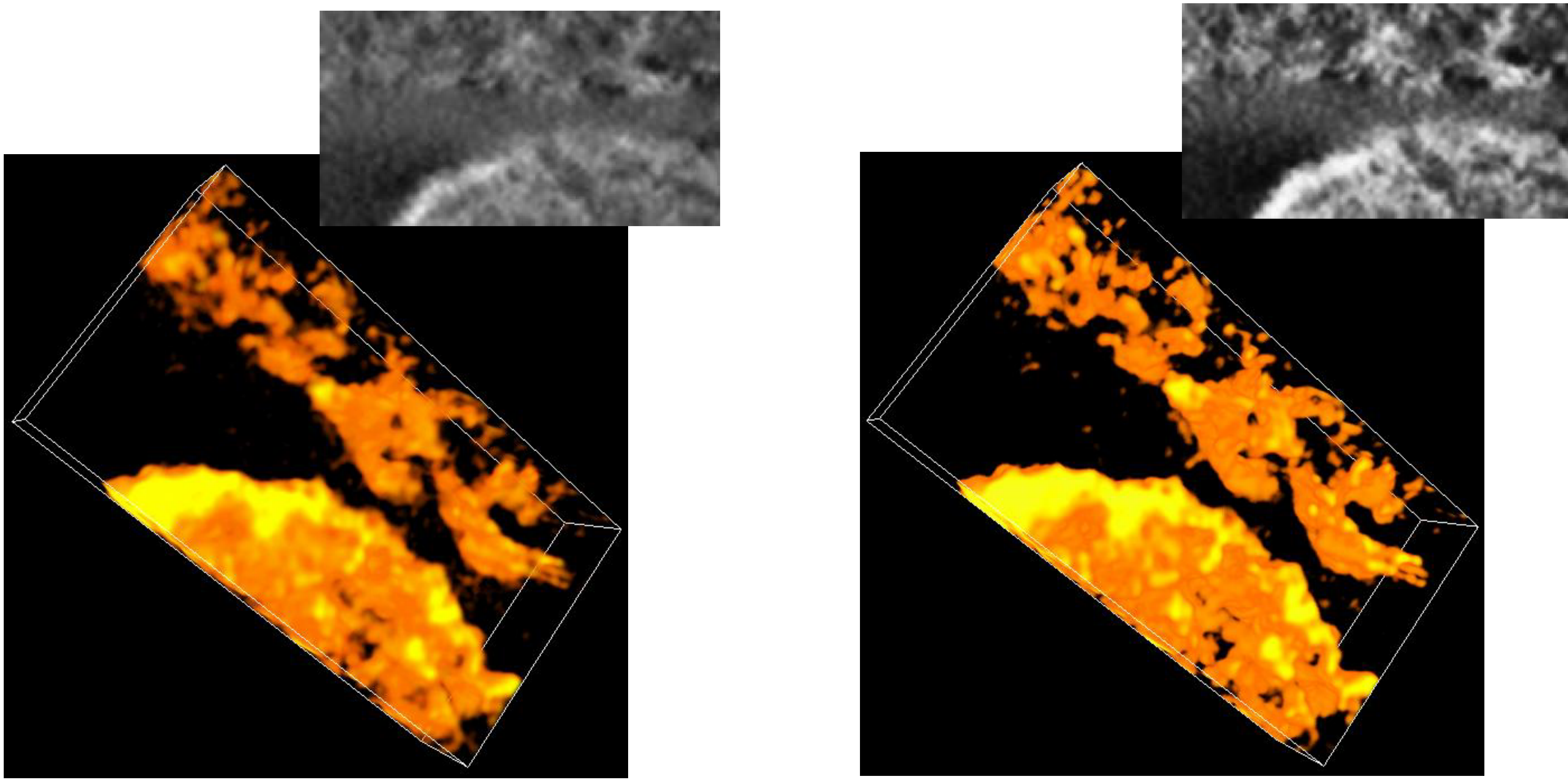






(Collaborator: Manfred Auer, Jim Hudspeth Rockefeller University and NYU Medical Sciences)

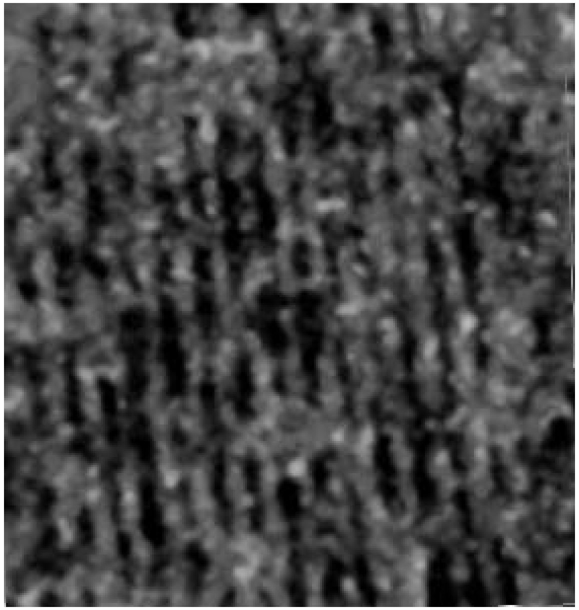
Image Contrast Enhancement (contd.)

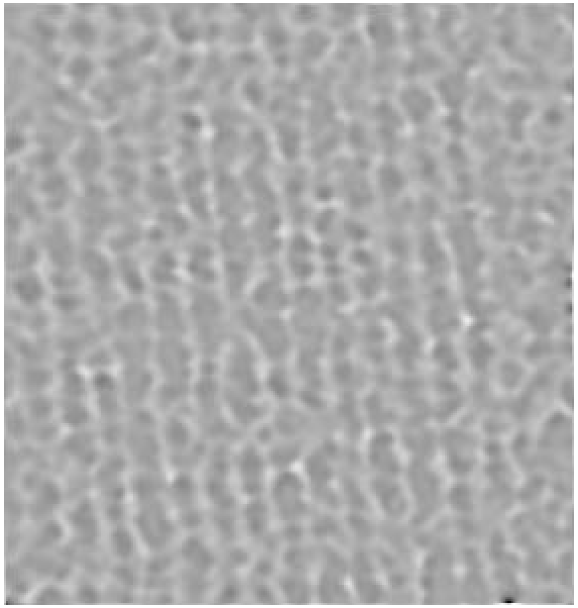


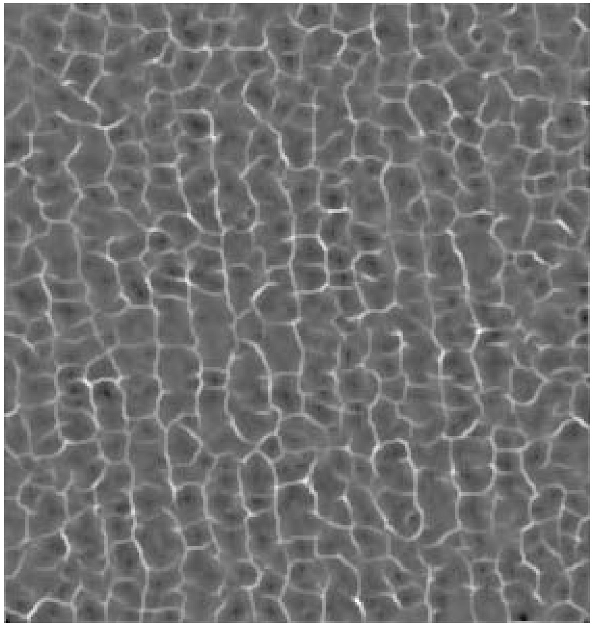
Tip structure of B280a (Left: original Right: enhanced)

Skeletons of ActinBundle (B280a)

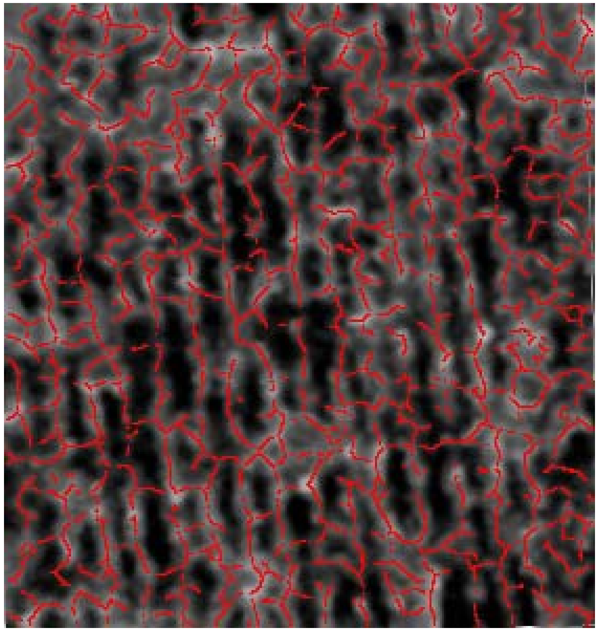
- 2D Electron tomogram


Original image

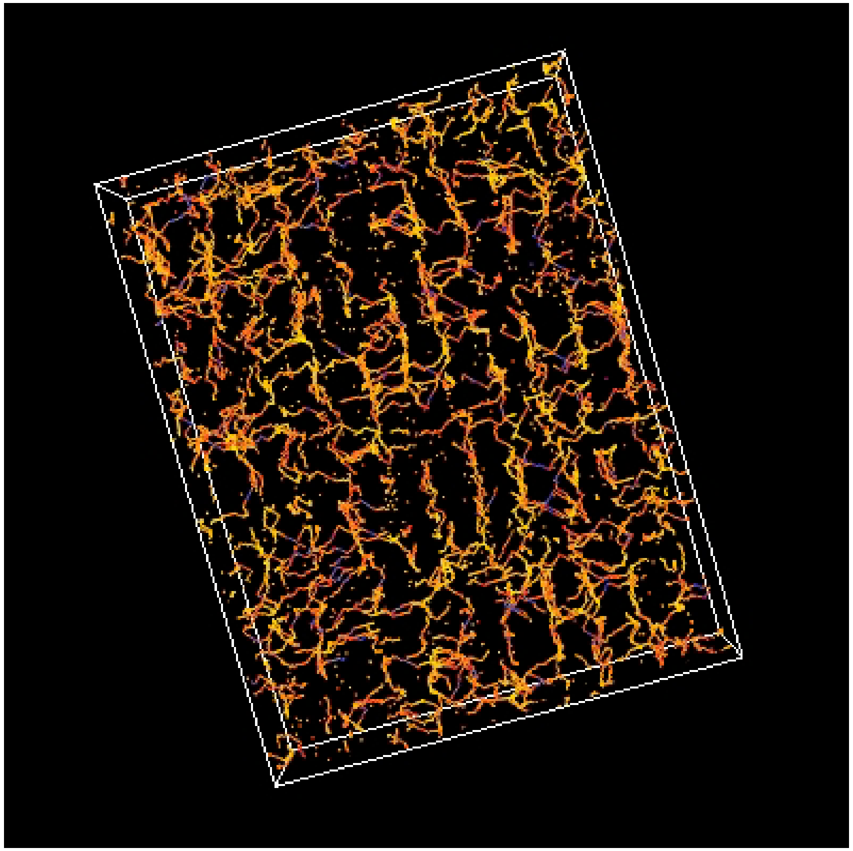

SMM (isotropic)

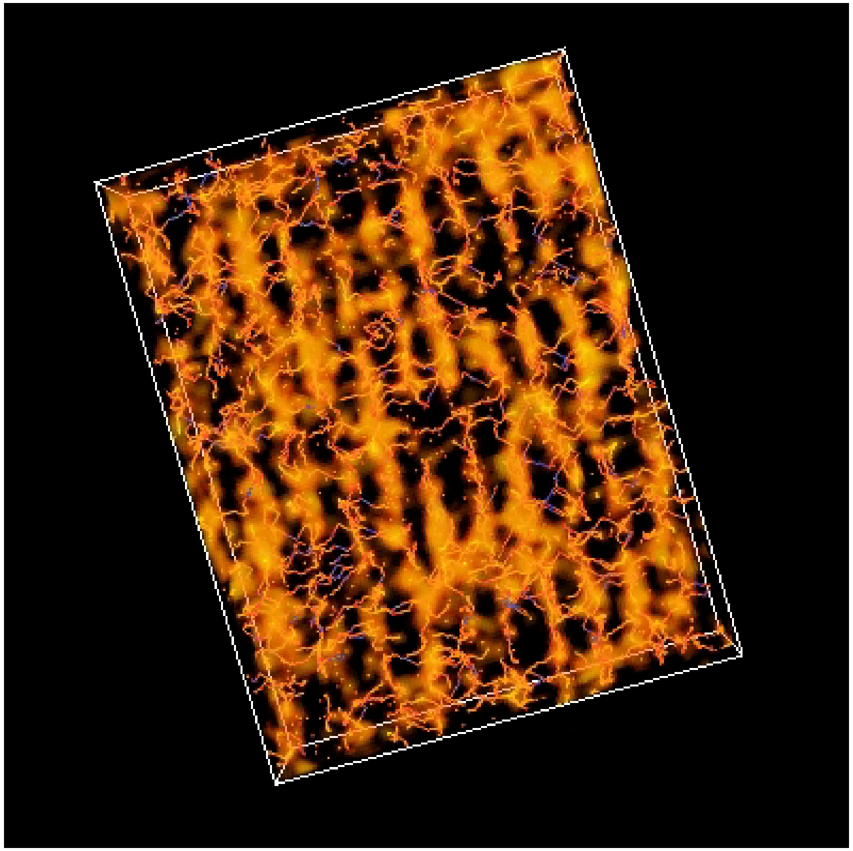

SMM (anisotropic)

Skeletons >>

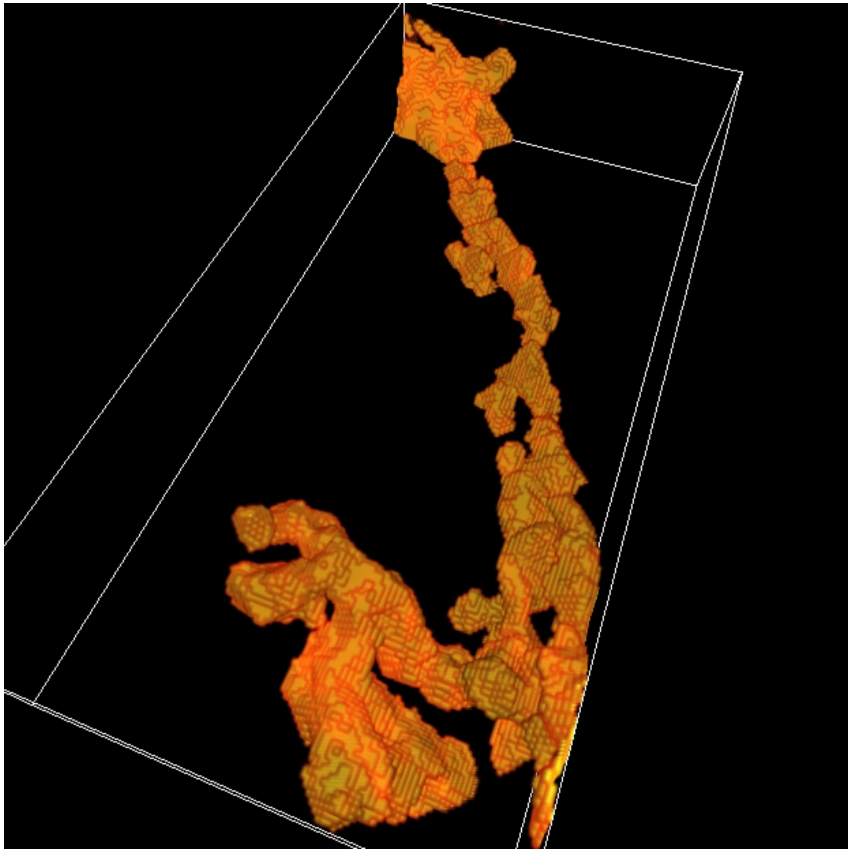


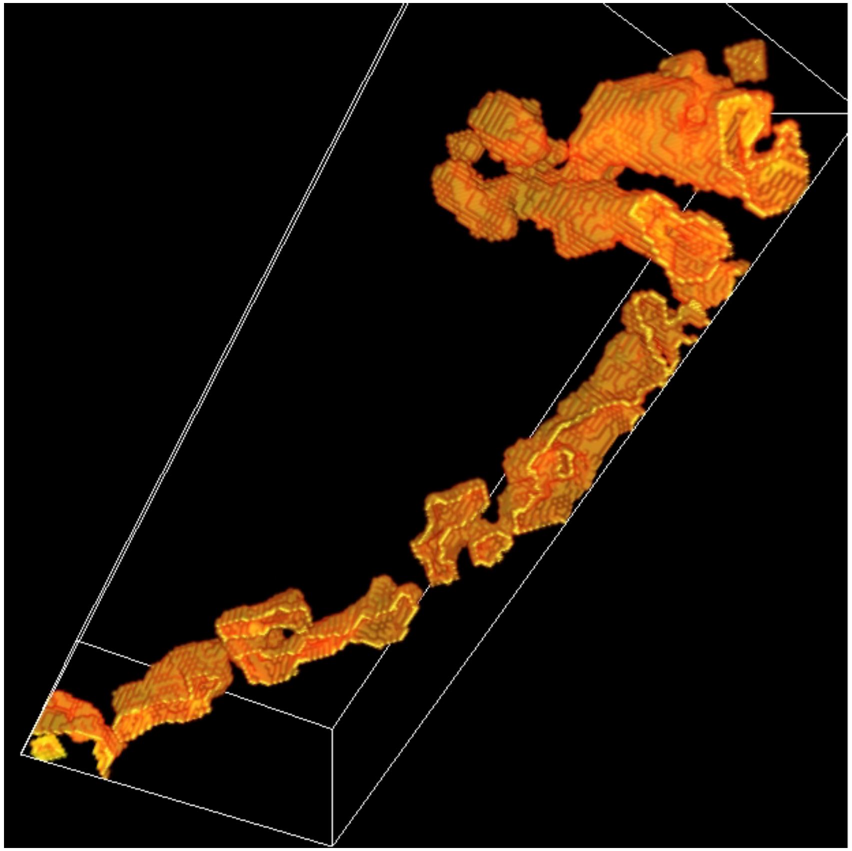
Skeletons of ActinBundle (B280a)



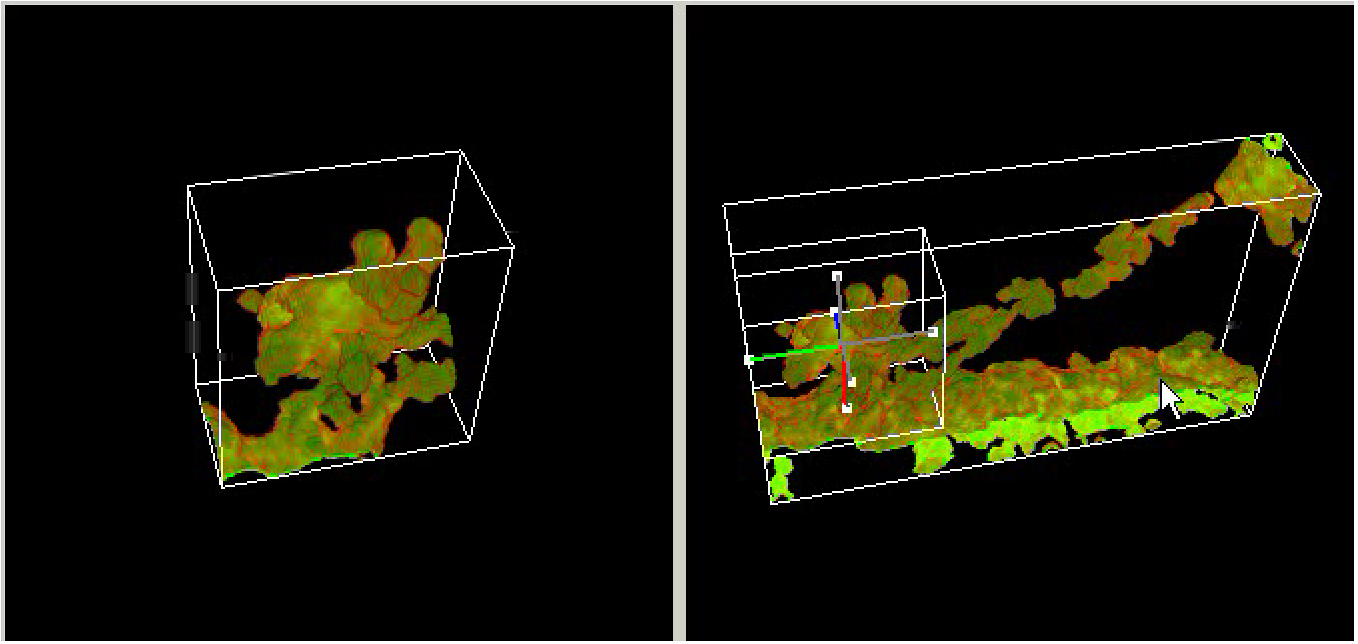


Segmentation of TipLink (B206a)





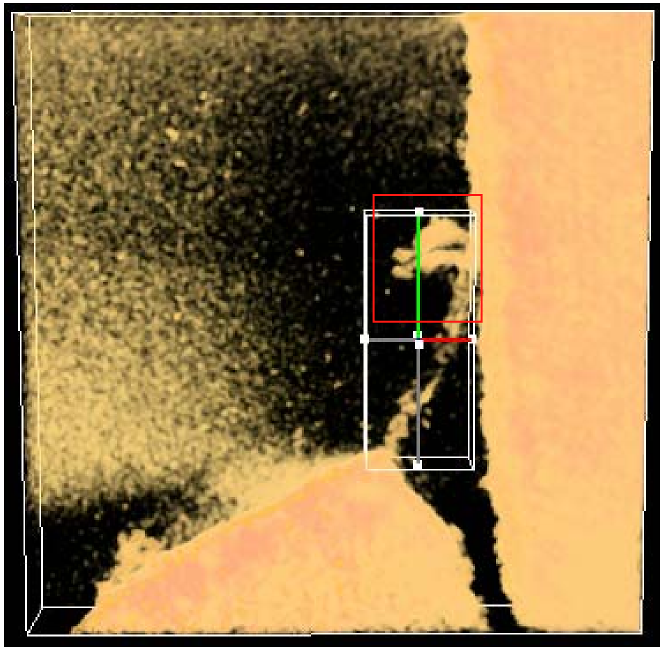
Segmented Tip Link (B206)

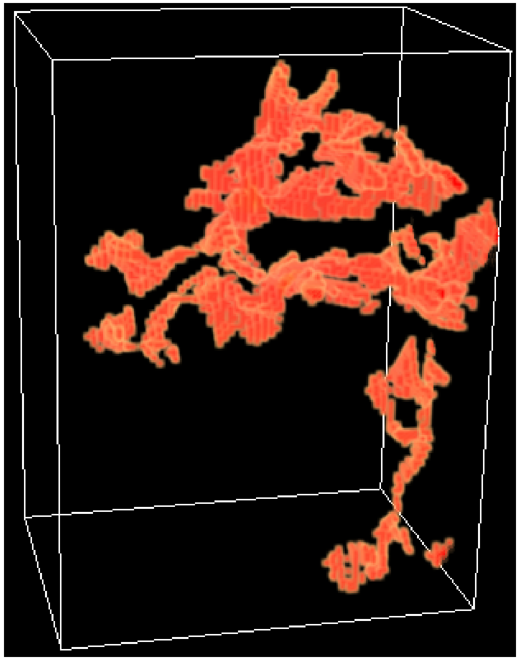


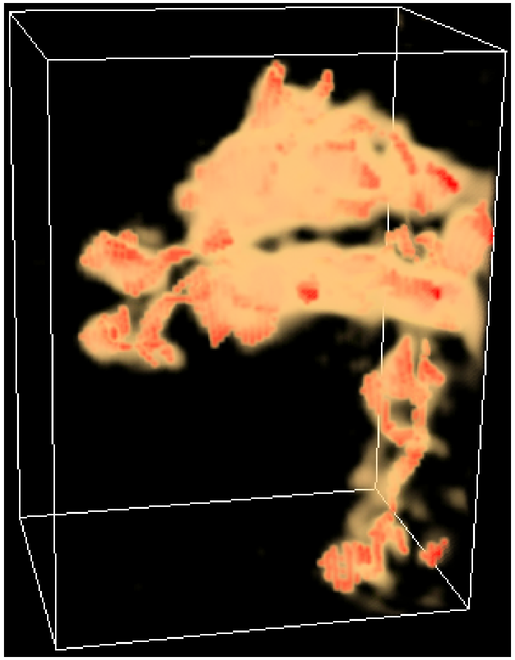
Bajaj, Zeyun, Auer, JSB, 2003 to appear.

Skletonization/InPainting

- 3D Electron Tomogram





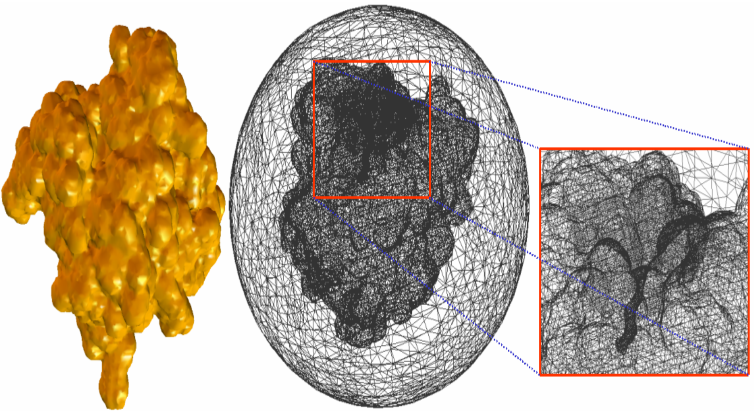
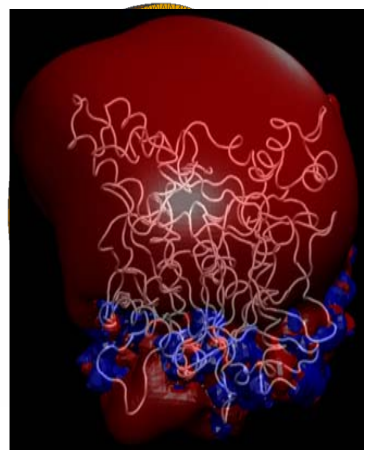
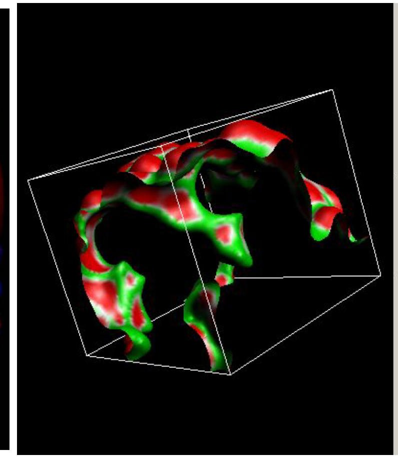
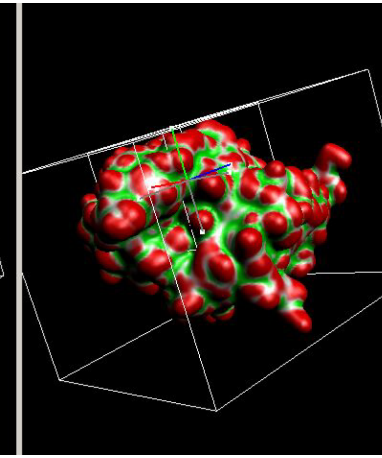


Overall volume

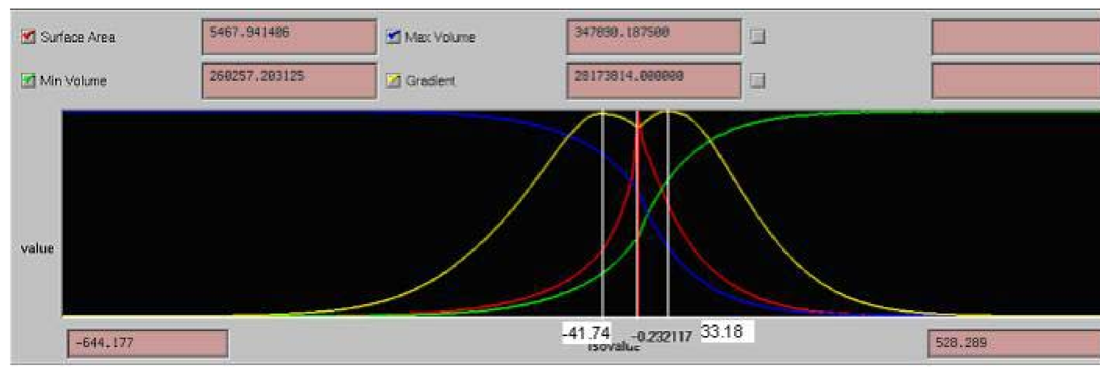
Skeletons

Skeletons with density map

Atomic Level Structure to Simulation to Analysis to Protein Function

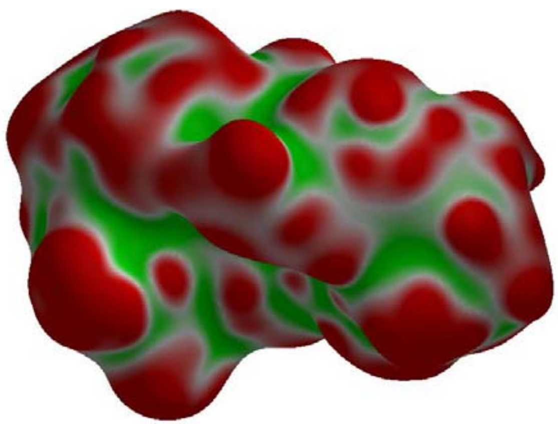
PDB → Finite Element Meshes with Properties → Poisson Boltzmann Calculations → Flexible Docking →Function Fingerprints

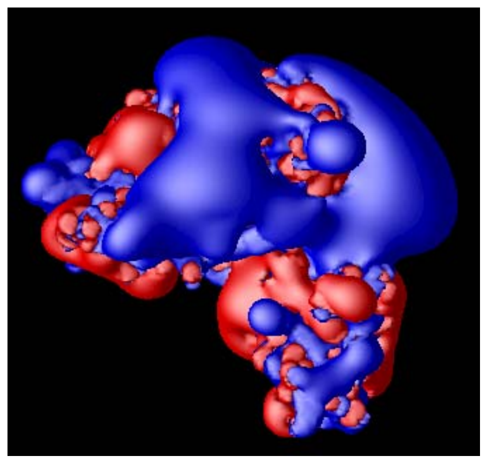


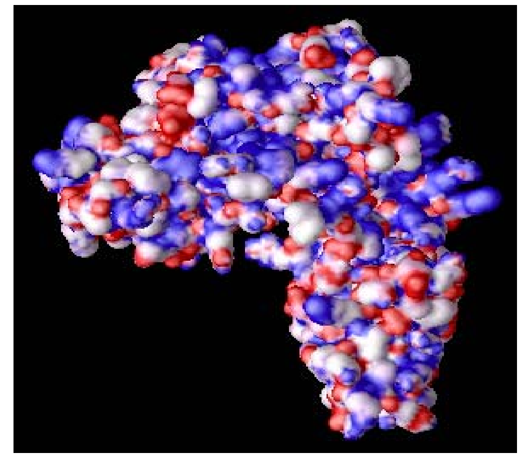
(Collaborators: N. Baker (Wash U), D. Goodsell(Scripps), A. McCammon (UCSD), A. Olson (Scripps), M. Sanner(Scripps))

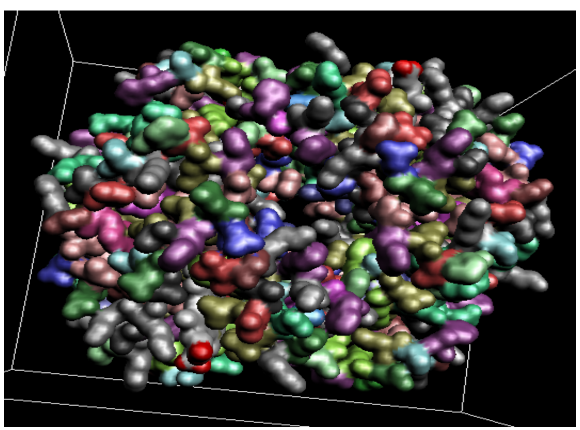
****Sponsored by NSF-NPACI-Interaction Environments (Bio-Alpha)**

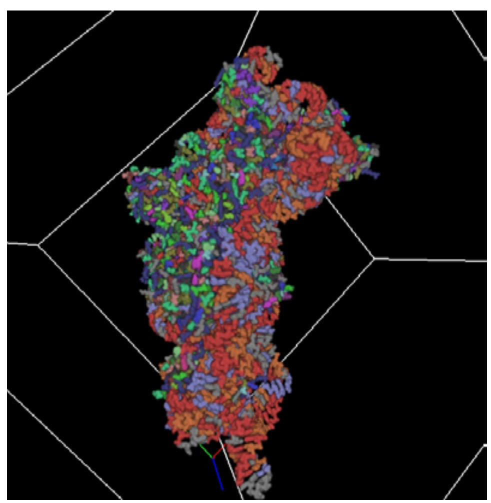
Compressed Volumetric Representations of Structures & Properties

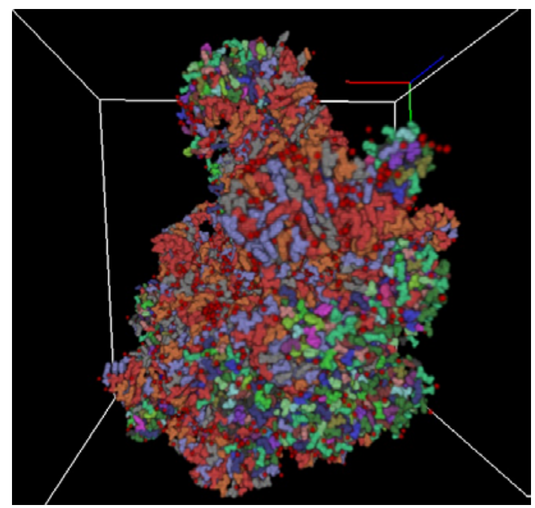












Volumetric Electron Density (Implicit Solvent Model)

The electron density in the unit volume at point r :

$$\rho(r; X) = N \int dv' \psi^*(x; X)\psi(x; X)$$

where x denotes the collection of electronic space and spin coordinates and X the collection of nuclear coordinates.

Common approximation = the summation of individual atomic electron charge distributions:

$$\rho_i(r) = exp(\frac{B_i r^2}{R_i^2} - B_i)$$

where $B_i < 0$ is a blobby parameter and R_i is the van der Waals radius of the atom

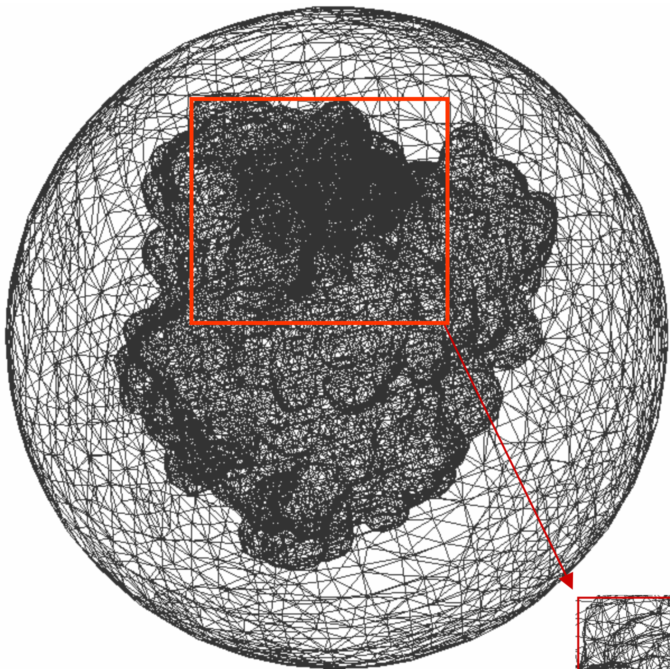
Volumetric Electrostatic Potential
(Baker, McCammon 2002: APBS)

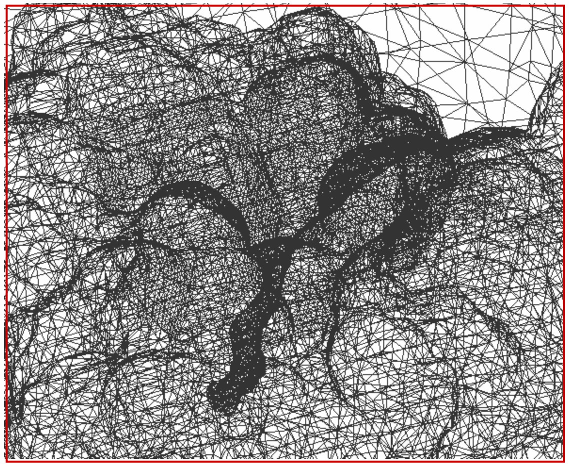
A common model for evaluating the molecules' electrostatic properties is the Poisson-Boltzmann equation.

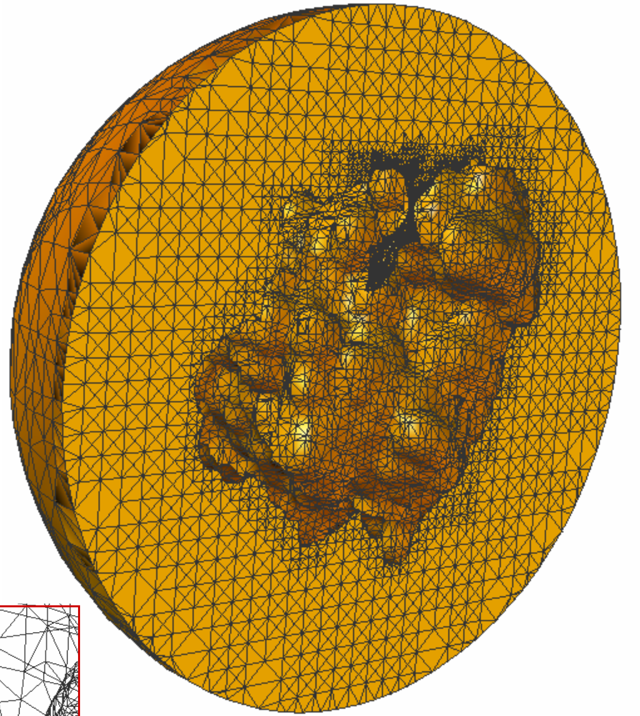
$$-\nabla \cdot [\varepsilon(\mathbf{r}\nabla V(\mathbf{r})] + k^2(\mathbf{r})\sinh(V(\mathbf{r})) = \rho(\mathbf{r})$$

where $\varepsilon(\mathbf{r})$ is the dielectric properties of the solute and solvent, k^2 is the ionic strength of the solution and the accessibility of ions to the solute, and $\rho(\mathbf{r})$ is the distribution of solute atomic partial charges.

Finite Element Models – AcetylCholinesterase (257^3, 66MB)





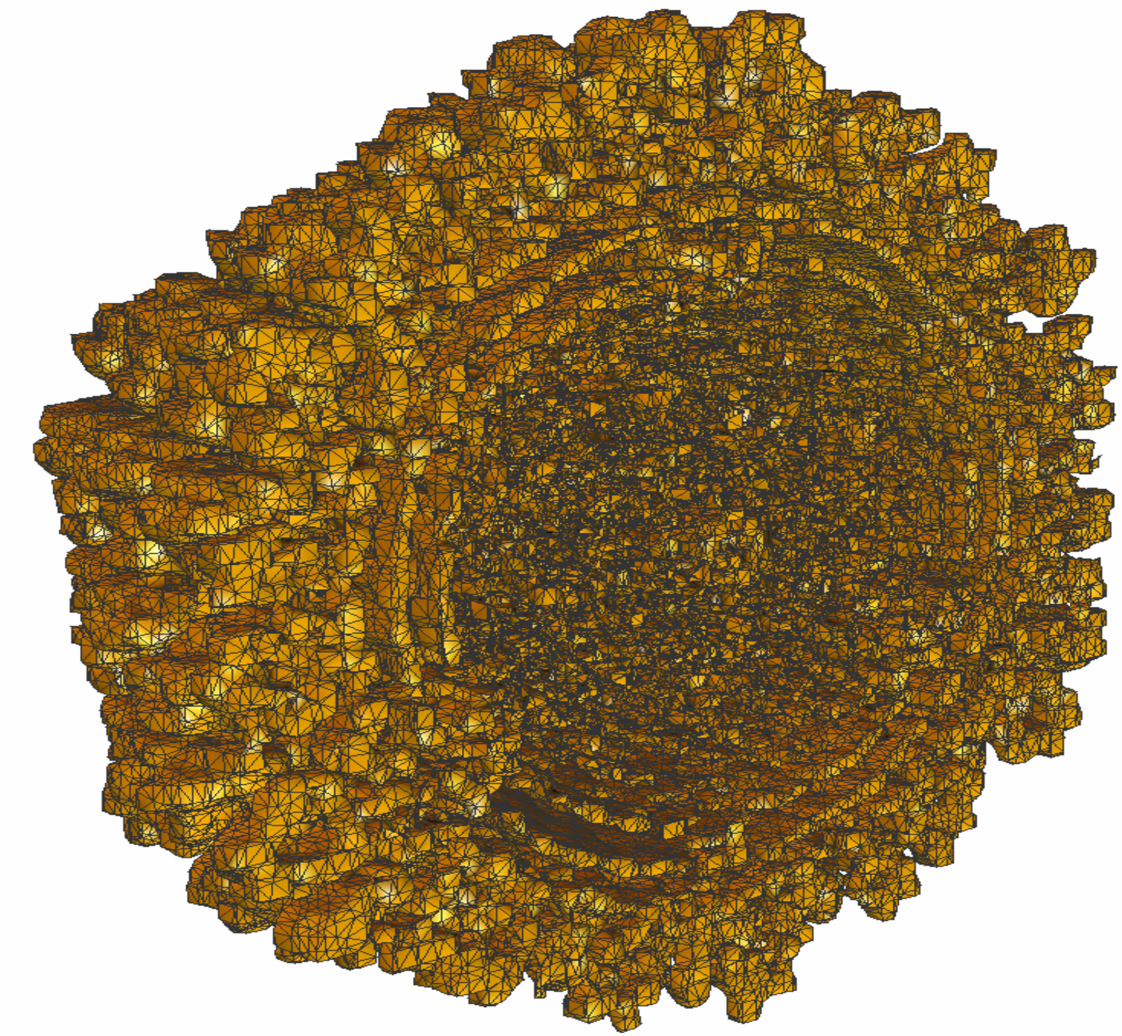


The active site groove is inside the red box. Adaptive meshes are generated in order to keep the accuracy of the groove, and reduce the number of elements at the same time.

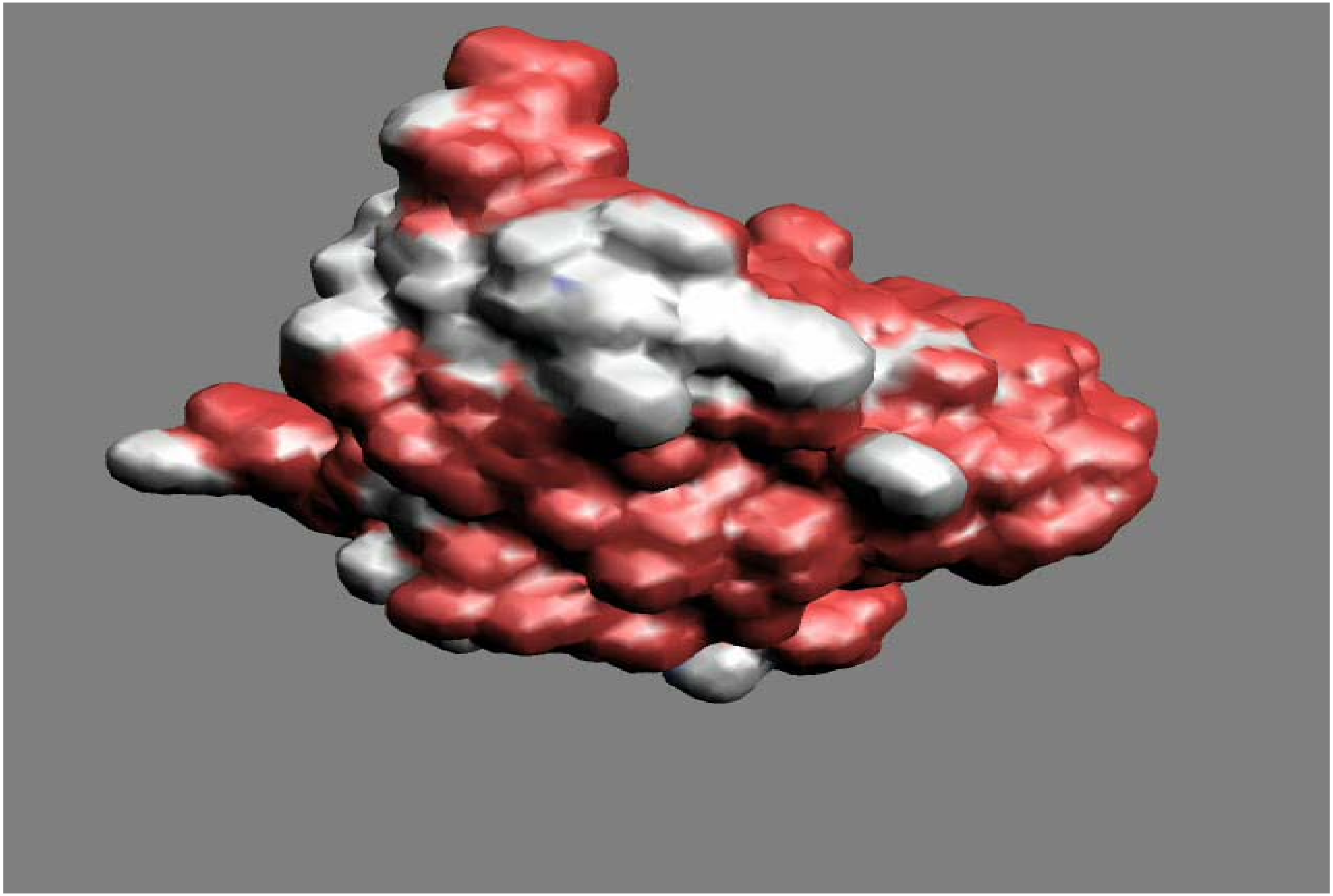
94847 vertices and 497327 tetra

Zhang, Bajaj, Sohn, ACM Solid Modeling 2003

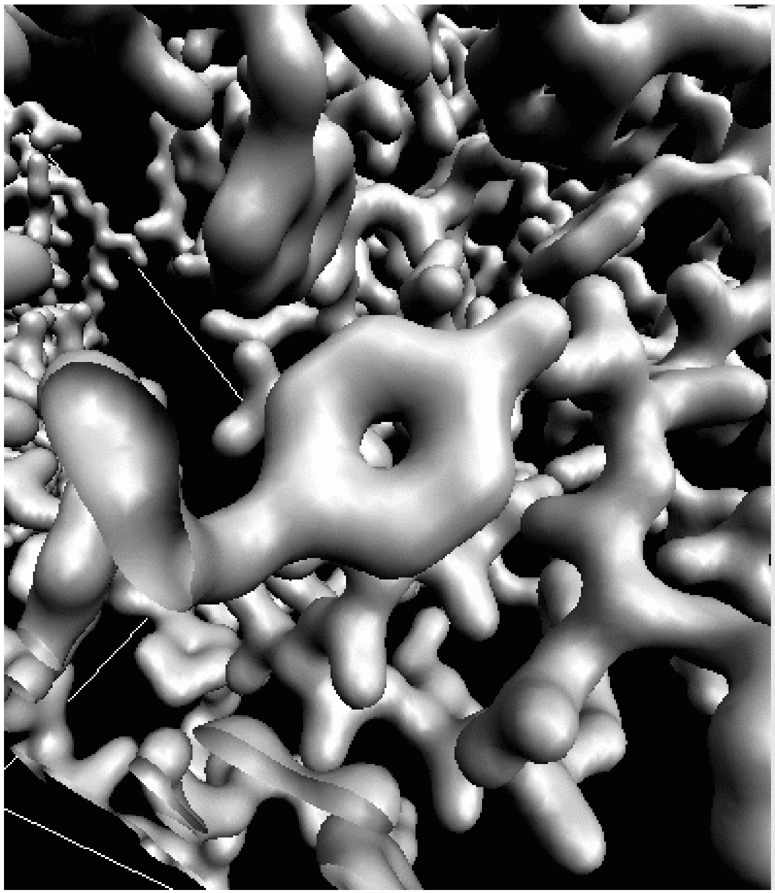
Finite Element Meshing

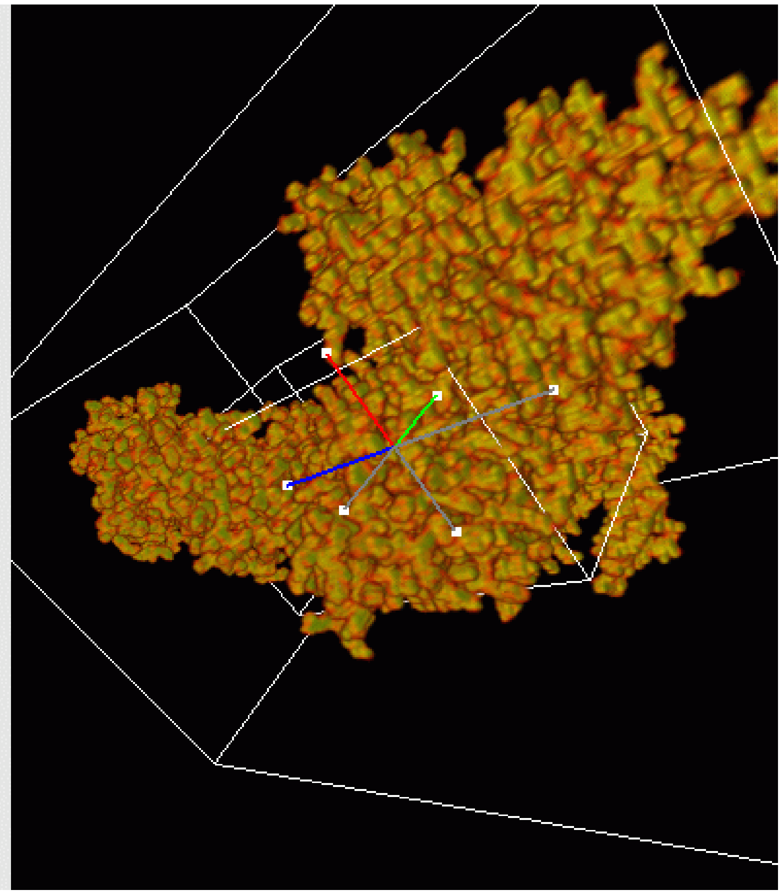


Electrostatic Potential on MACHE

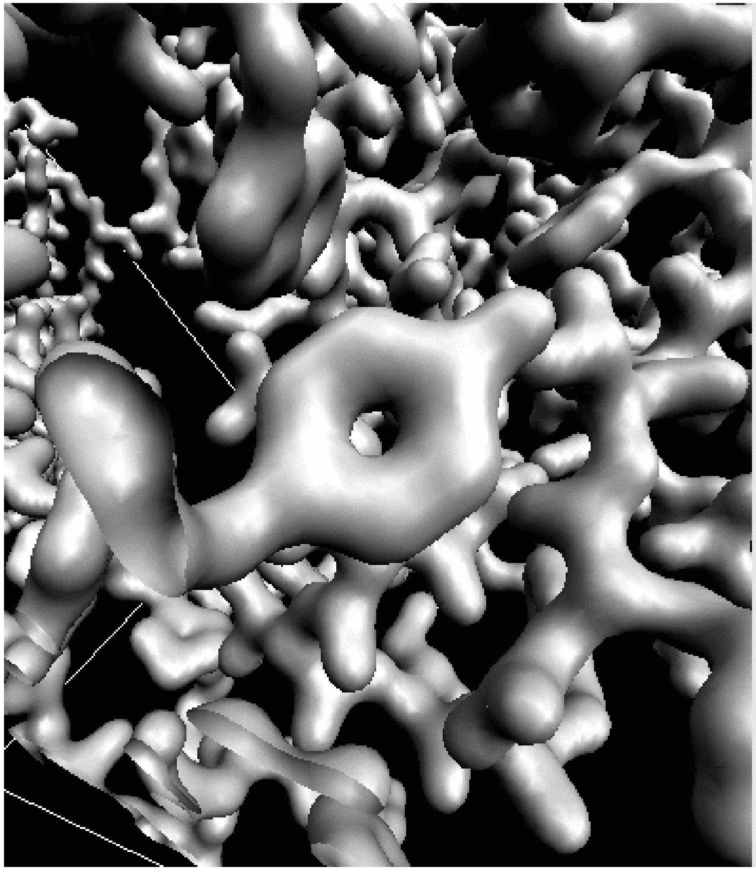


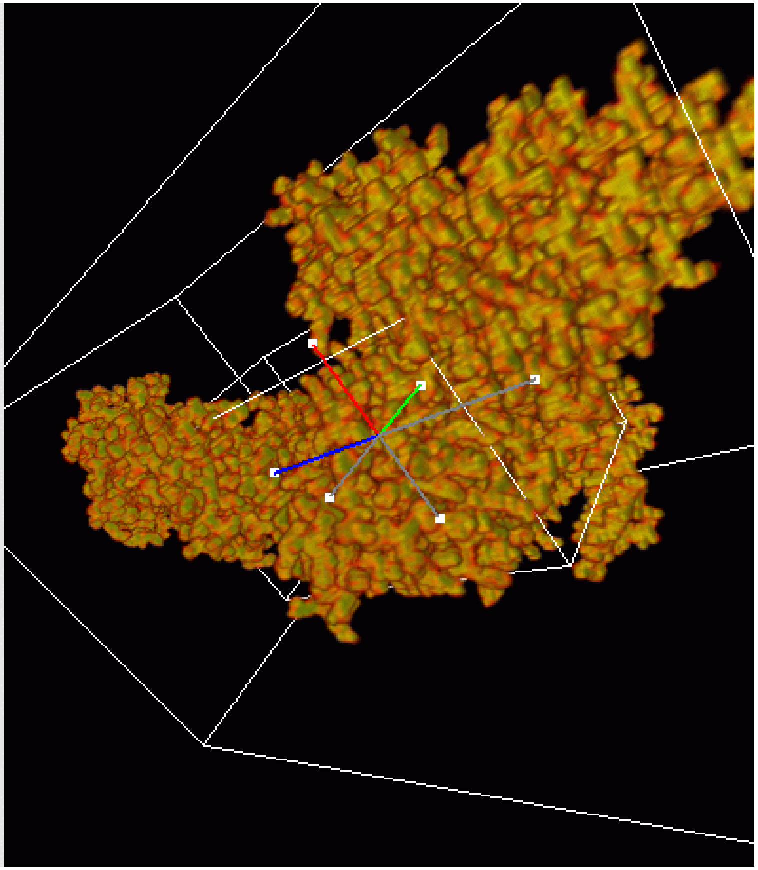
HIV-1 Reverse Transcriptase In Complex With A Polypurine Tract RNA (12,139 atoms)





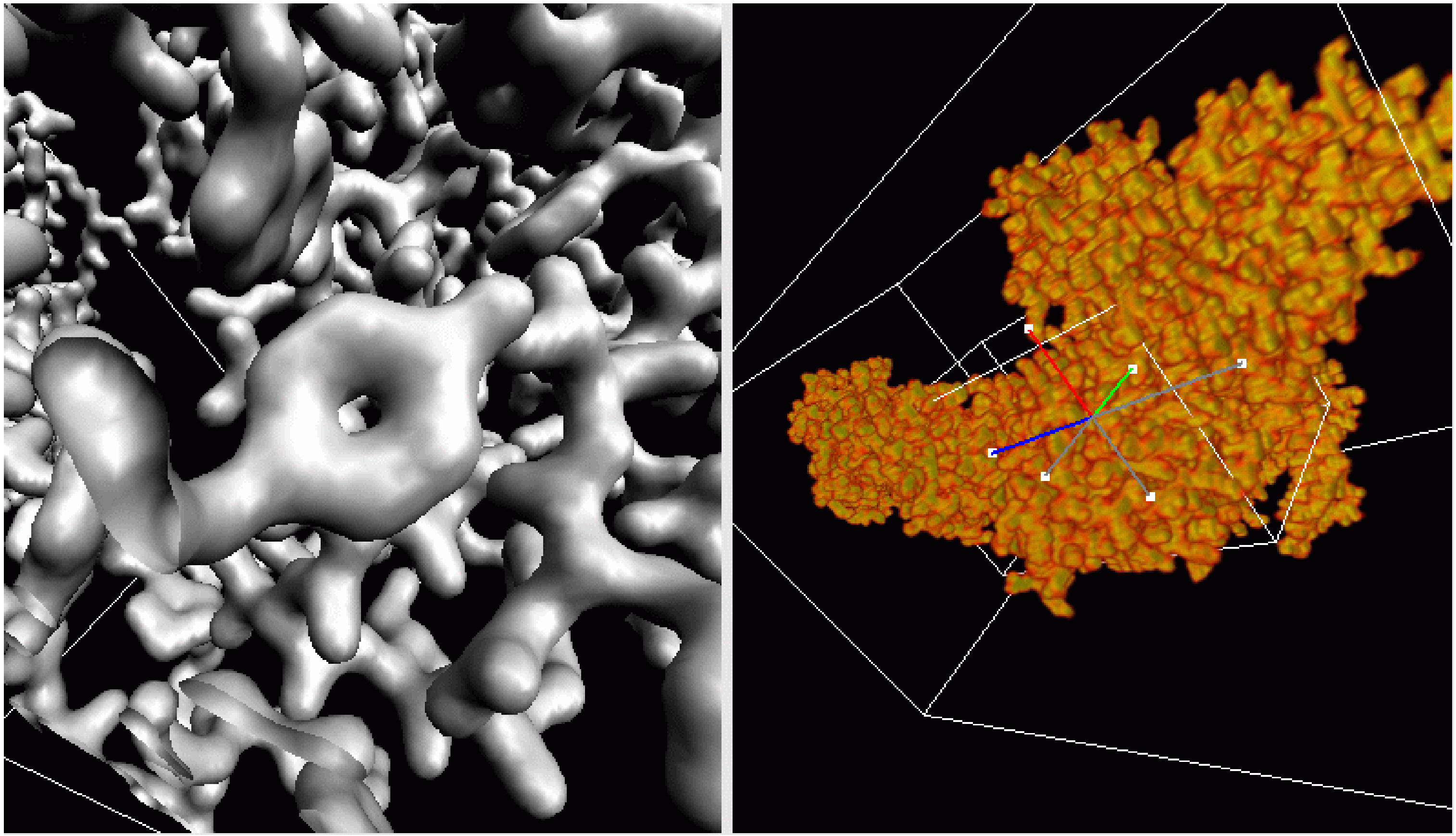
HIV-1 Reverse Transcriptase In Complex With A Polypurine Tract RNA (12,139 atoms)





Compression: 18.5:1   Error: 2.9%

HIV-1 Reverse Transcriptase In Complex With A Polypurine Tract RNA (12,139 atoms)



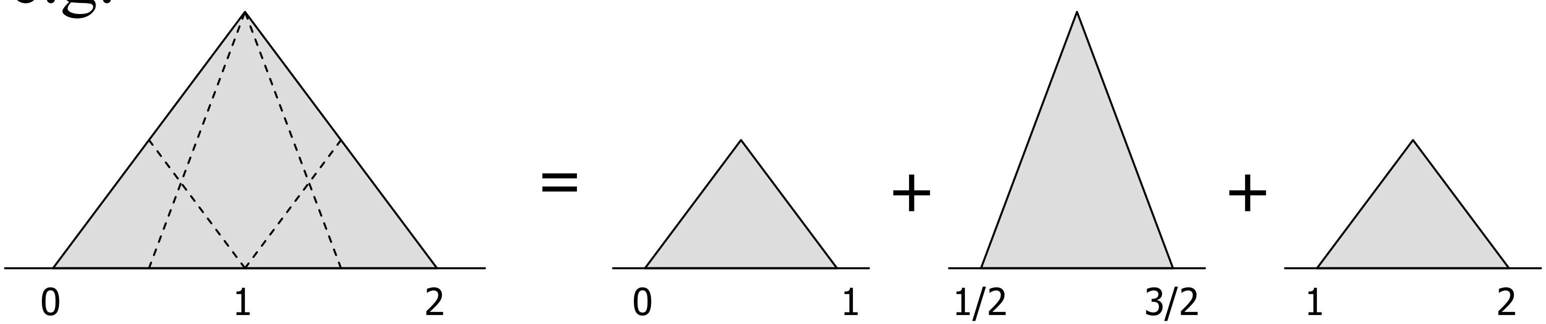
Compression: 49:1   Error: 6.9%

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

Background (Classical Wavelet Representations)

Key idea: refinement

e.g.



$\phi(x)$ $\quad\quad$ $\frac{1}{2}\phi(2x)$ $\quad$ $\phi(2x-1)$ $\quad$ $\frac{1}{2}\phi(2x-2)$

$$\phi(x) = \sum_k h_0[k]\phi(2x-k) \quad \text{with} \quad h_0[k] = \left\{\tfrac{1}{2}, 1, \tfrac{1}{2}\right\}$$

Refinable functions are called scaling functions

Wavelet Representations

Wavelets are also linear combinations of scaling functions

$$\psi(x) = \sum_k h_1[k]\phi(2x-k)$$

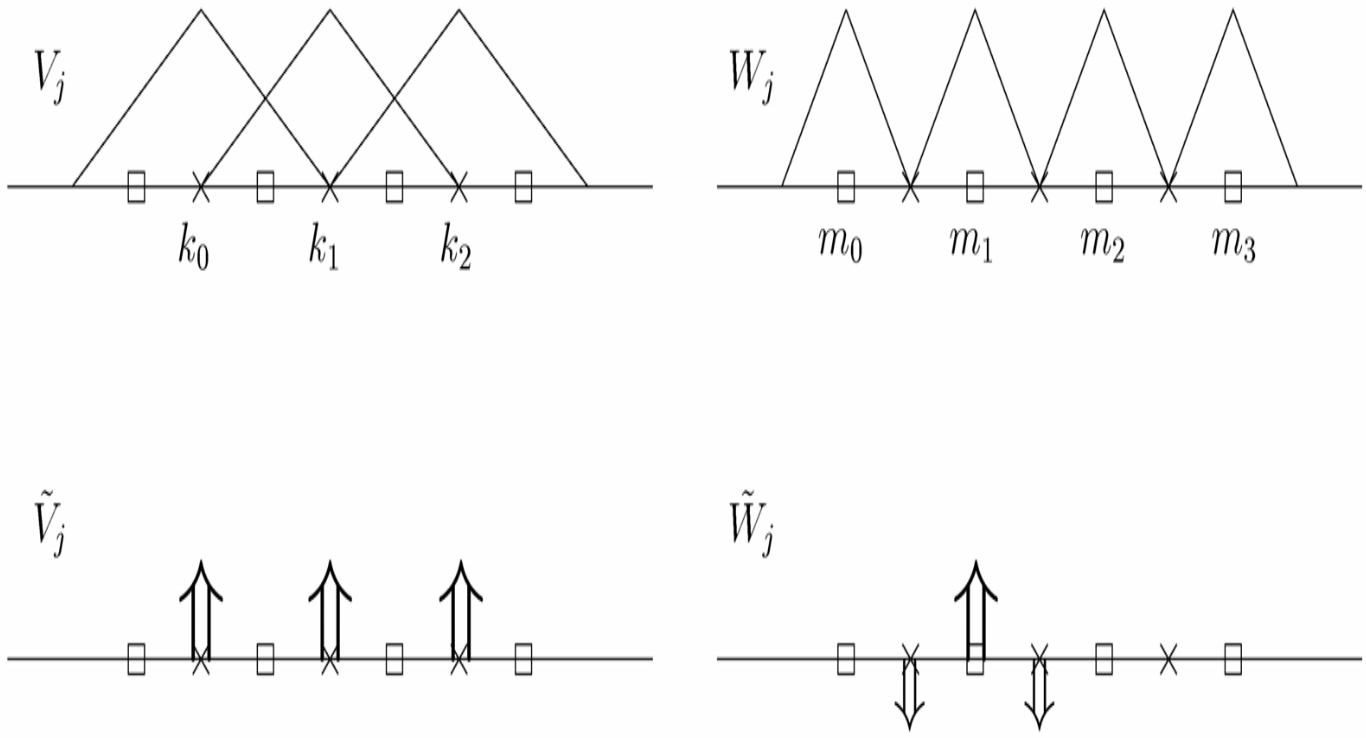
Usual design criteria for $h_0[k]$ and $h_1[k]$

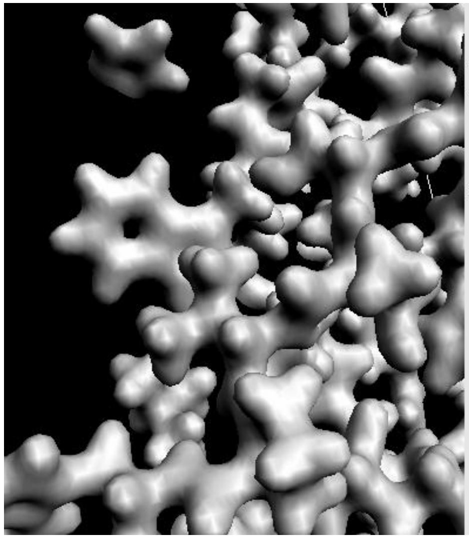
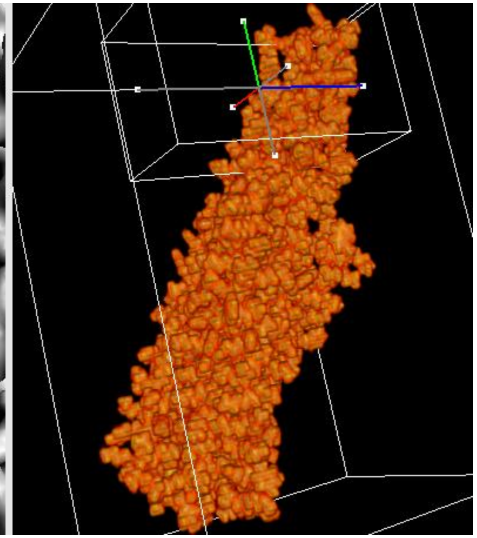
– finite length => makes wavelet and scaling functions compactly supported

– Vanishing moments:

$$\int \psi(x)x^m dx = 0$$

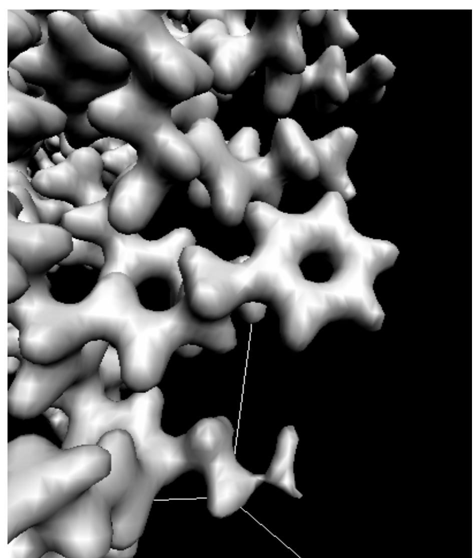
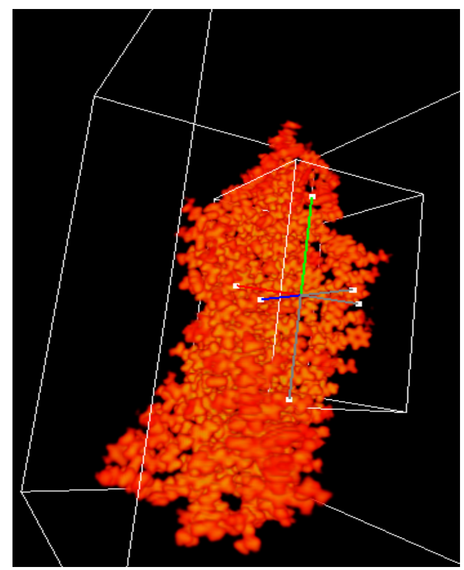
2^{nd} Generation Wavelets Based on Hierarchical Basis and a Lifting Scheme

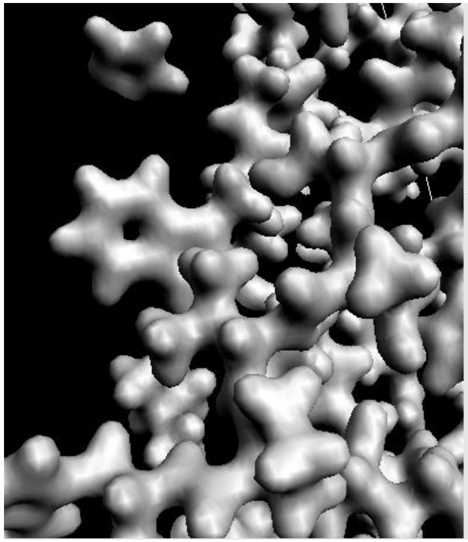


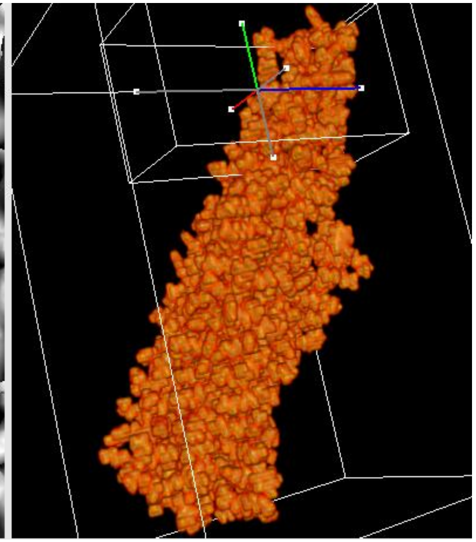



Original data set to be Compressed by Linear Hierarchal Basis

Original data set To be Compressed by Haar Wavelets

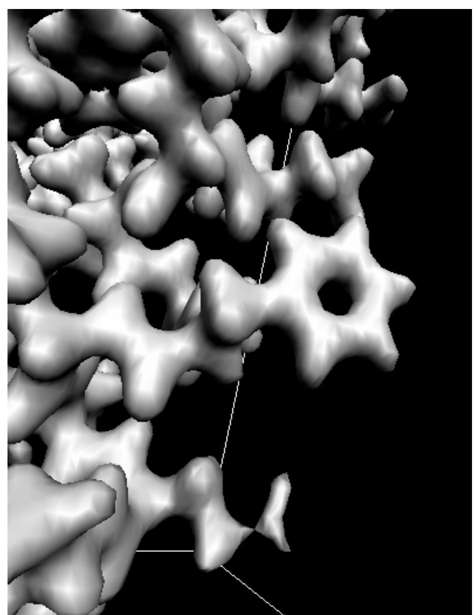



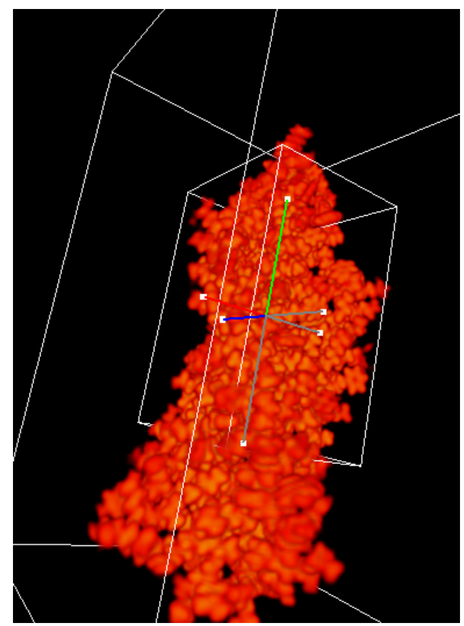


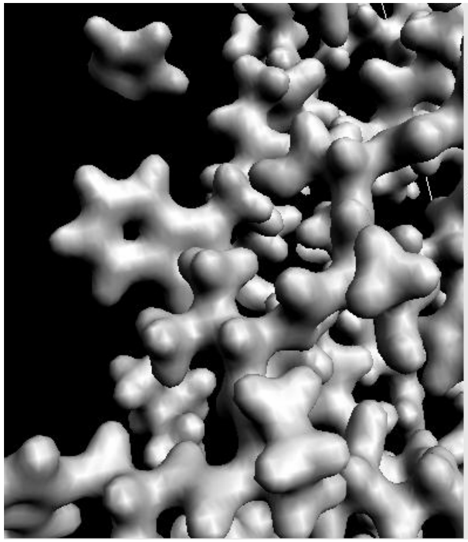


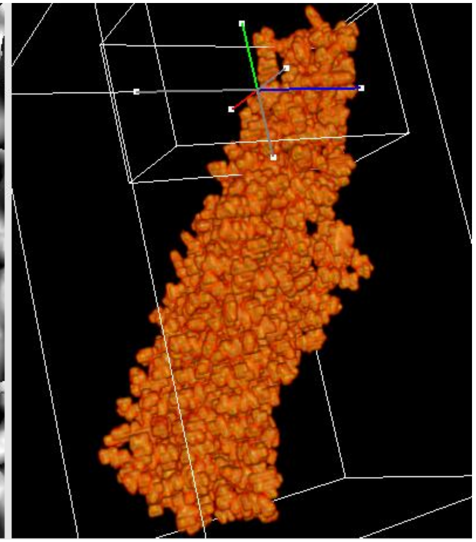
Linear Hierarchal Basis
Total Compression(TC):37

Haar Wavelets
TC:20



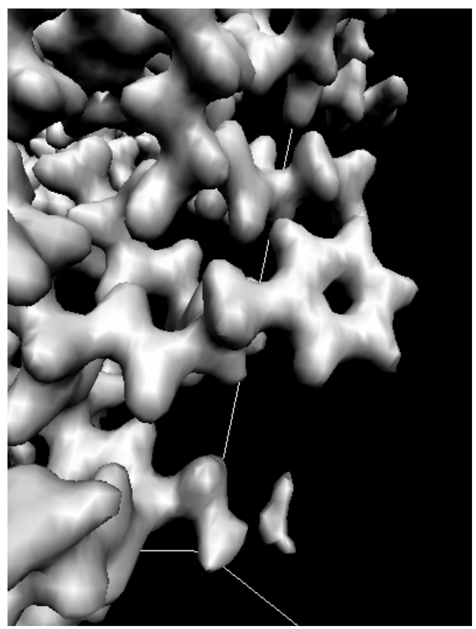


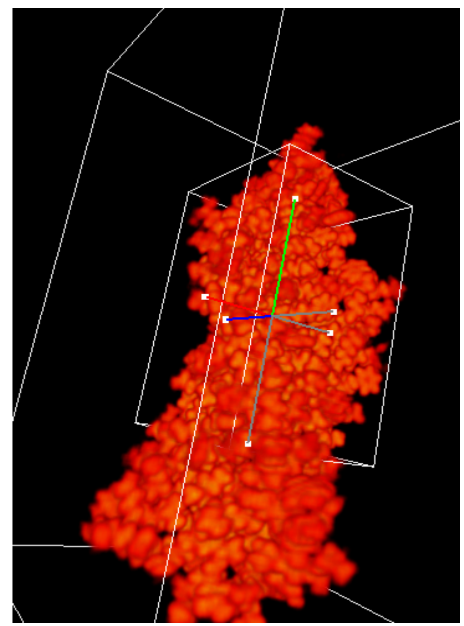


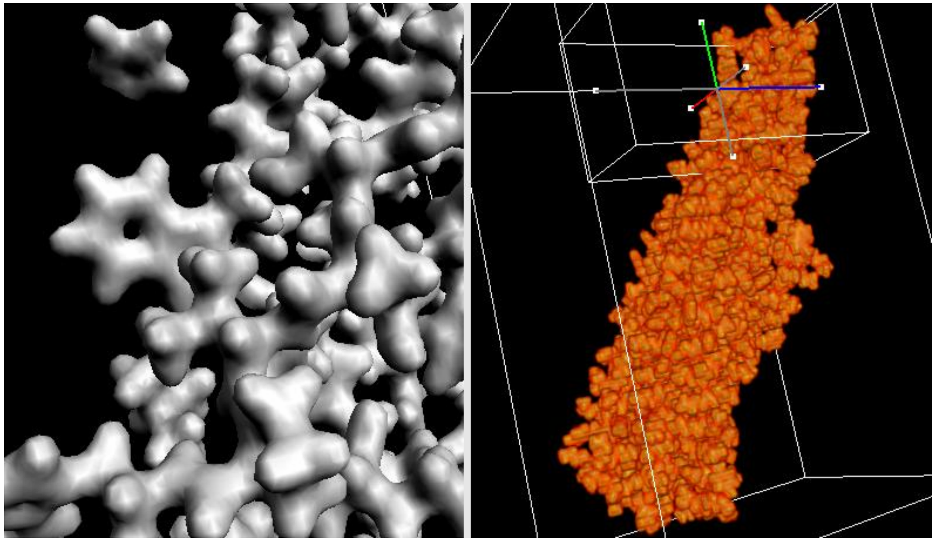


Linear Hierarchal Basis
Total Compression:37

Haar Wavelets
TC:33

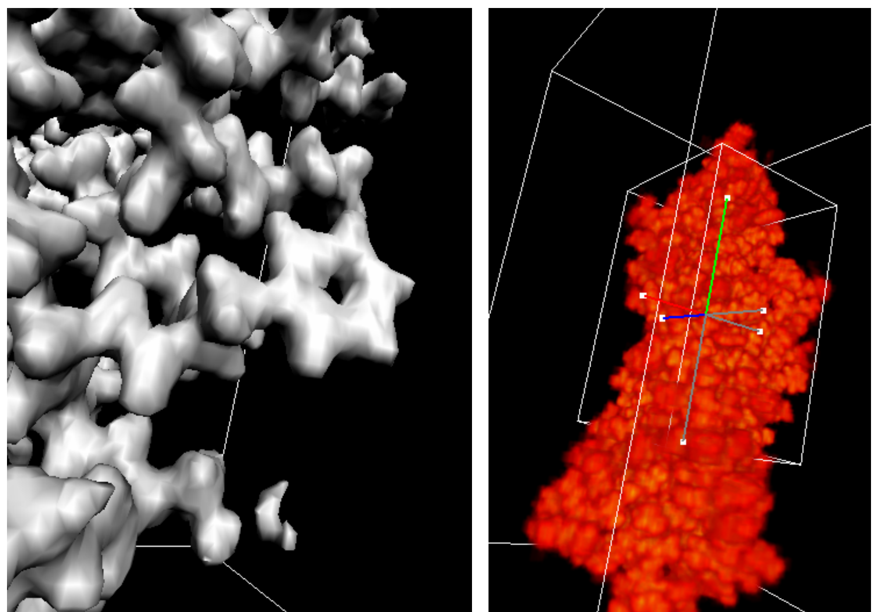


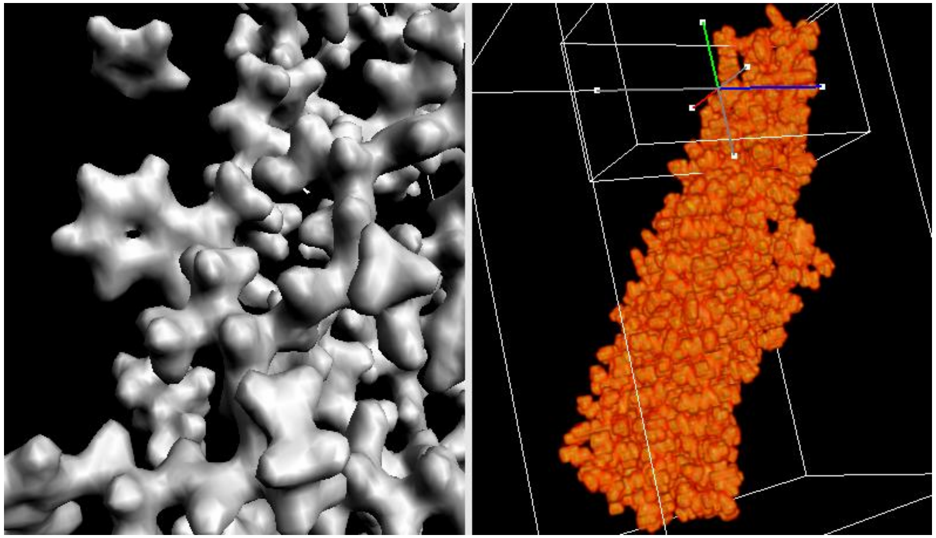




Linear Hierarchal Basis
TC:70.8

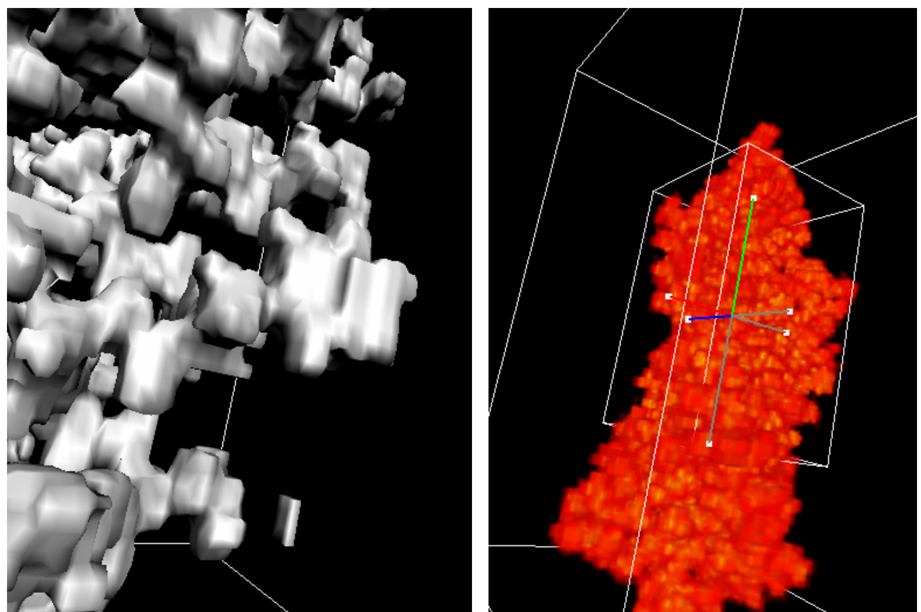
Haar Wavelets
TC:70.8

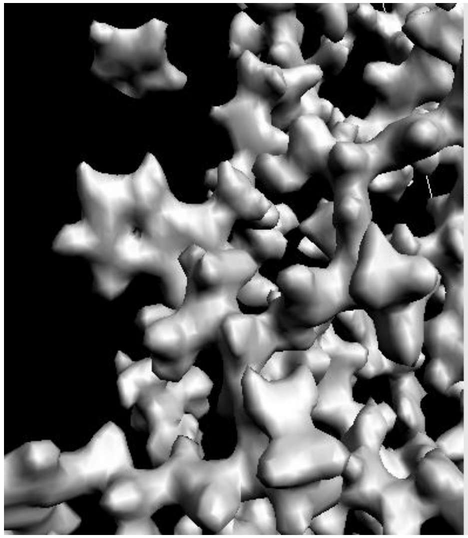


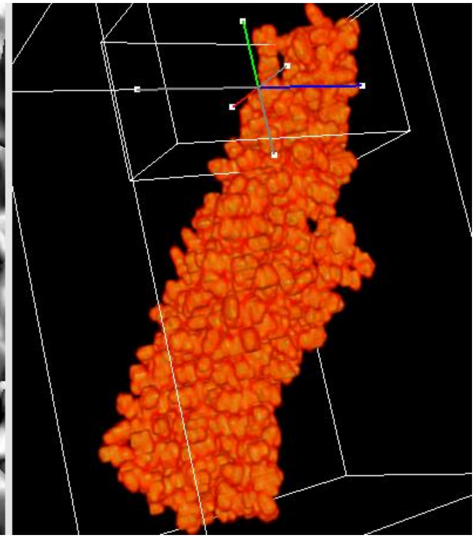


Linear Hierarchal Basis
TC:206.9

Haar Wavelets
TC:206.9

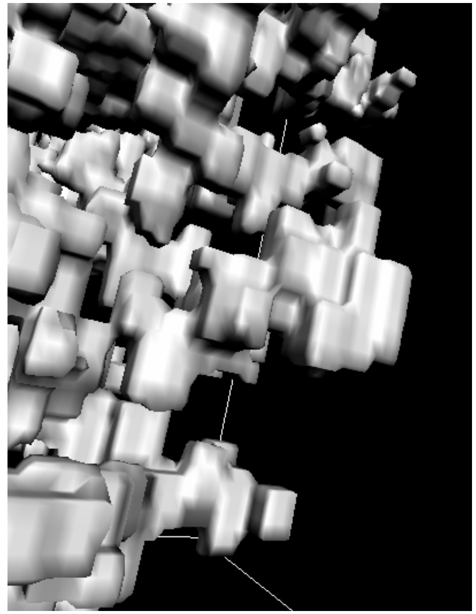


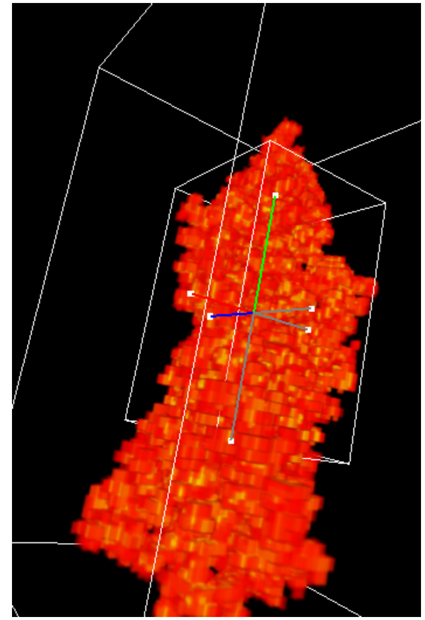




Linear Hierarchal Basis
TC:571

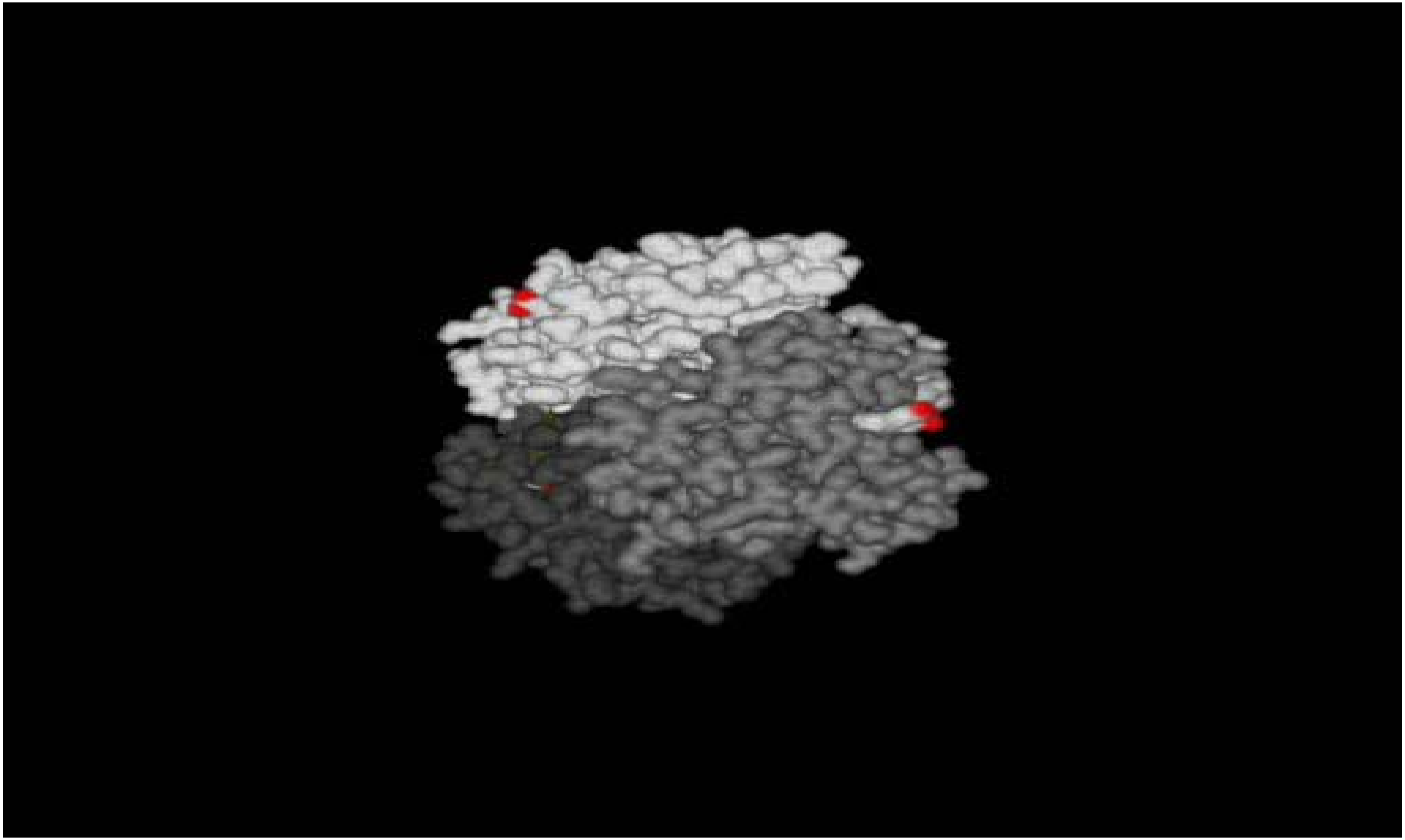
Haar Wavelets
TC:571



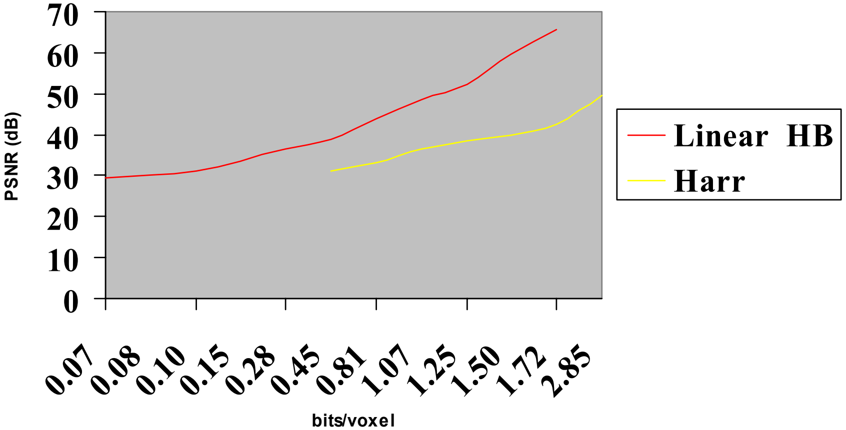


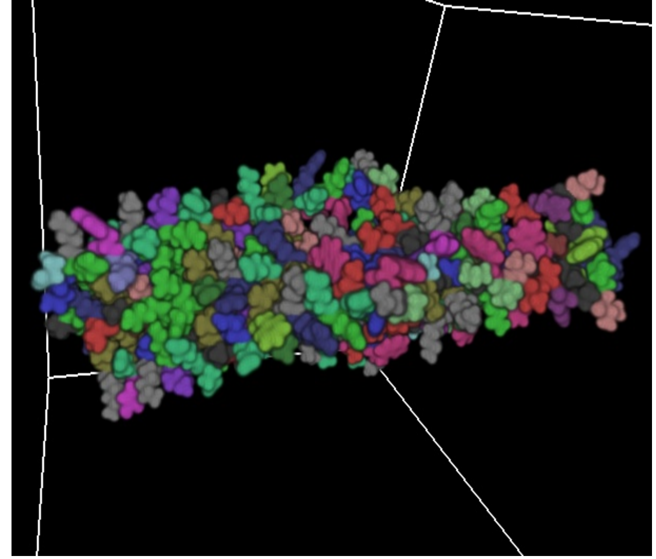
Visualization of Hemoglobin Dynamics

Interrogative Volumetric Video (VolVis2002)

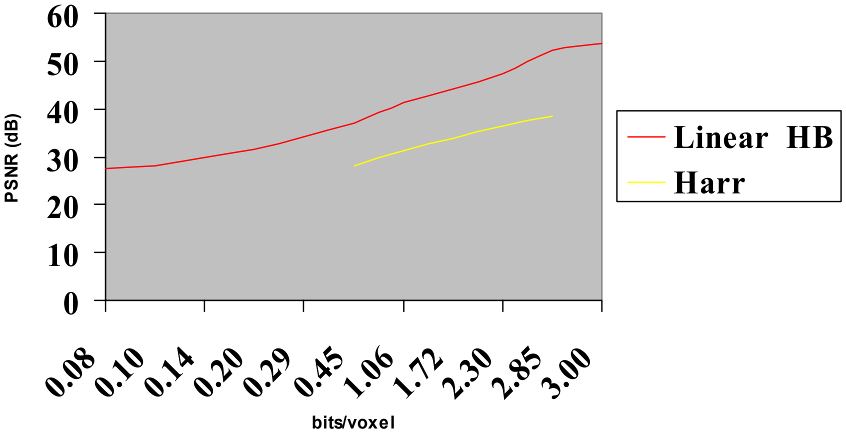


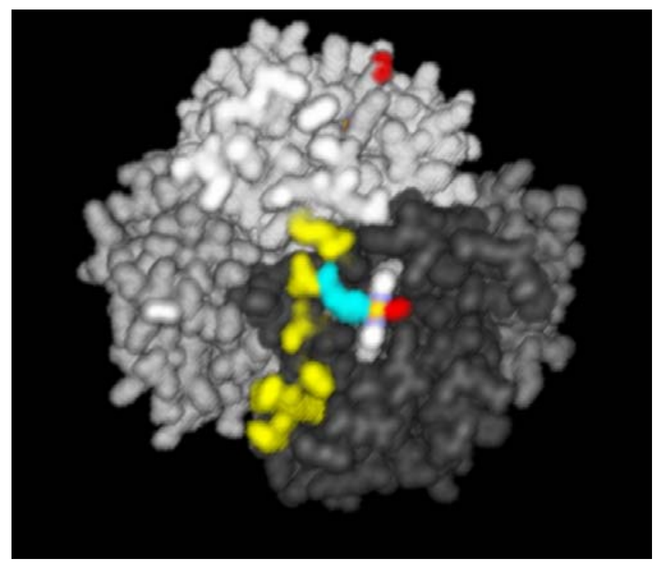
Rate Distortion (2EZP)



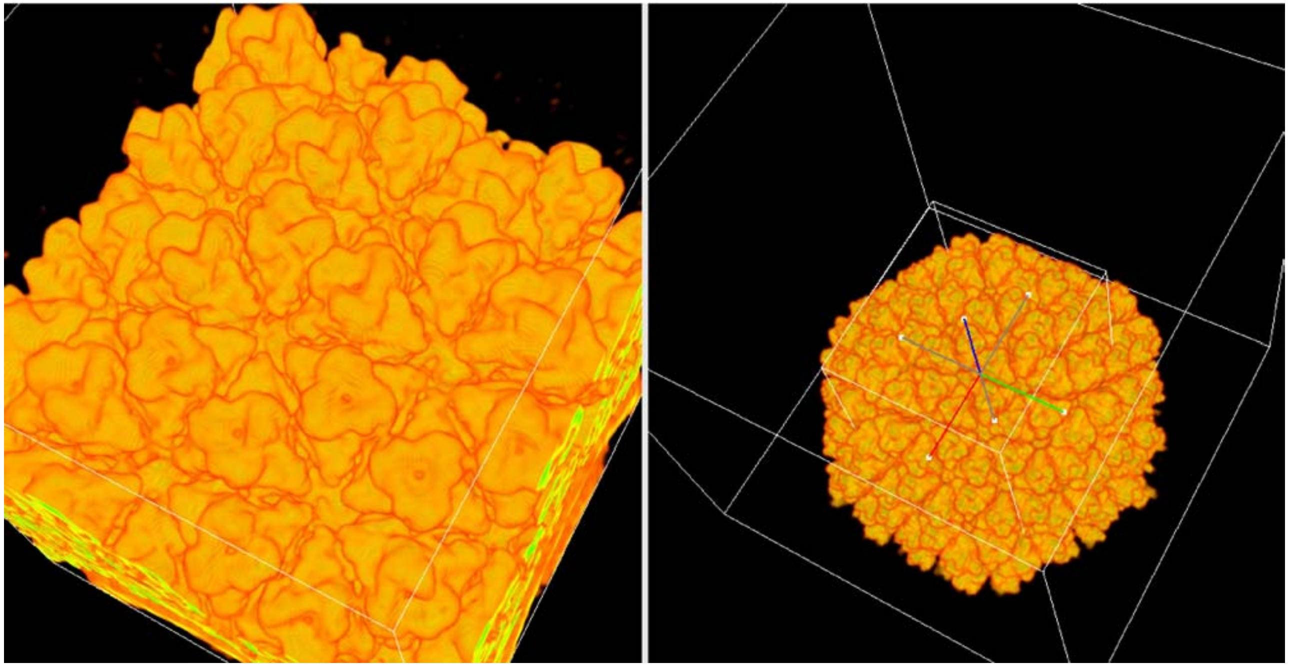


Rate Distortion (Hemoglobin)



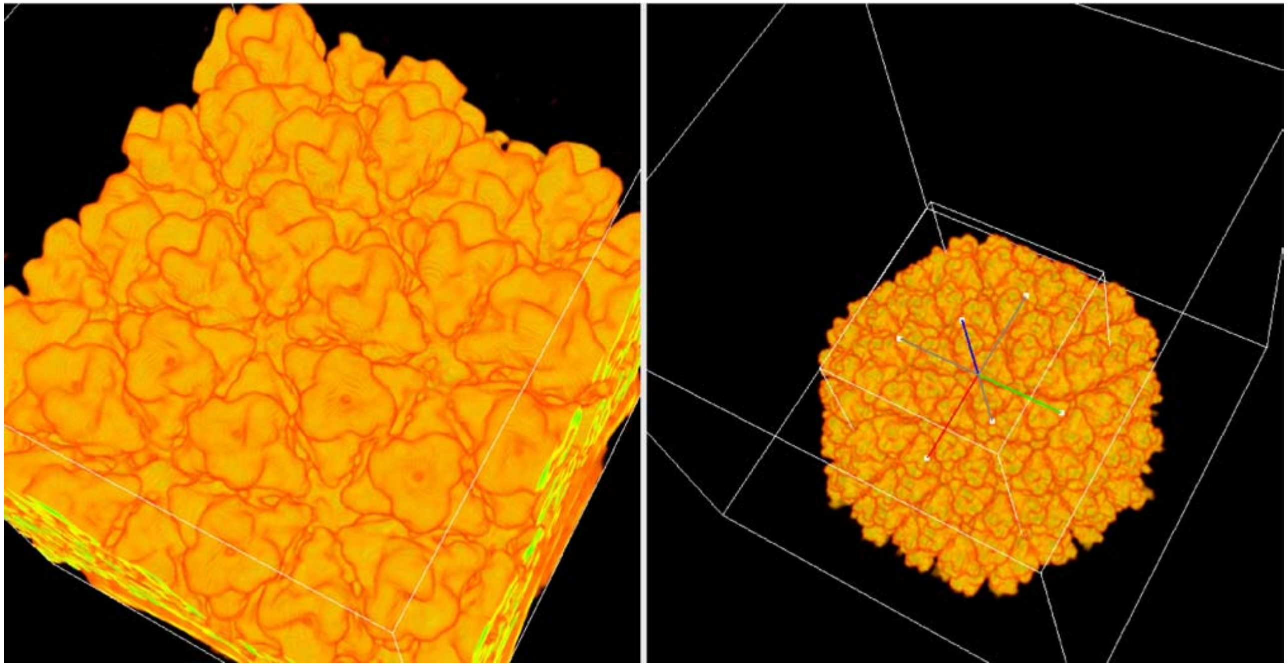


Rice Dwarf Virus(Smoothed)



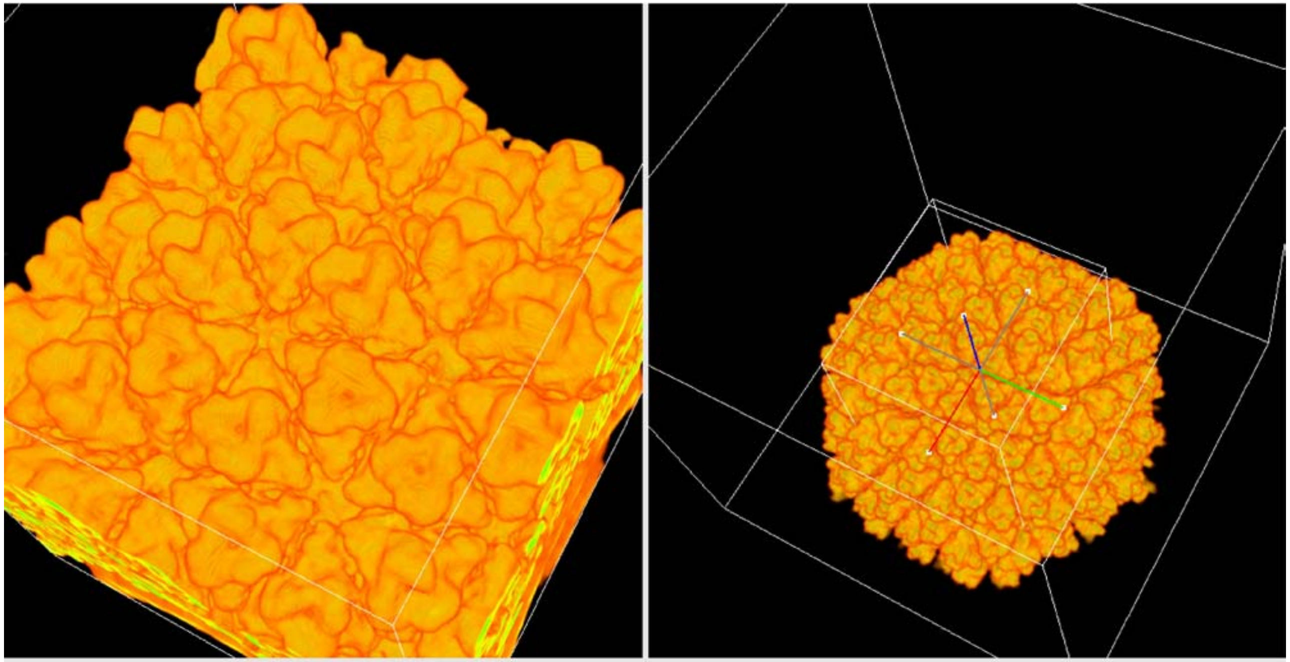
Original

Rice Dwarf Virus(Smoothed)



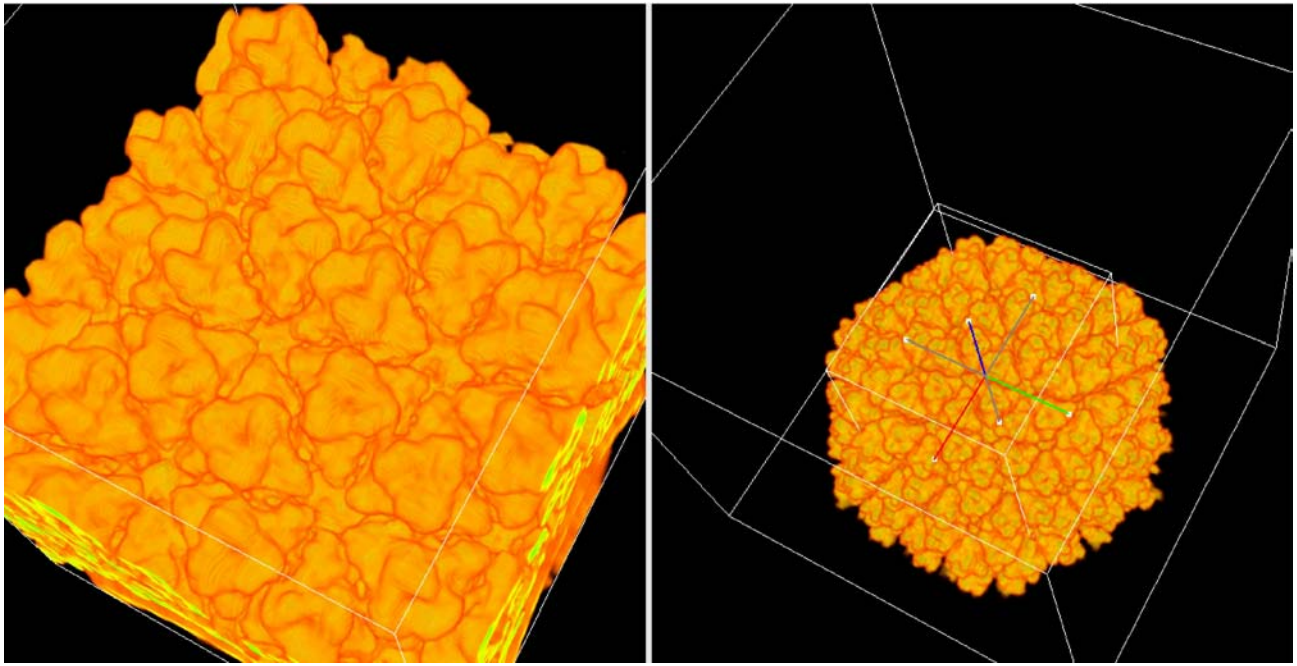
TC=31.6, PSNR =42.4dB

Rice Dwarf Virus(Smoothed)



TC=120.3, PSNR =36.5dB

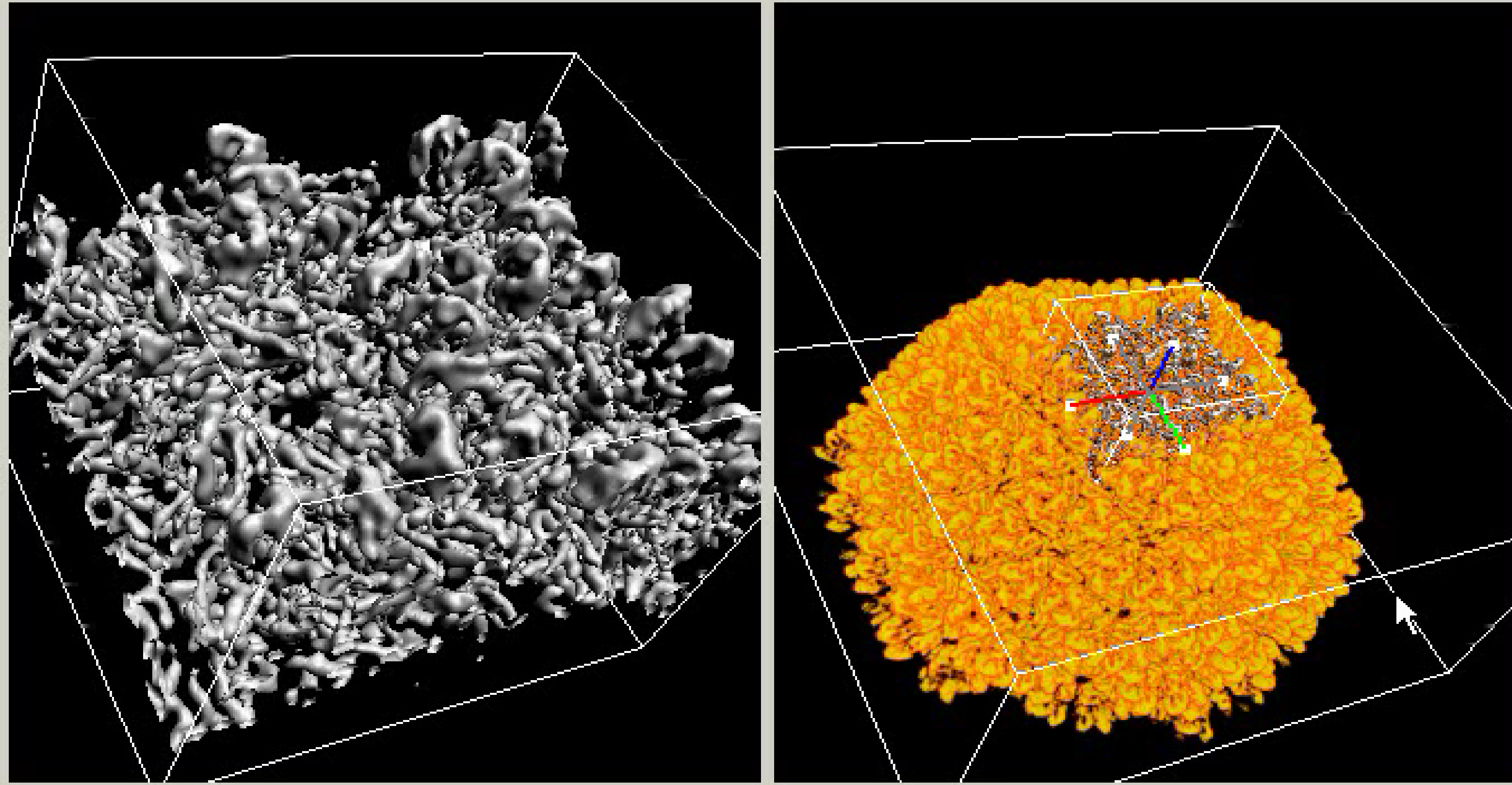
Rice Dwarf Virus(Smoothed)



TC=328.3, PSNR =31.6dB

Multi- Resolution Volume Exploratoration

(http://www.ices.utexas.edu/CCV/software/)



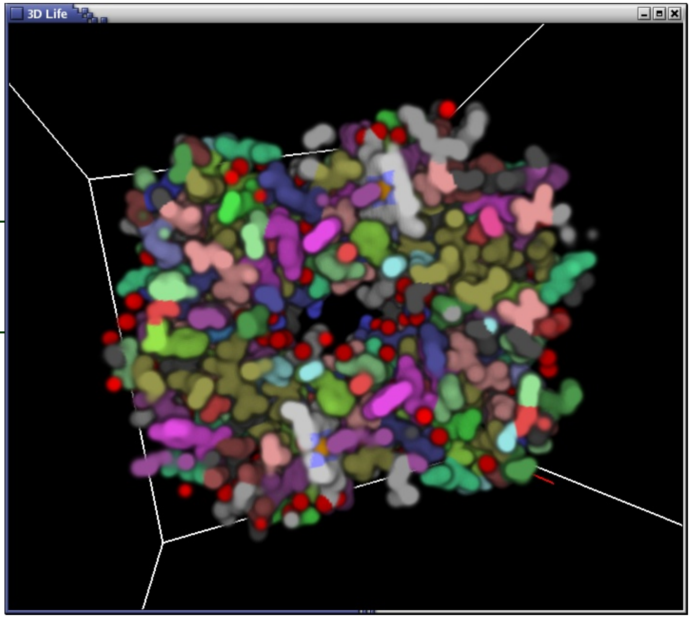
Volume Rover (CORBA client)

Multi-Level Visualization

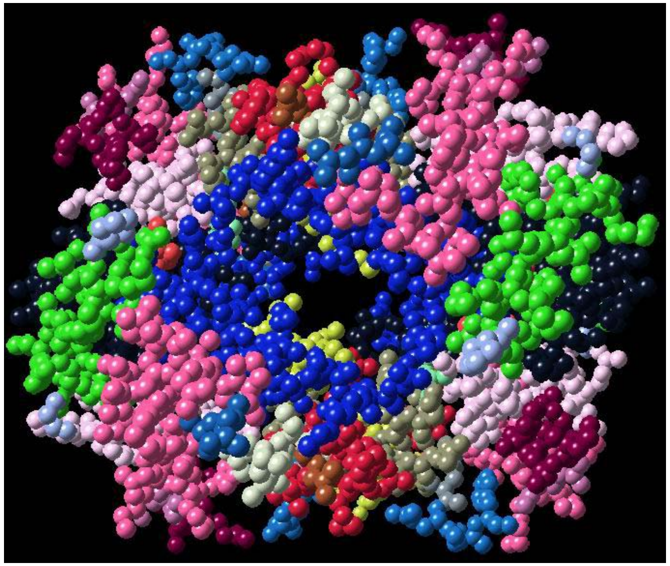
Volume Rendering

Hemoglobin

Coloring
Vua Residues

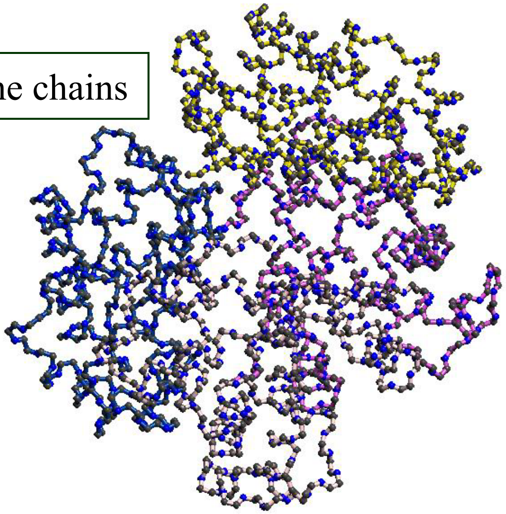


CPK

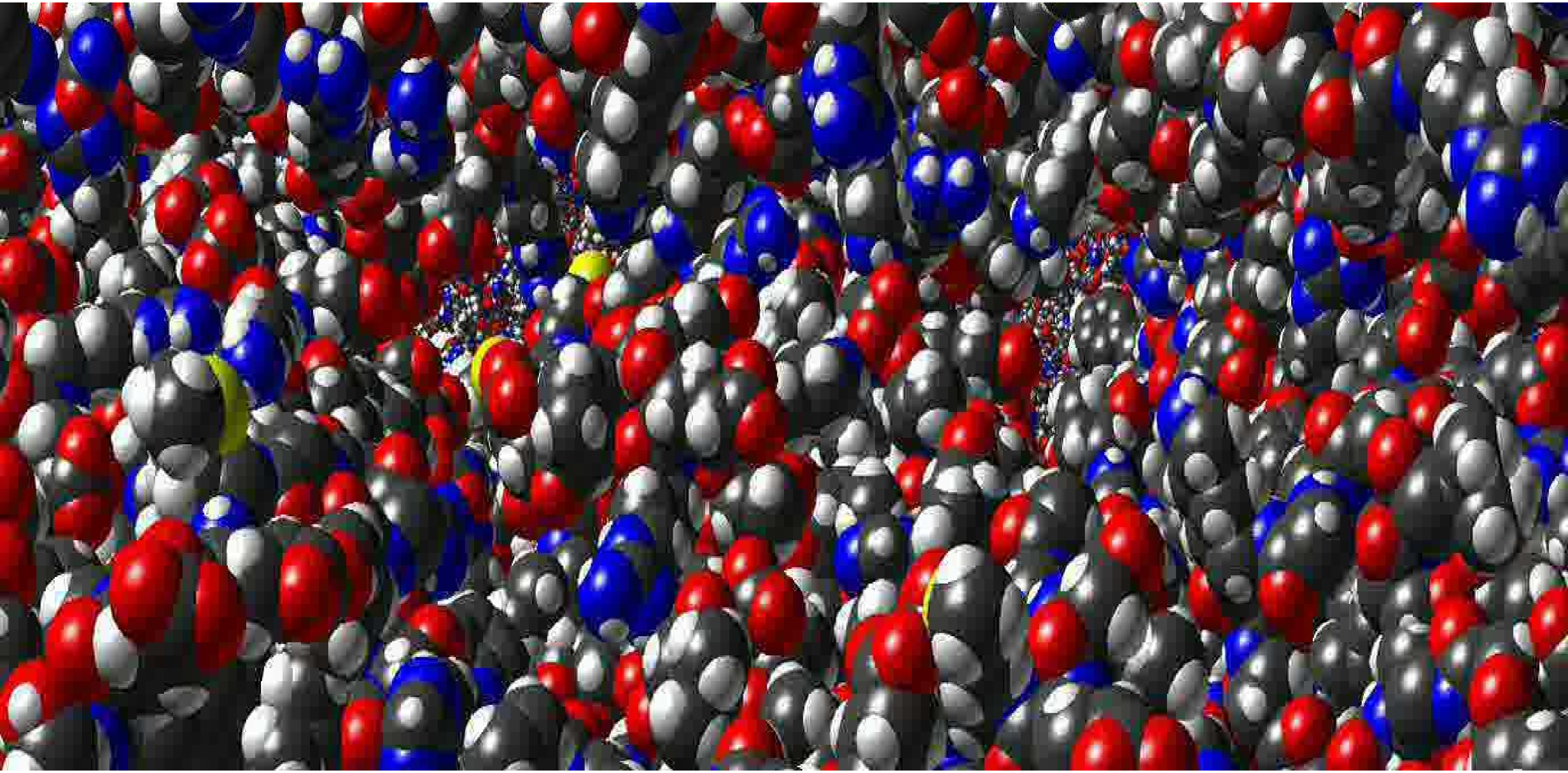


Backbone chains

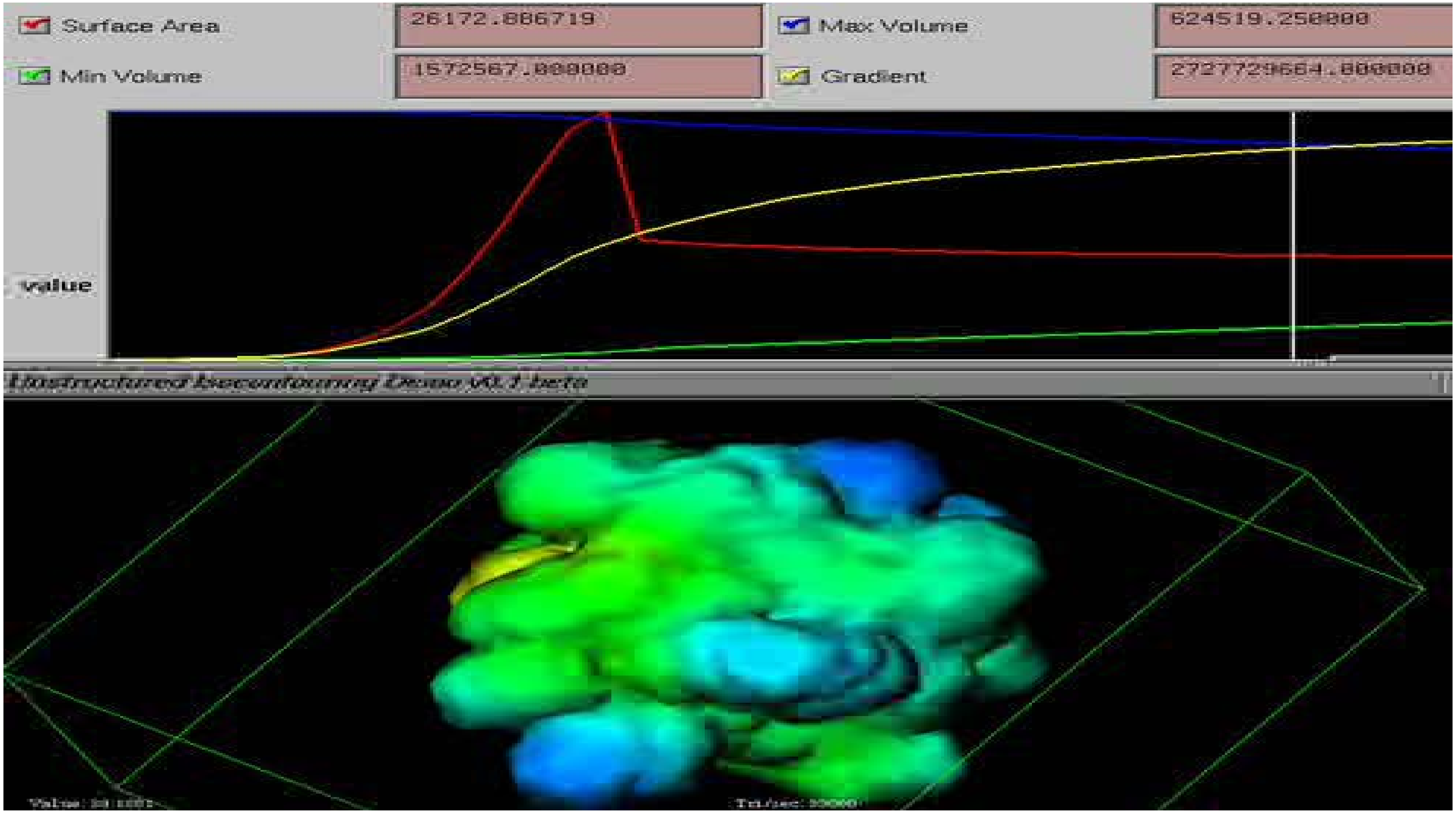
Coloring via Secondary structures



Microtubule
(Graphics Accelerated Texture-Impostors)

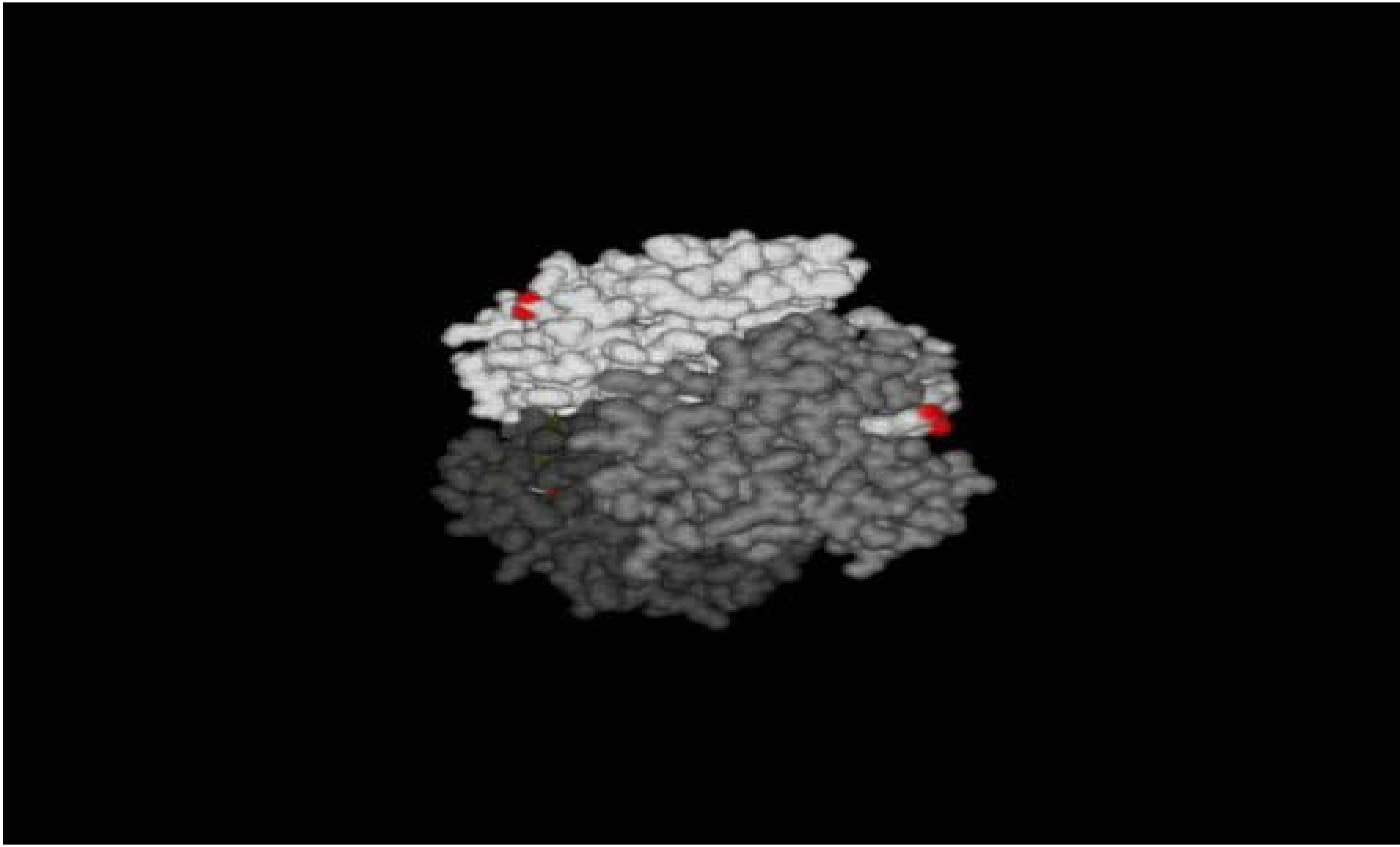


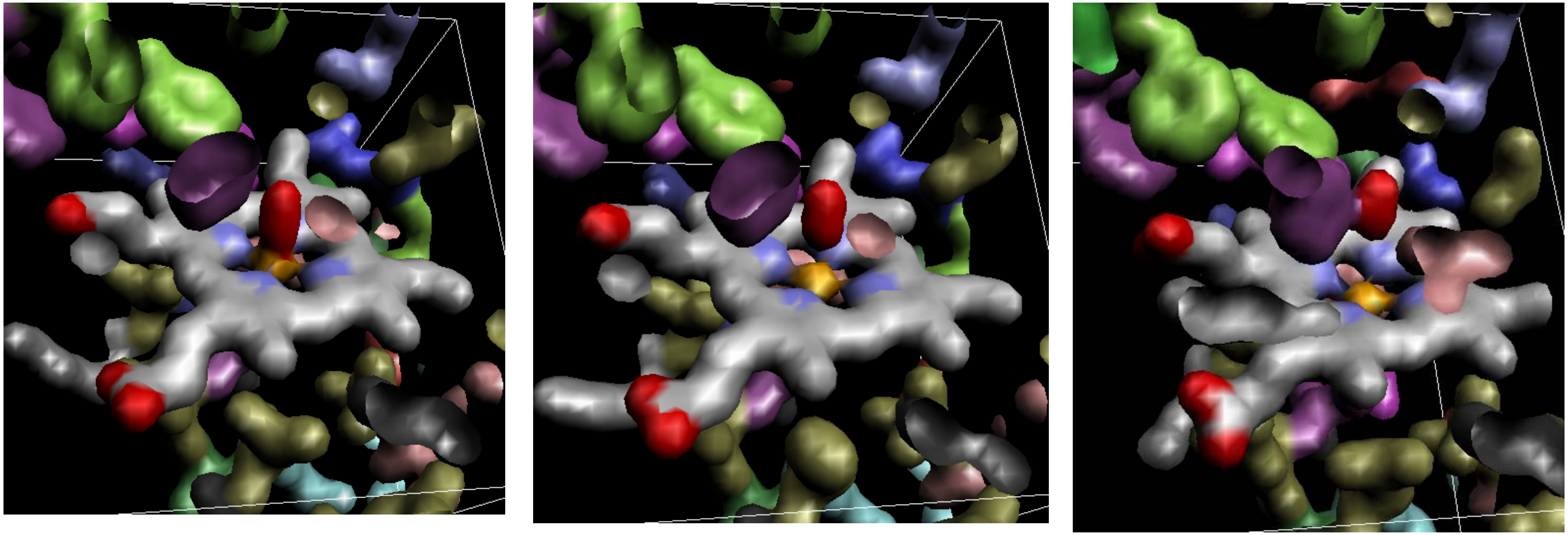
(TAQT: Topology Analyses & Quantitative Tools)



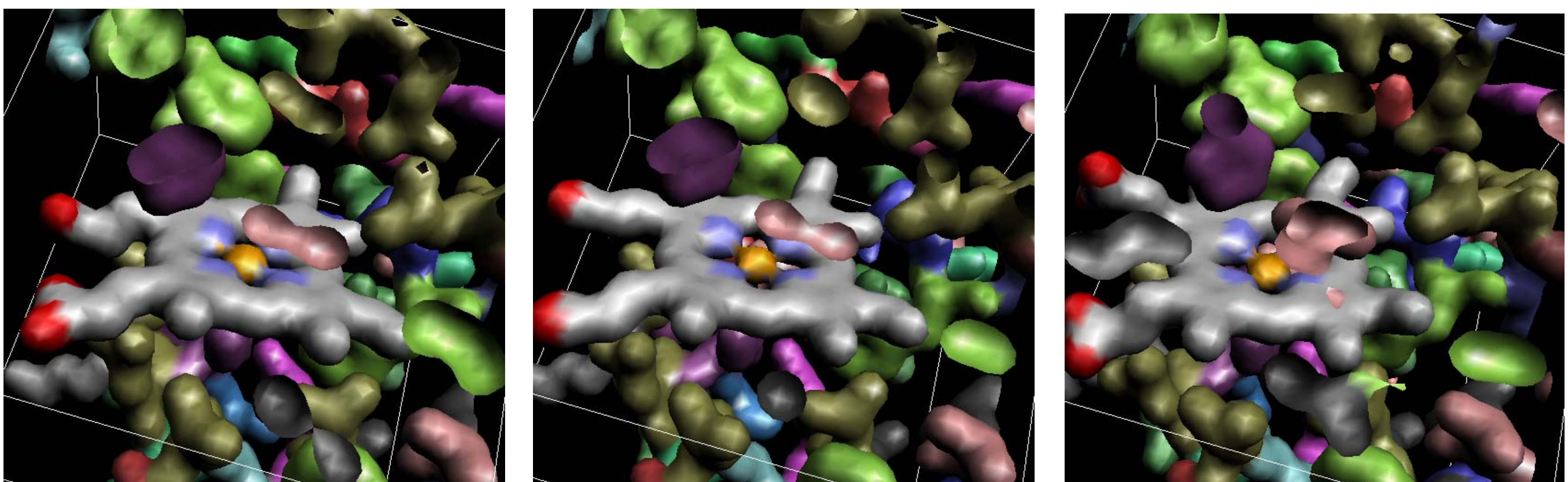
The Contour Spectrum (IEEE Vis '97)

Quantitative Visualization of Hemoglobin Dynamics
Interrogative Volumetric Video (VolVis2002)

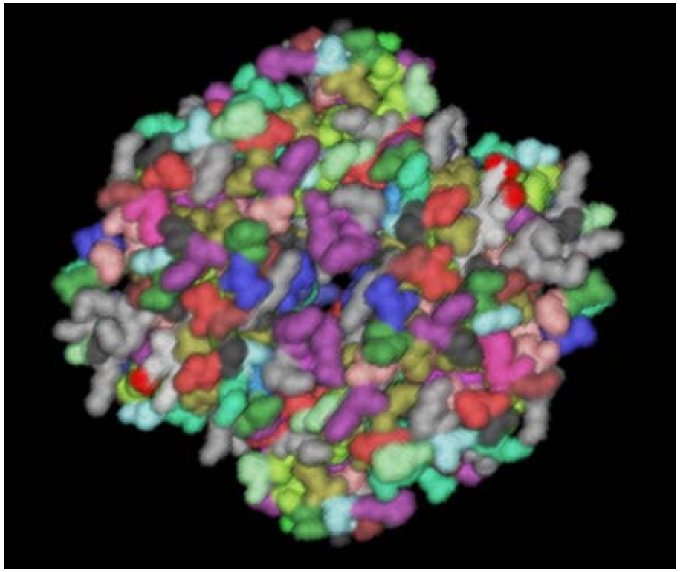




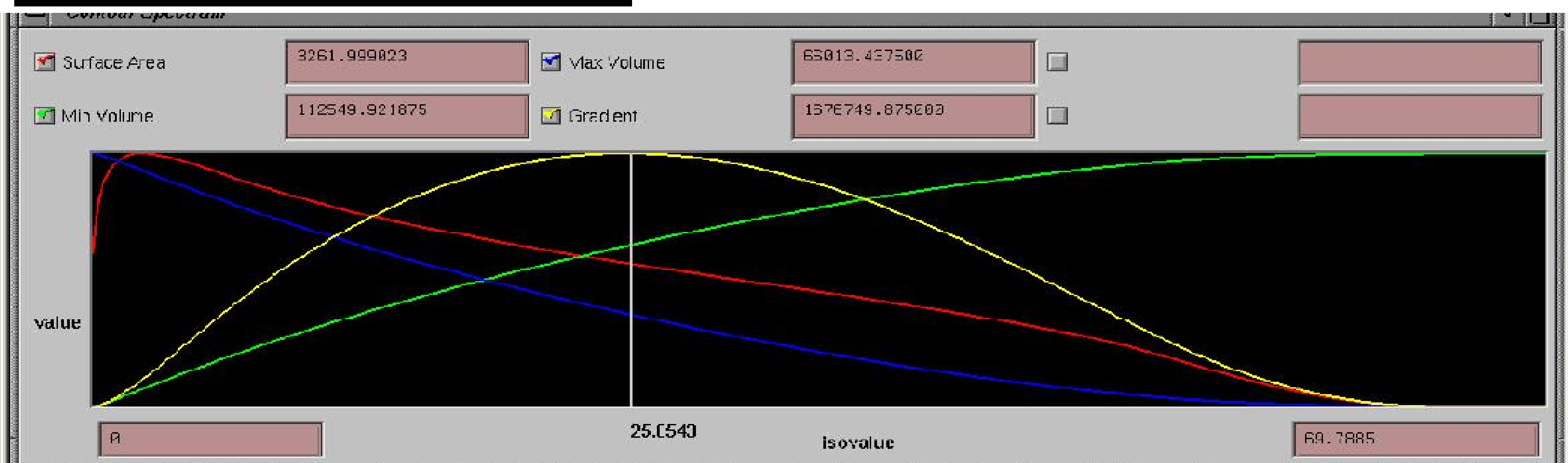
Time 1 Time 15 Time 30



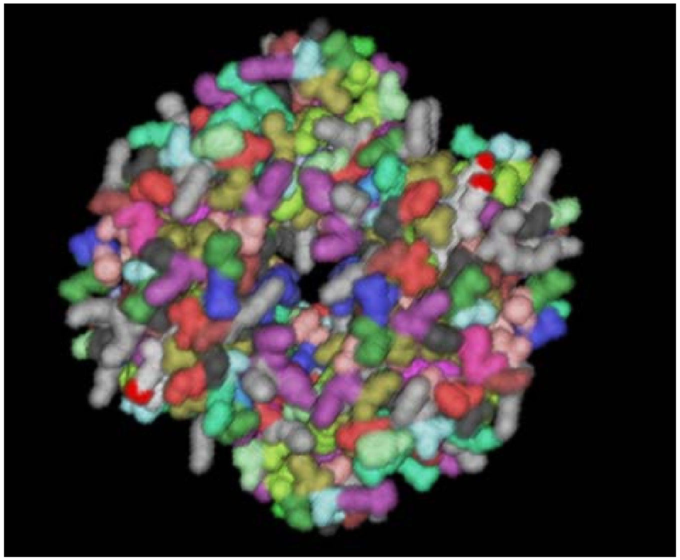
Static Contour Spectrum



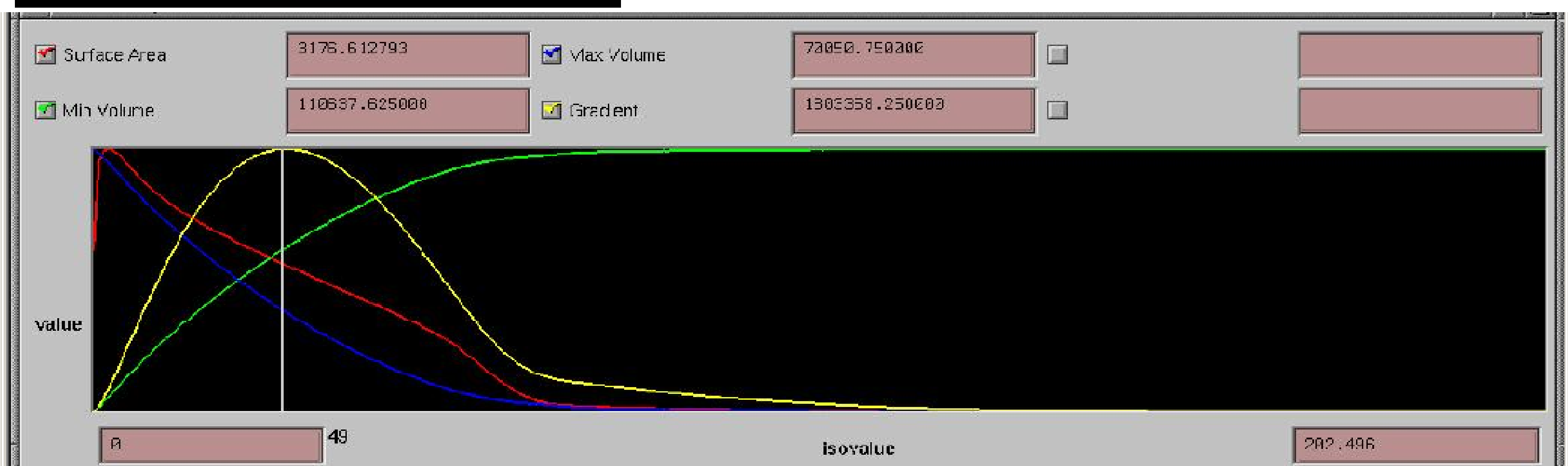
Time = 1



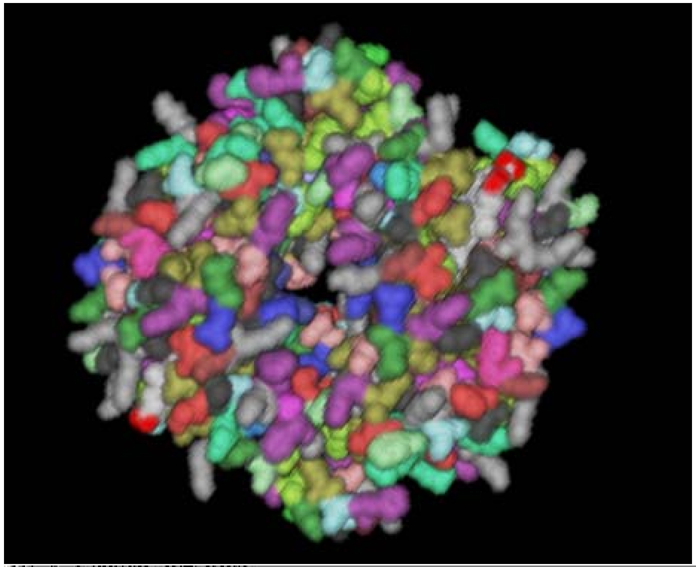
Static Contour Spectrum



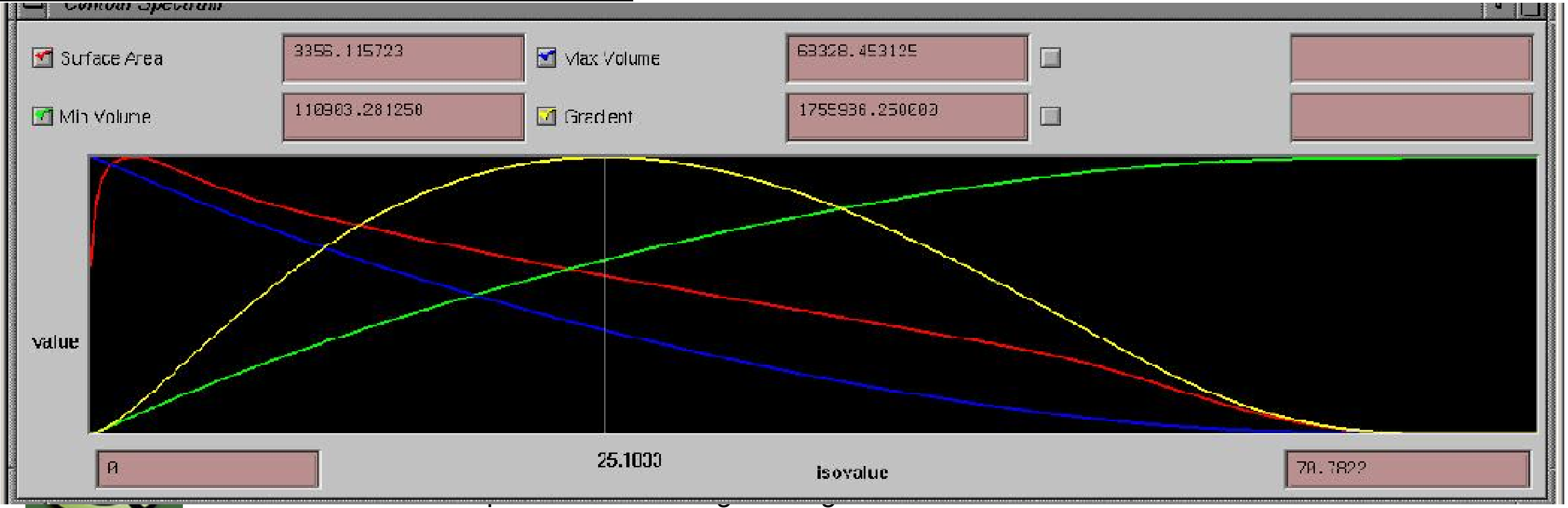
Time = 15



Static Contour Spectrum

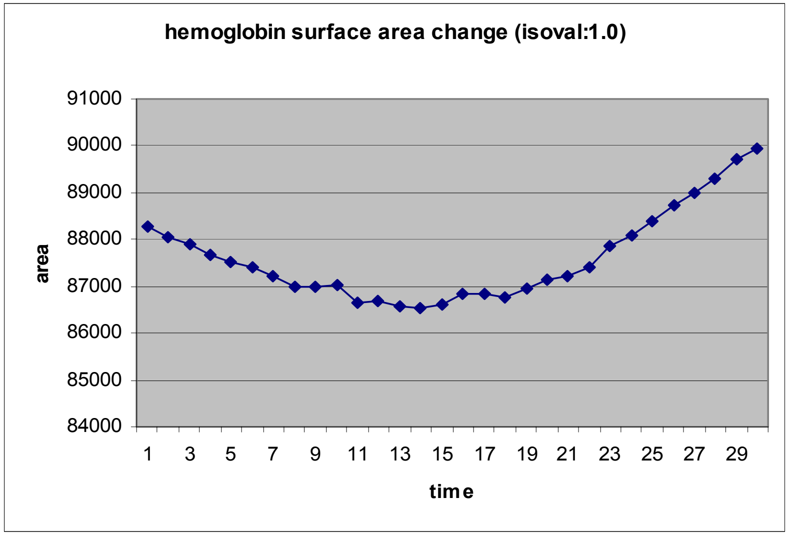


Time = 30

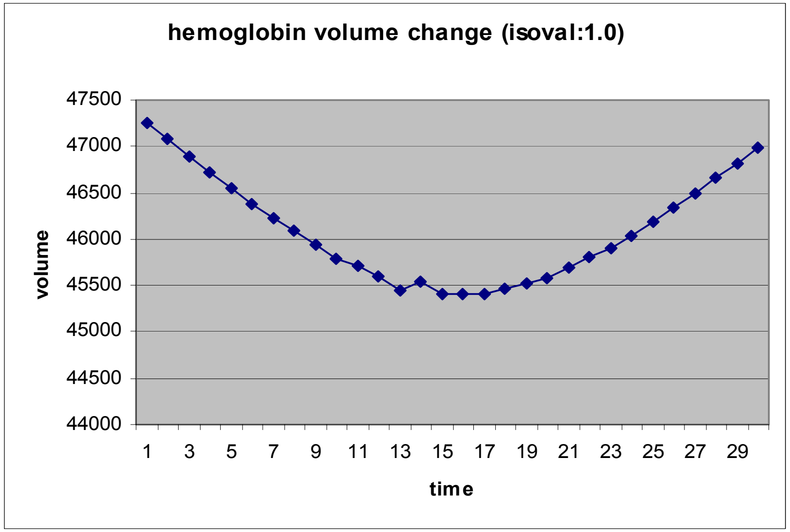


Time-Varying Contour Spectrum

Hemoglobin Surface Area/Volume Change over Time



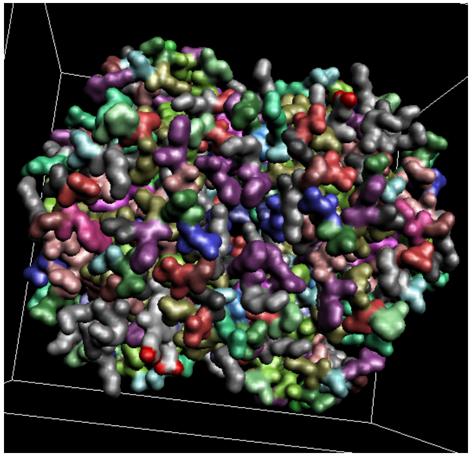
Surface Area



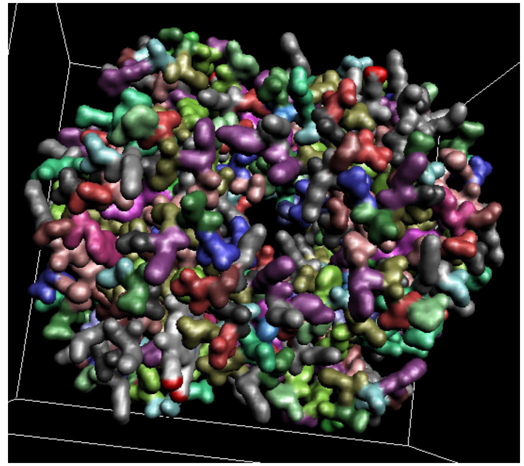
Volume

Surface Area

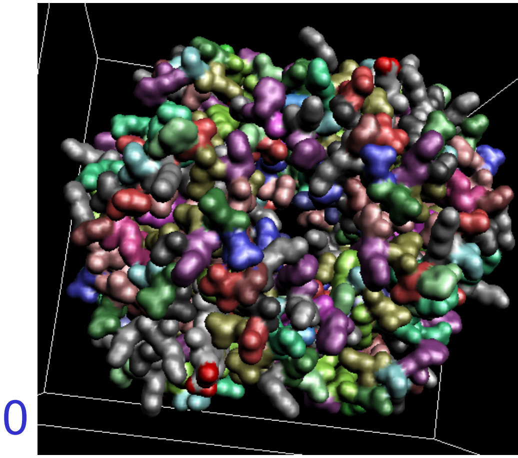
t=1



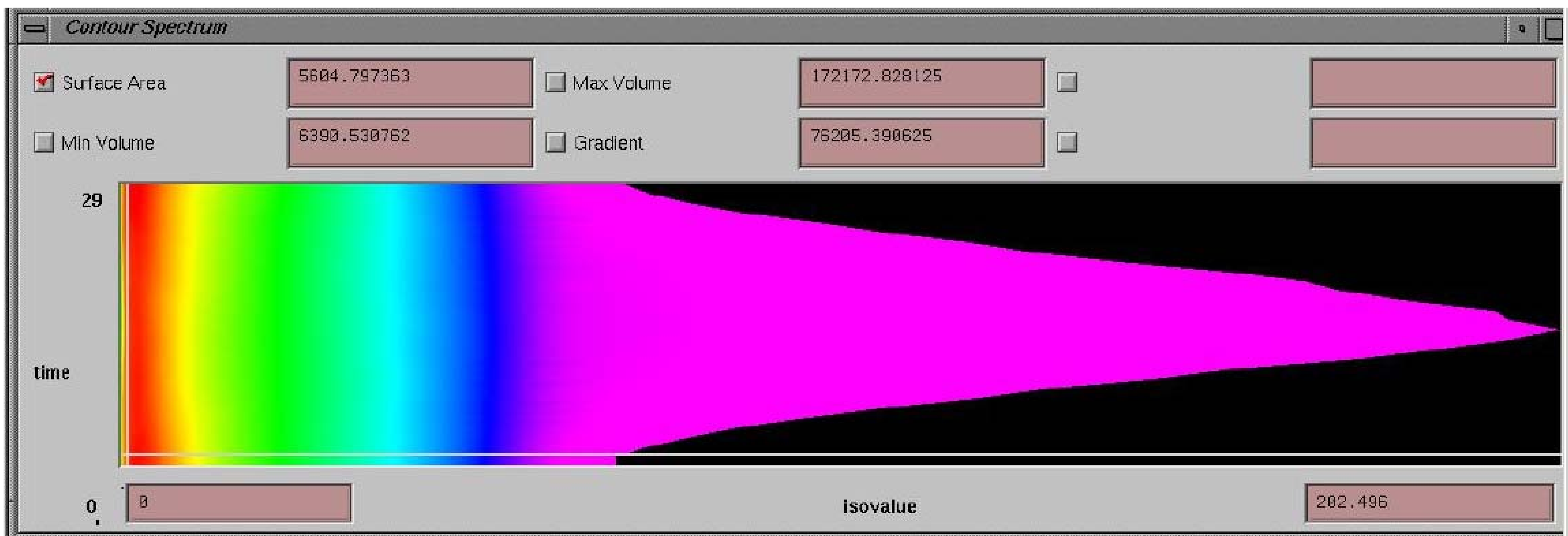
t=15



t=30



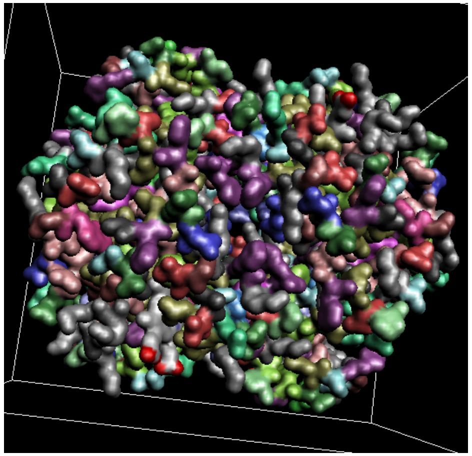
time



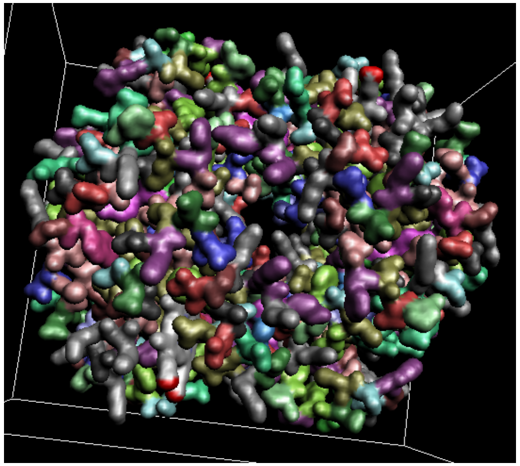
Isovalue 17.44 is selected for isosurface rendering

Volume

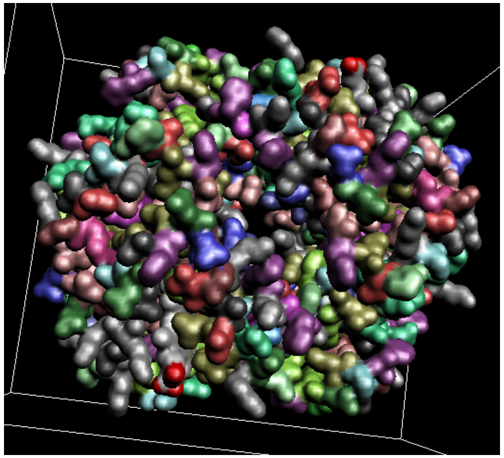
t=1



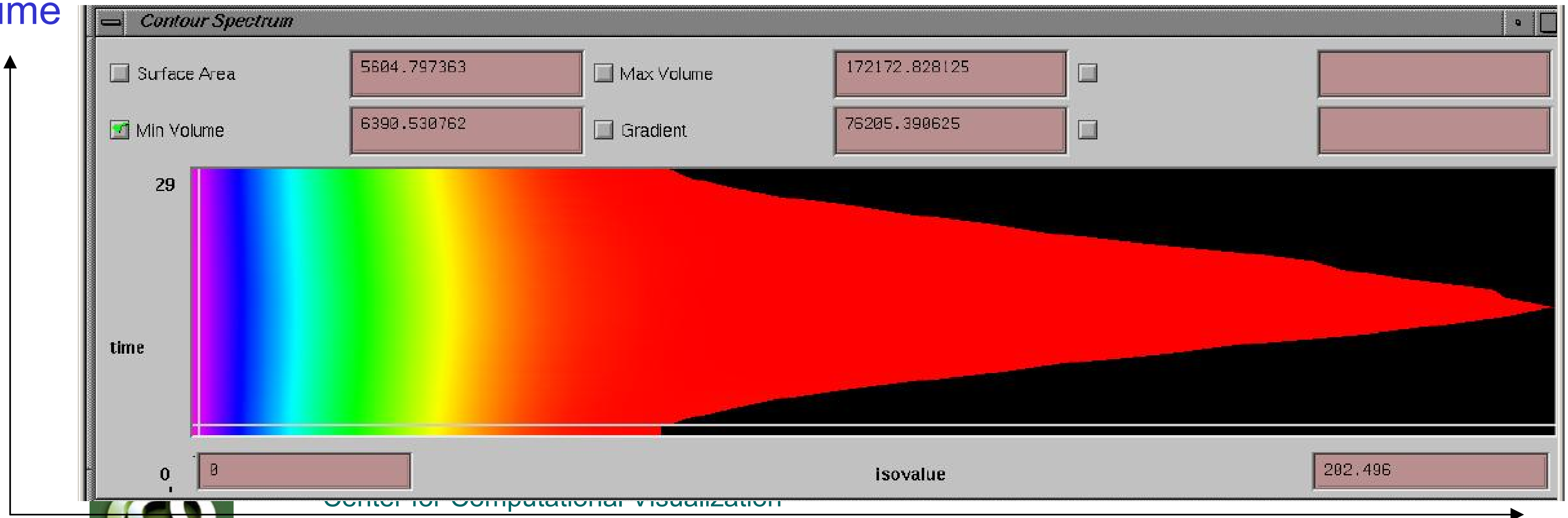
t=15



t=30

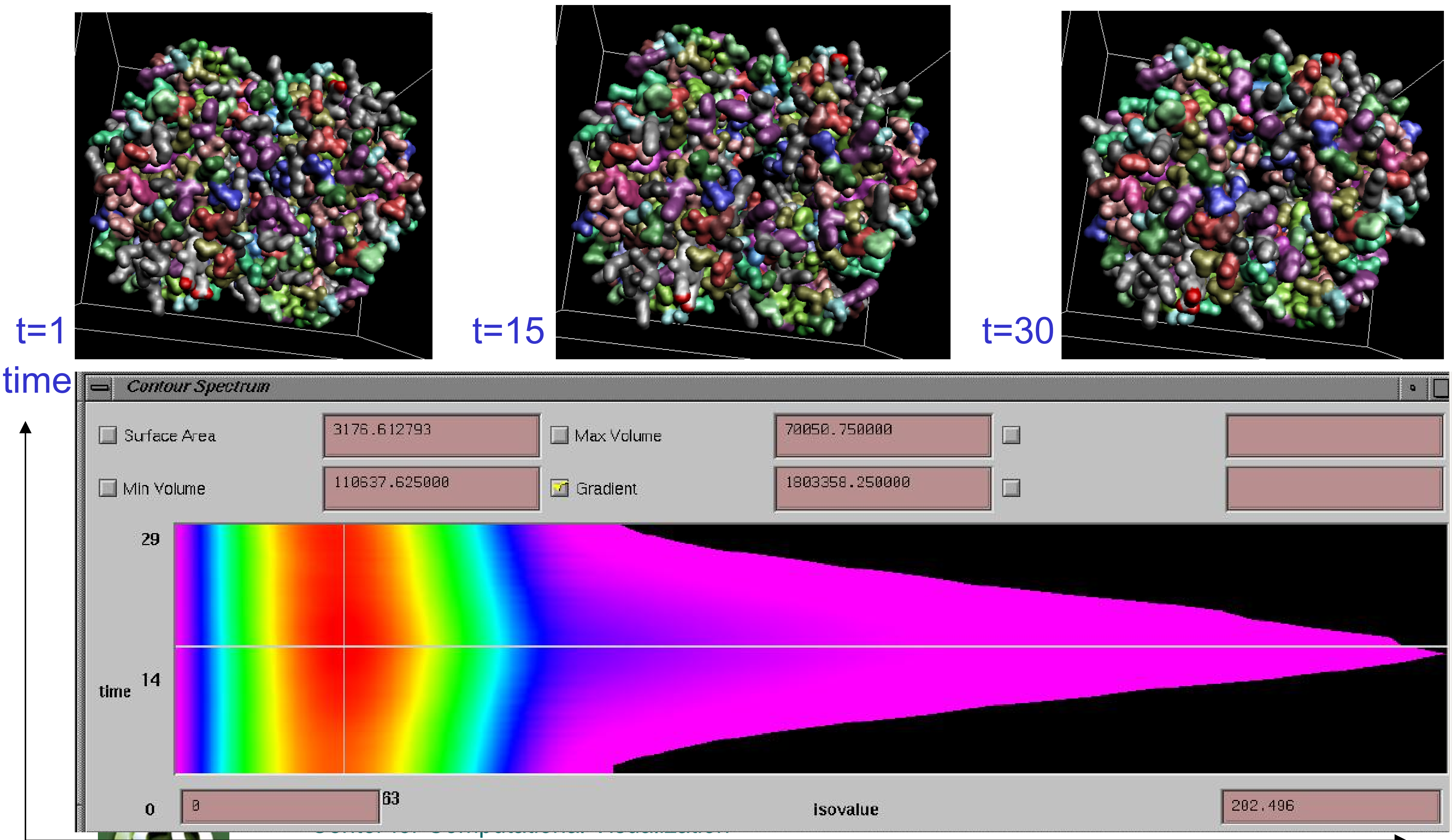


time



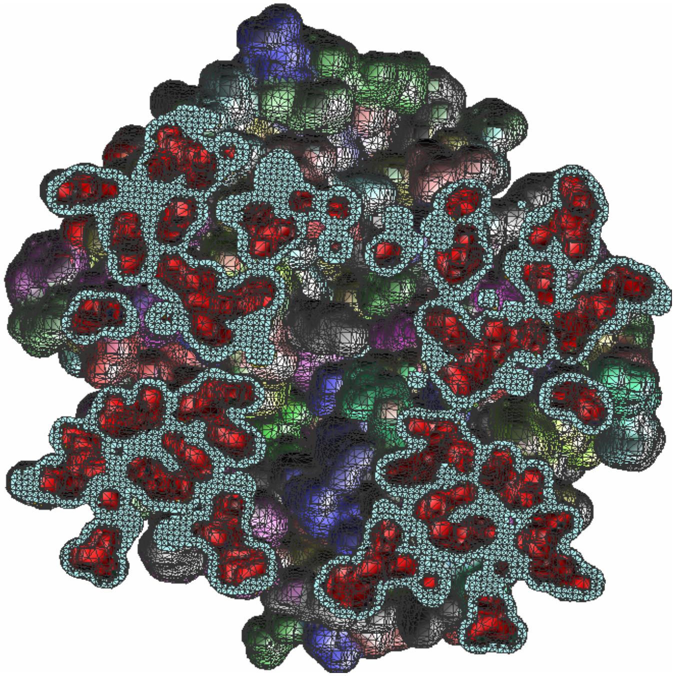
Isovalue 17.44 is selected for isosurface rendering

Gradient Magnitude



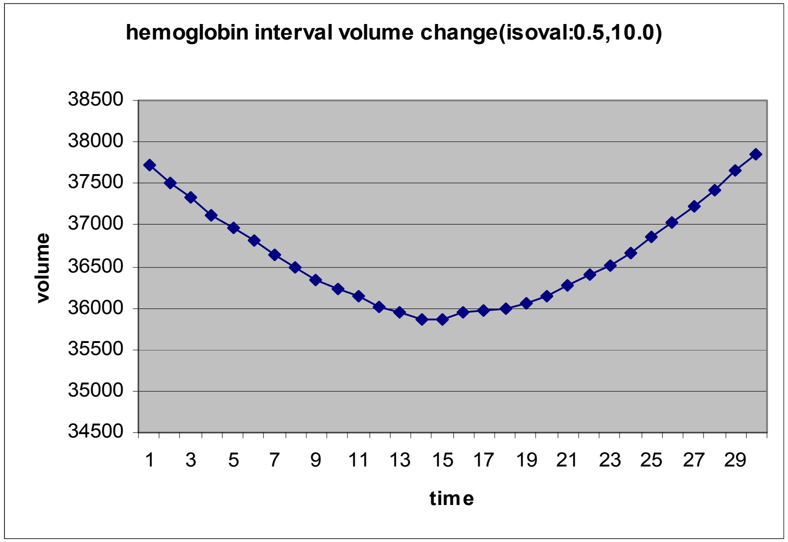
Isovalue 17.44 is selected for isosurface rendering

Time-Varying Contour Spectrum



Interval Volume Crosssection

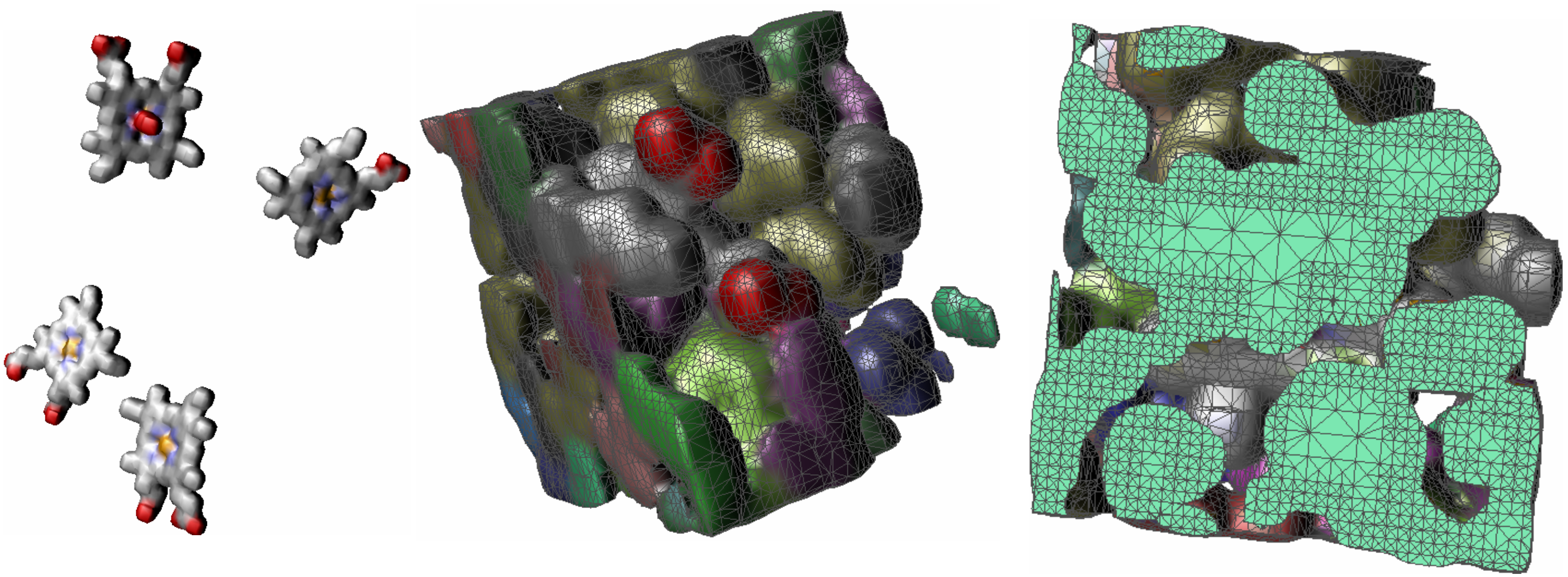
Interval Volume Change over Time



Isovalue 0.5(outer surface) and 10.0(inner surface)

Time-Varying Contour Spectrum

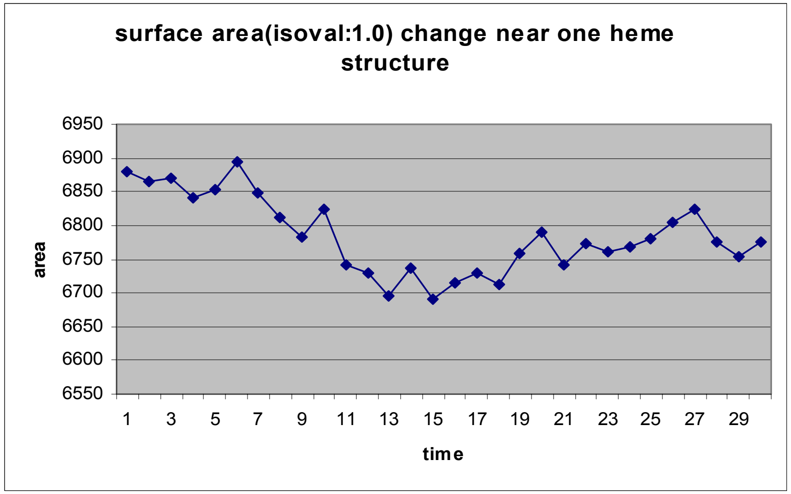
- Quantification around Heme Structure



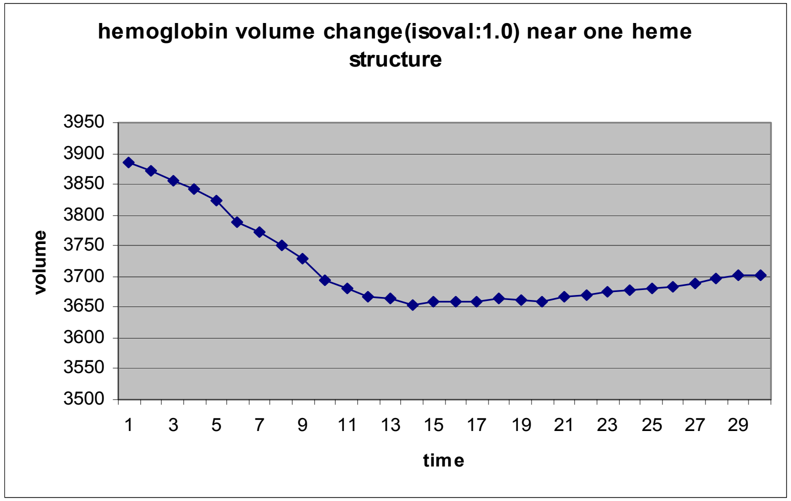
Isovalue 0.5

Time-Varying Contour Spectrum

- Quantification around Heme Structure



Surface Area

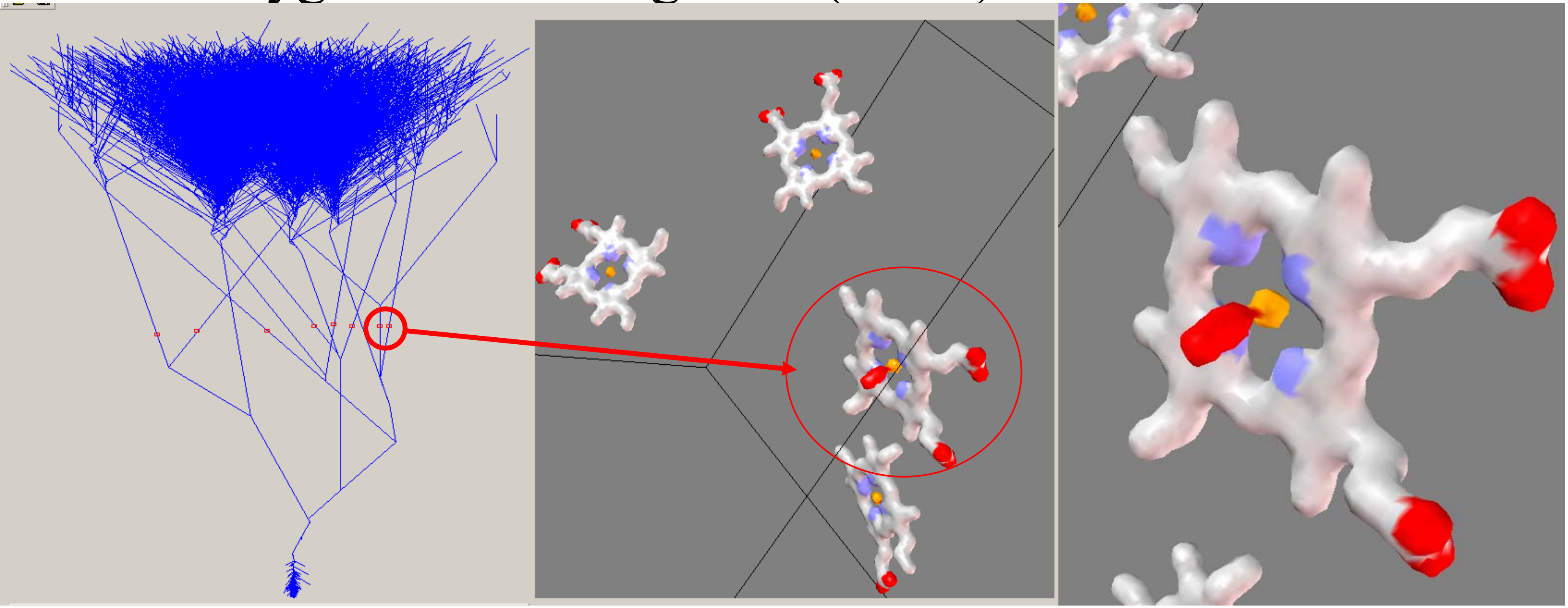


Volume

Isovalue is 1.0

Analysis using the TIME CONTOUR TREE

- Oxygenated Hemoglobin (T=1)

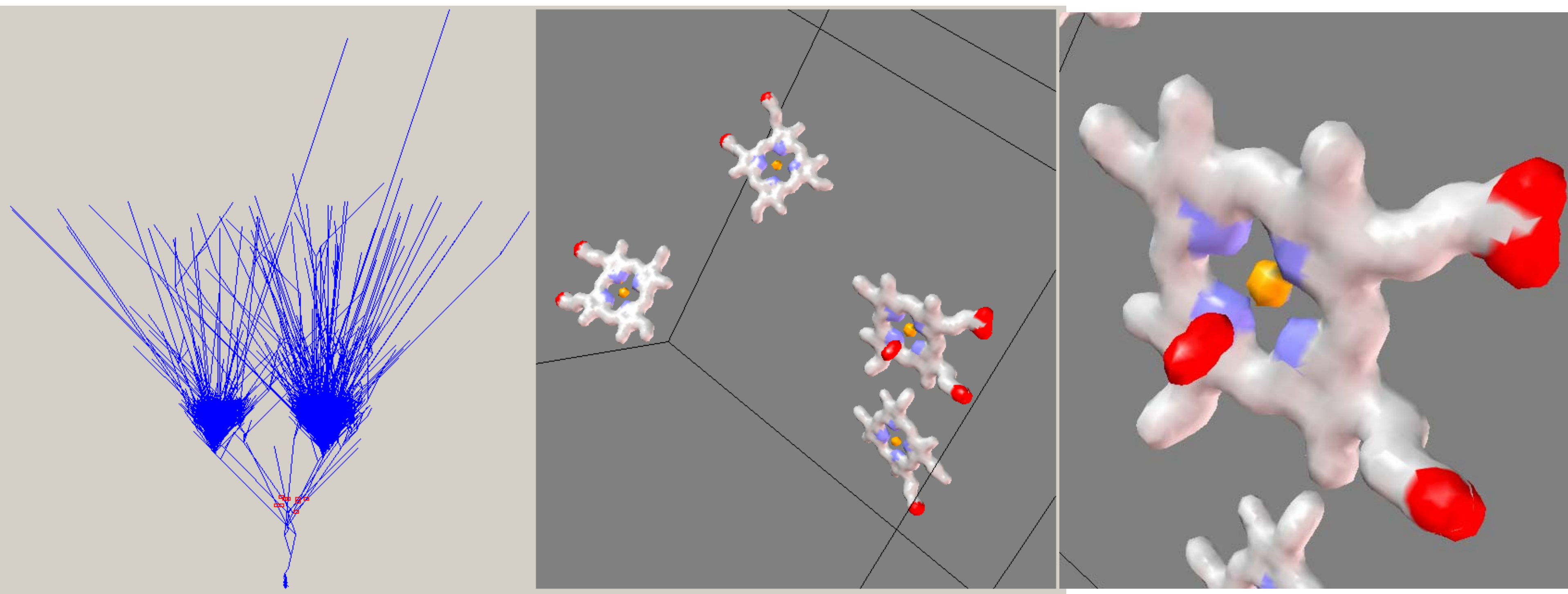


<isovalue = 31>

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

Analysis using the TIME CONTOUR TREE

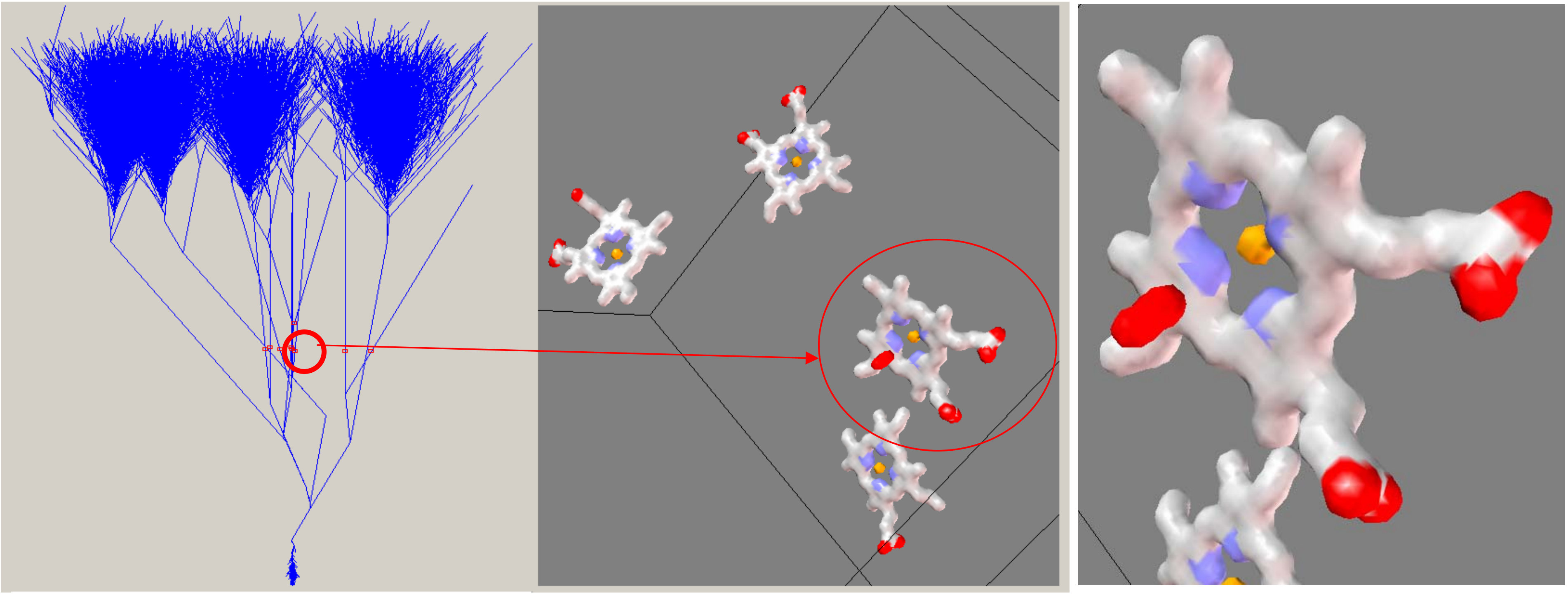
- Intermediate step (T=15)



<isovalue = 31>

Analysis using the TIME CONTOUR TREE

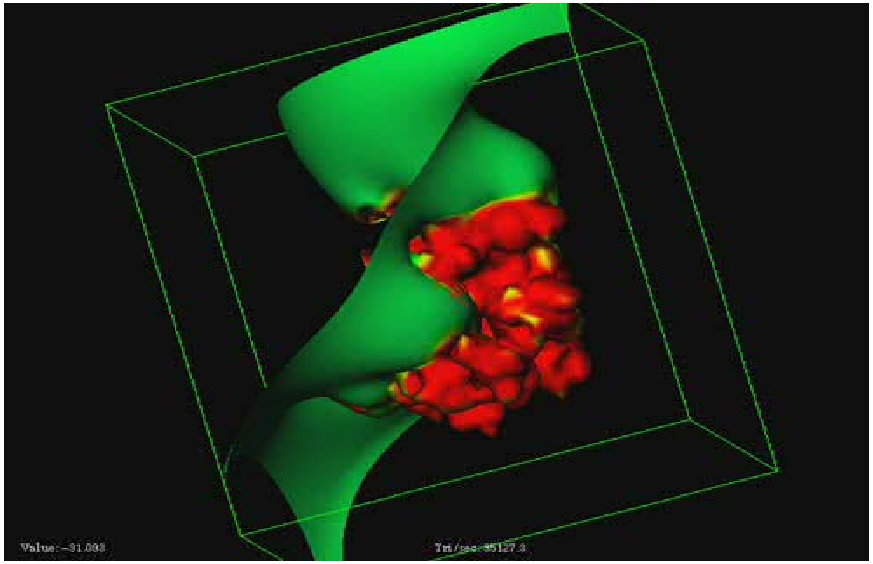
- Deoxygenated Hemoglobin (T=30)

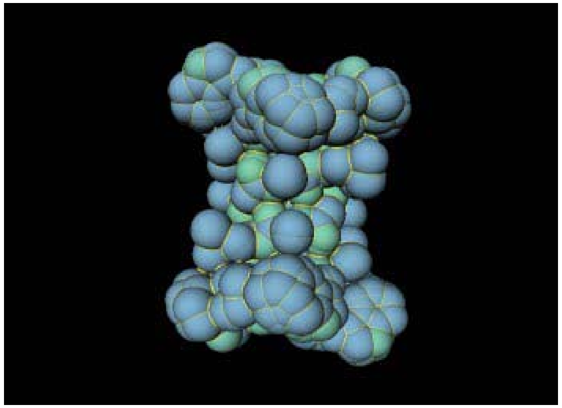


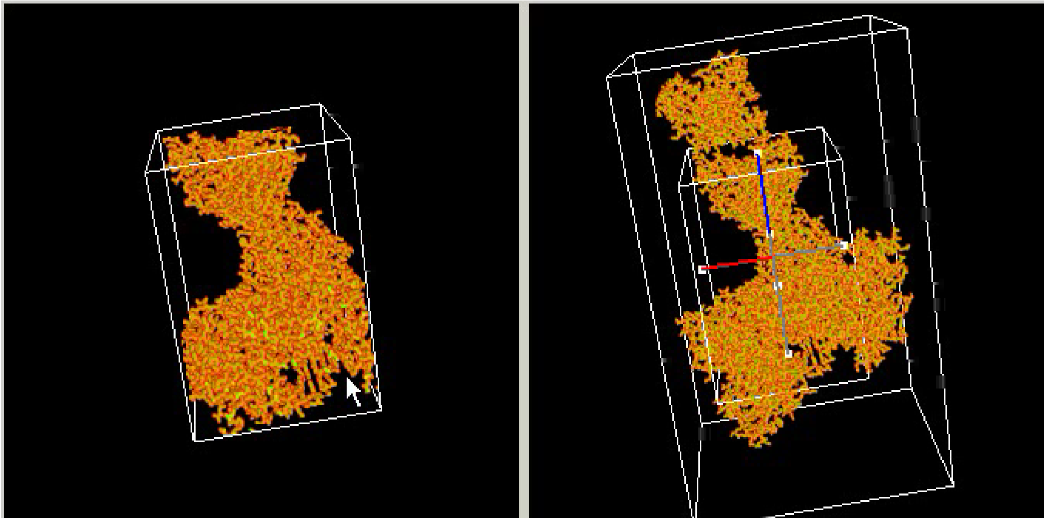
<isovalue = 31>

Modeling, Analysis and Visualization Software
(http://www.ices.utexas.edu/CCV/software/)

- Desktop and Parallel Tools
  - Isocontouring and volume rendering software on COTS
  - Multi-Display Clients using programmable graphics hardware
  - Integration with the Grid Underway for Remote Visualization Services







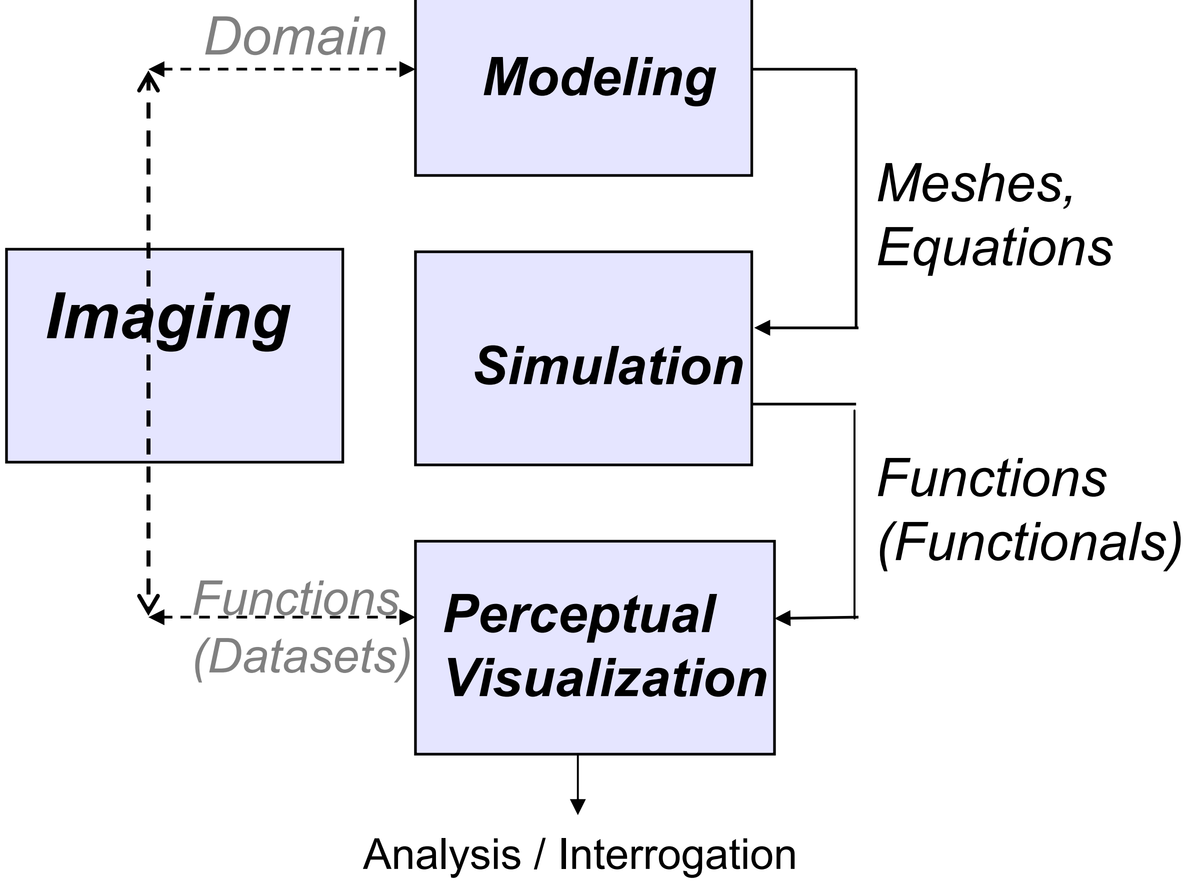
Whats in the Future ?

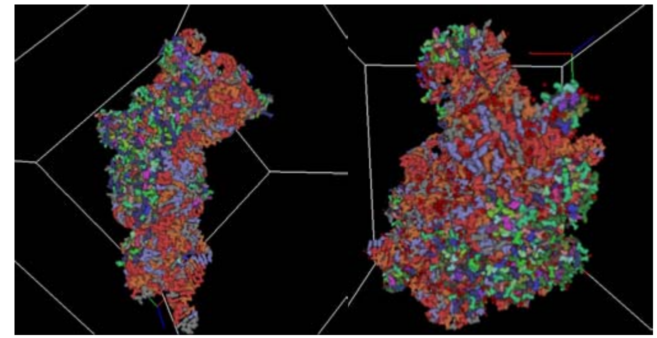
- Computational Modeling for Nano-Machines and Nano-Medicine
  - Psuedo-atomic model generation for bio-molecular machines, and their assemblage *properties*
  - Mechanisms for capturing knowledge of macromolecular *flexibility* and inferring *functionality*
  - Understanding *interactions* between molecular assemblies, biological and synthetic through biochemistry/biophysics simulations

Acknowledgements

- **CCV**
  - **Julio Castrillon**
  - **Peter Djeu**
  - **SK Vinay**
  - **Zeyun Yu**
  - **Bong-Soo Sohn**
  - **Young-In Shin**
  - **Sangmin Park**
  - **Jessica Zhang**
  - **Greg Johnson**
  - **Zaiqing Xu**
  - **KL Chandrasekhar**
  - **Qiu Wu**
  - **Jasun Sun**
  - **Anthony Thane**
  - **Shashank Khandelwal**
- Computational resources
  - CCV/ICES/UT
  - NPACI/SDSC
- Sponsors
  - NSF
  - UT/MDACC/Whitaker
  - NPACI/NSF
  - DOE-LLNL/Sandia

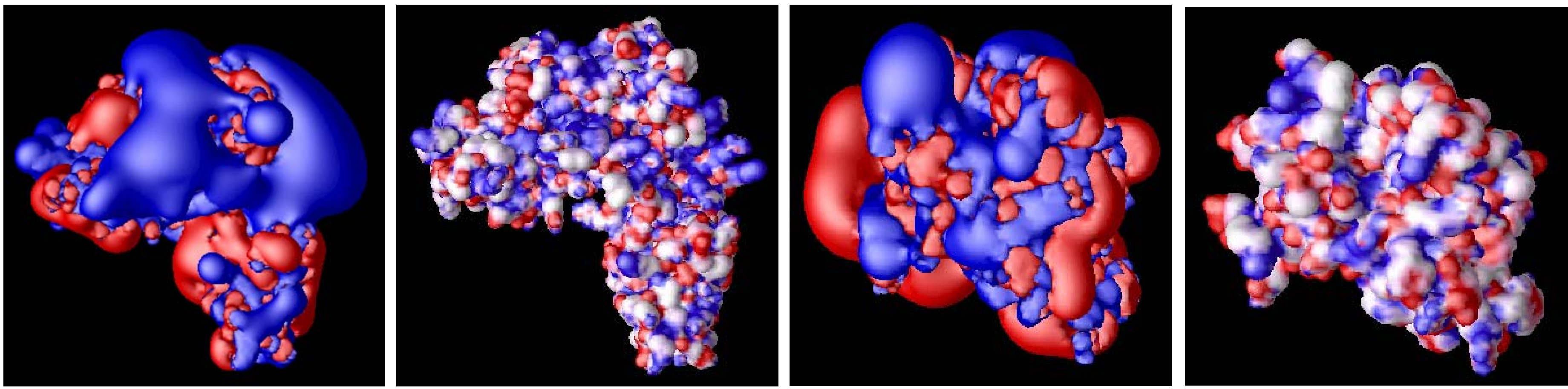
Computational Visualization





- To identify and perceive *information* for model calibration or scientific discovery
- Model *Analysis, Visualization* and *interrogation* with maximum *fidelity*

Molecular Electrostatics Visualization



Blue positive

White neutral

Red negative

Isosurfaces of Electrostatics potential, and rendered as a Function on an Isosurface of Electron Density

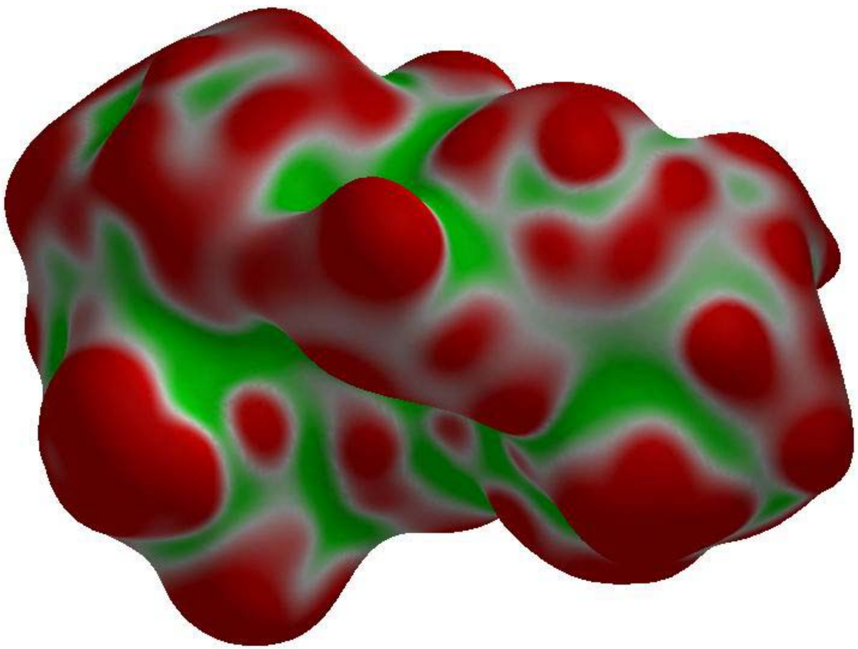
Surface curvatures

The two main surface curvatures in differential geometry are the ***Mean Curvature*** H and the ***Gaussian curvature*** K.

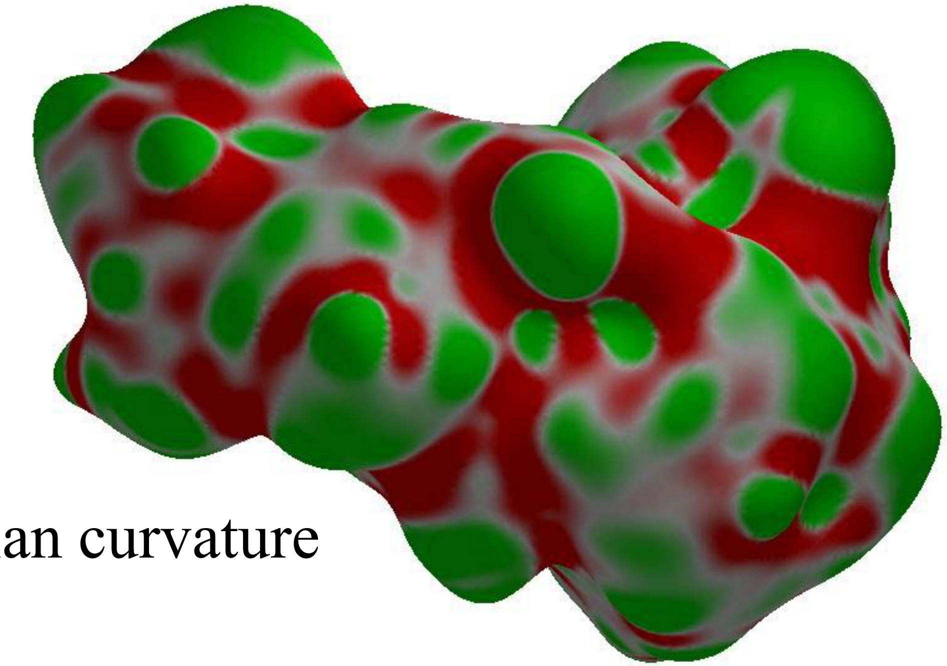
Let k_{min} and k_{max} be the minimum and maximum curvatures at a point. Then,

$\mathbf{H} = \frac{1}{2}(k_{min} + k_{max})$ and
$\mathbf{K} = k_{min}k_{max}$

Protein Kinase from Rat (1a06)

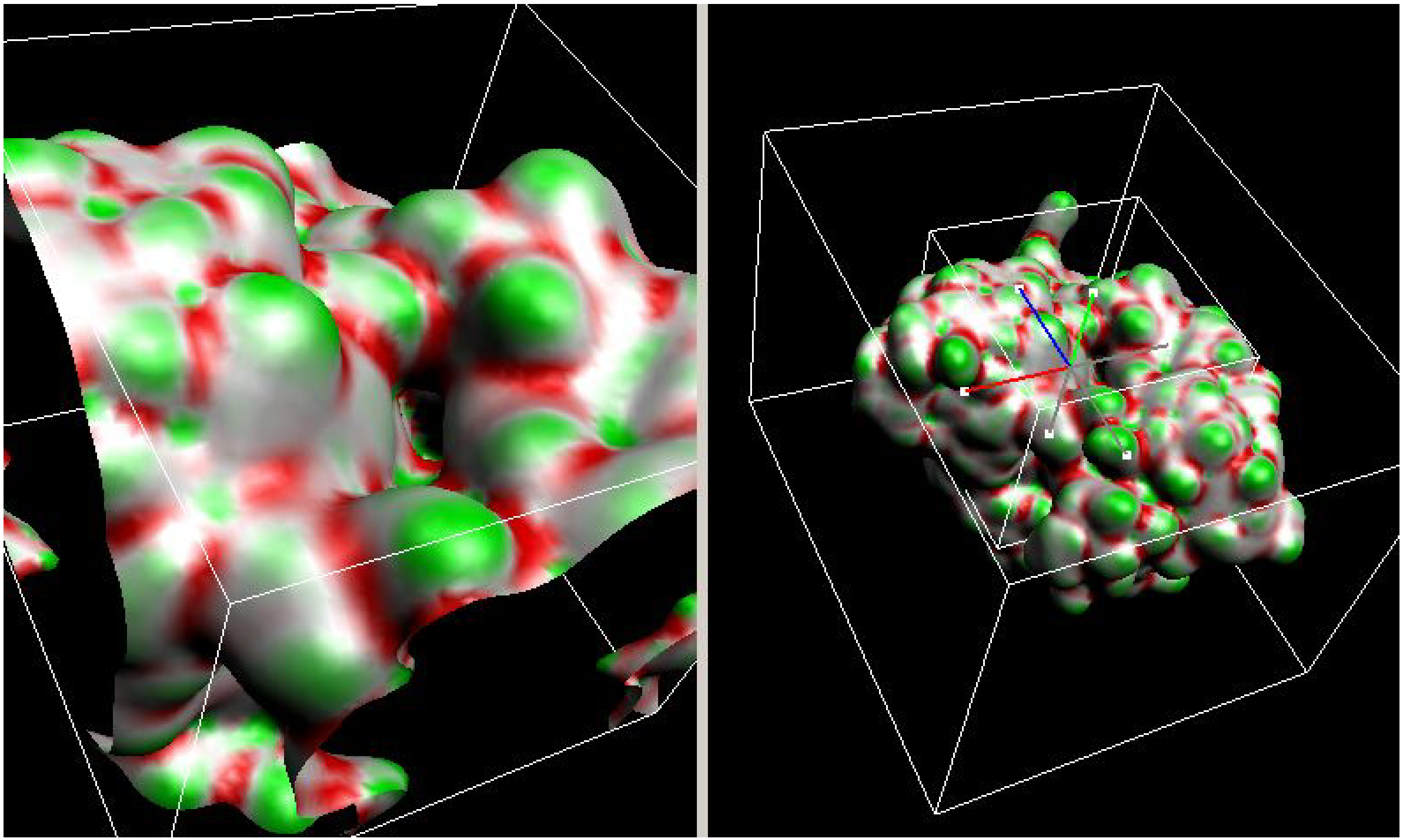


Mean curvature

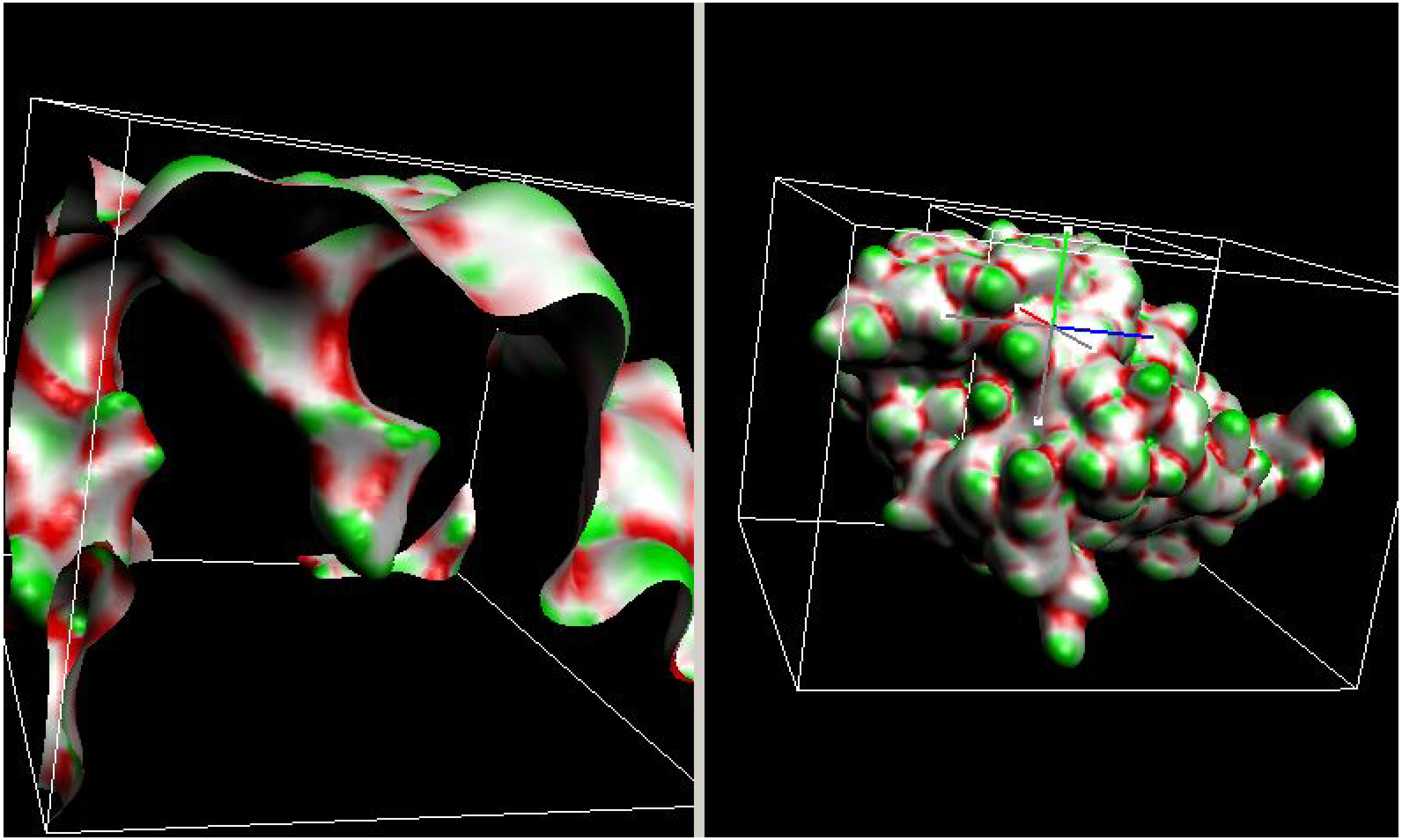


Gaussian curvature

Gaussian curvatures on Mouse AcetylCholinesterase

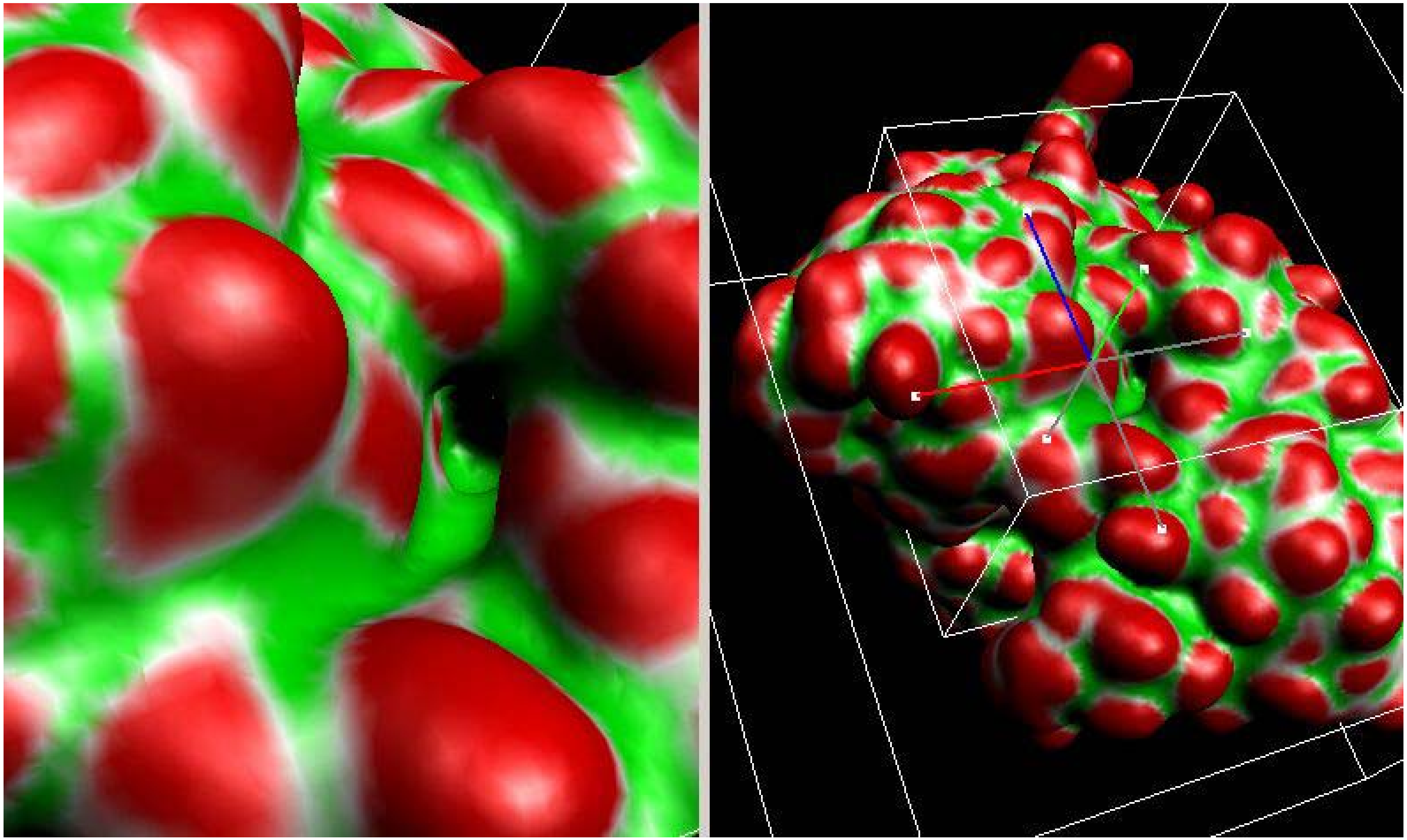


Mouse AcetylCholinesterase



Gaussian curvatures

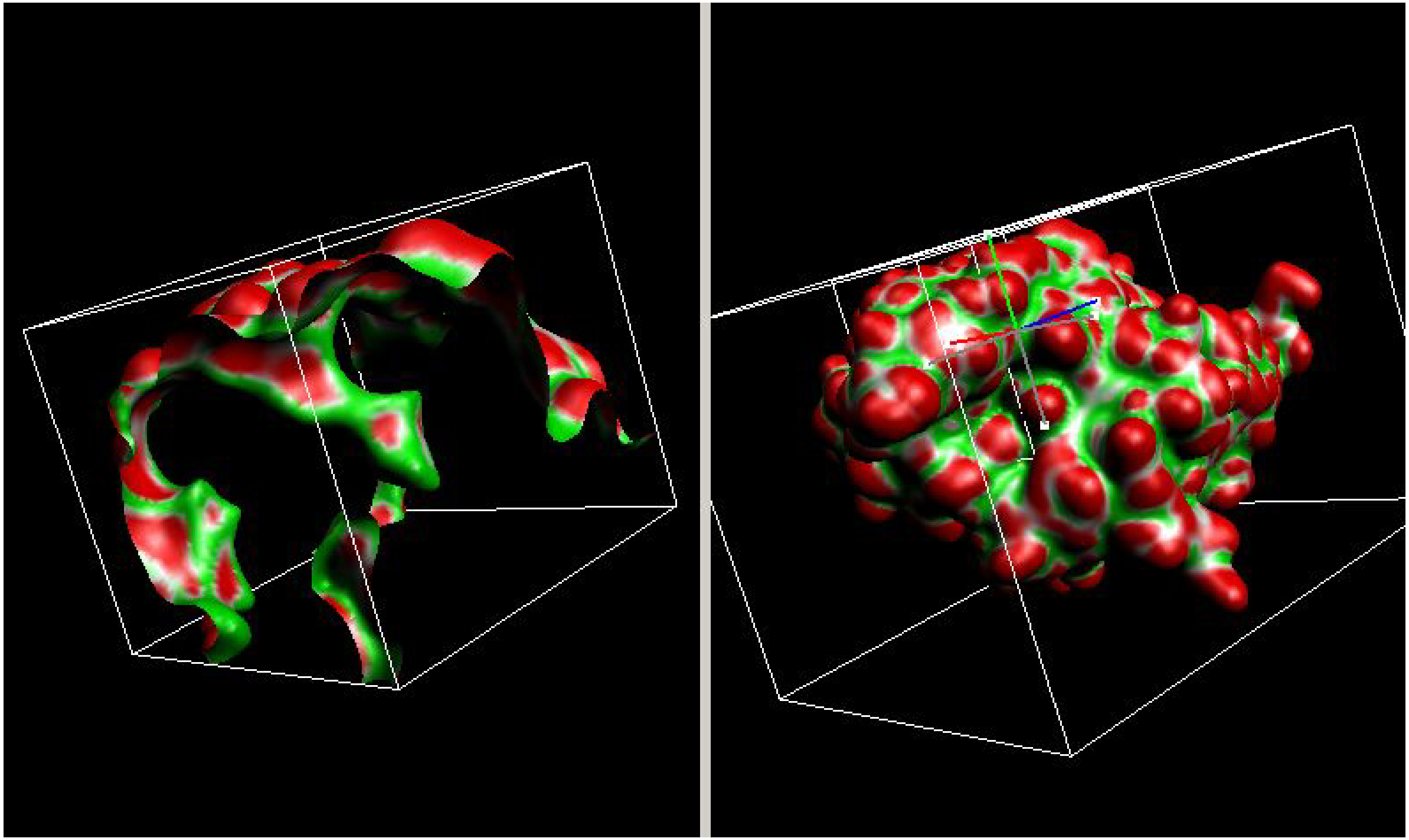
Mouse AcetylCholinesterase



Mean curvatures

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

Mouse AcetylCholinesterase



Mean curvatures