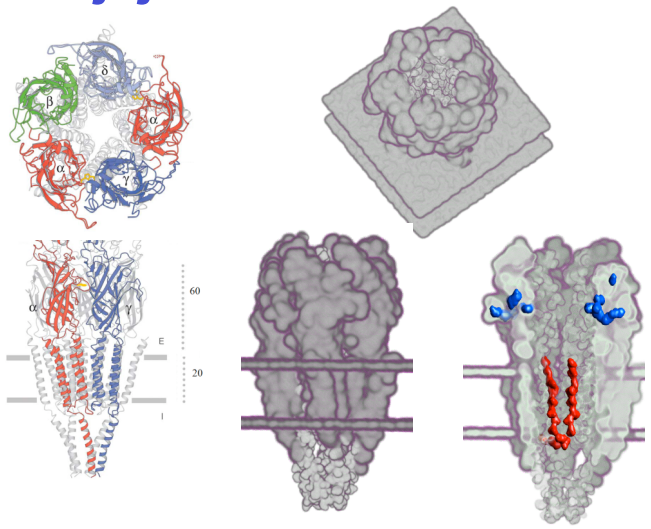
Lecture 4: Geometric Modeling and Visualization

Molecular Structures (Models) from PDB, VIPER

Chandrajit Bajaj

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

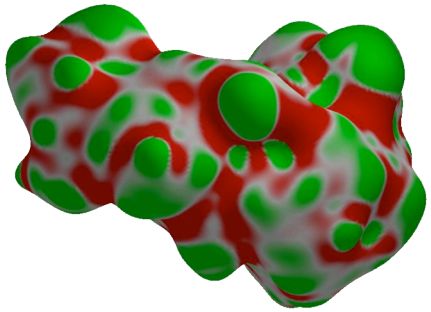


Center for Computational Visualization
Institute of Computational and Engineering Sciences
Department of Computer Sciences
University of Texas at Austin

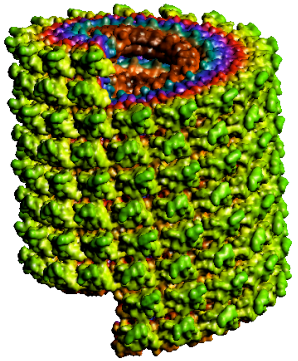
September 2007

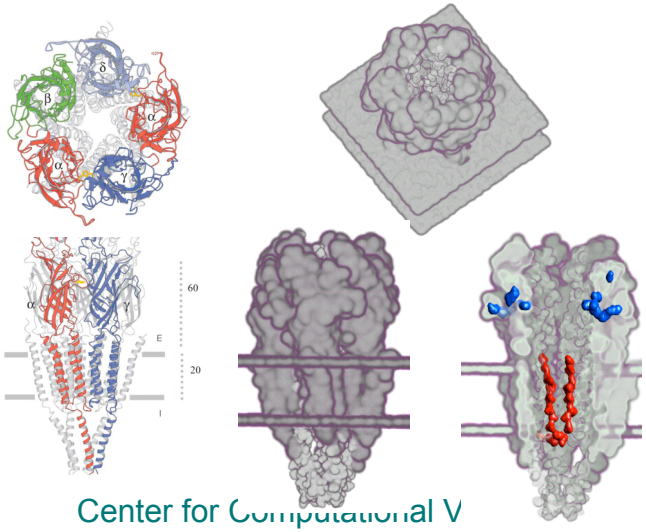
Tools for 3D Molecular Structure Determination

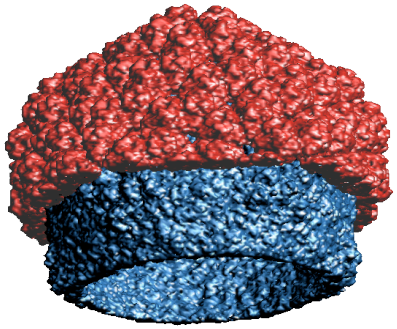
- X-ray crystallography (diffraction)
  - Atomic resolution
  - Difficulties (experimental, computational)



- Nuclear magnetic resonance (NMR) (spectroscopy)
  - Atomic resolution
  - Limited to small structures

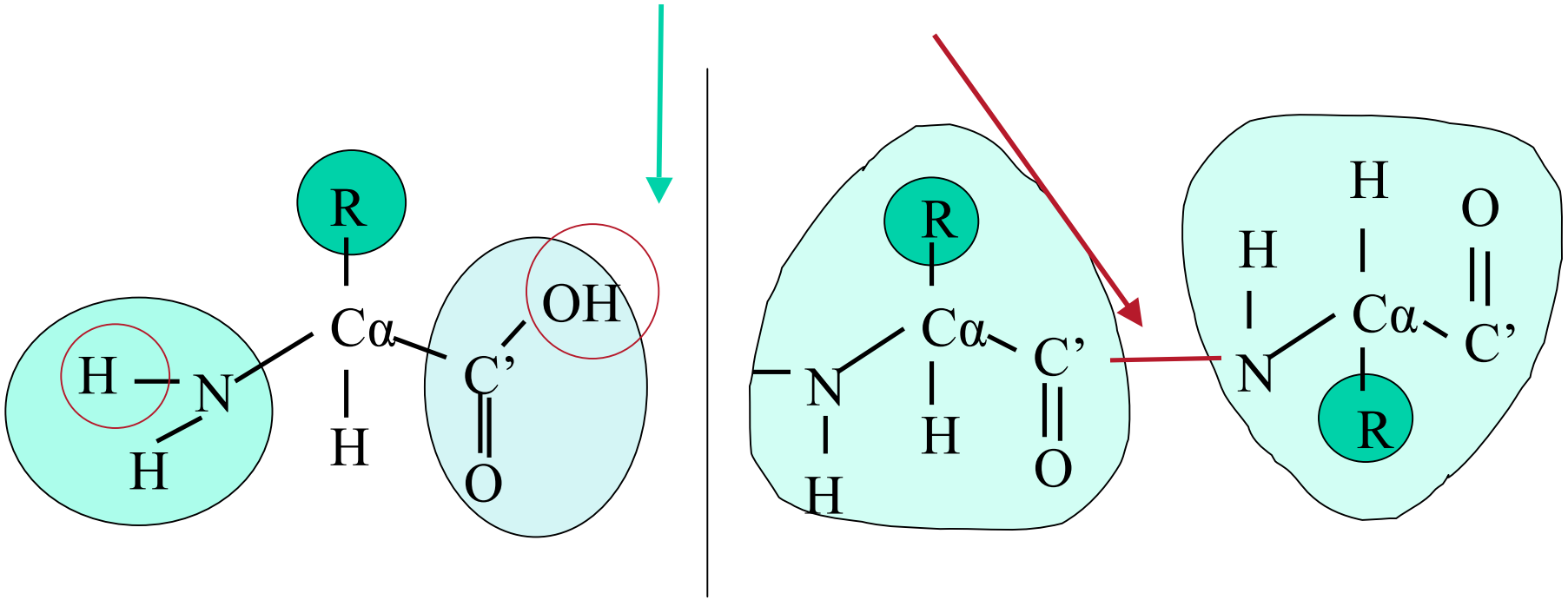




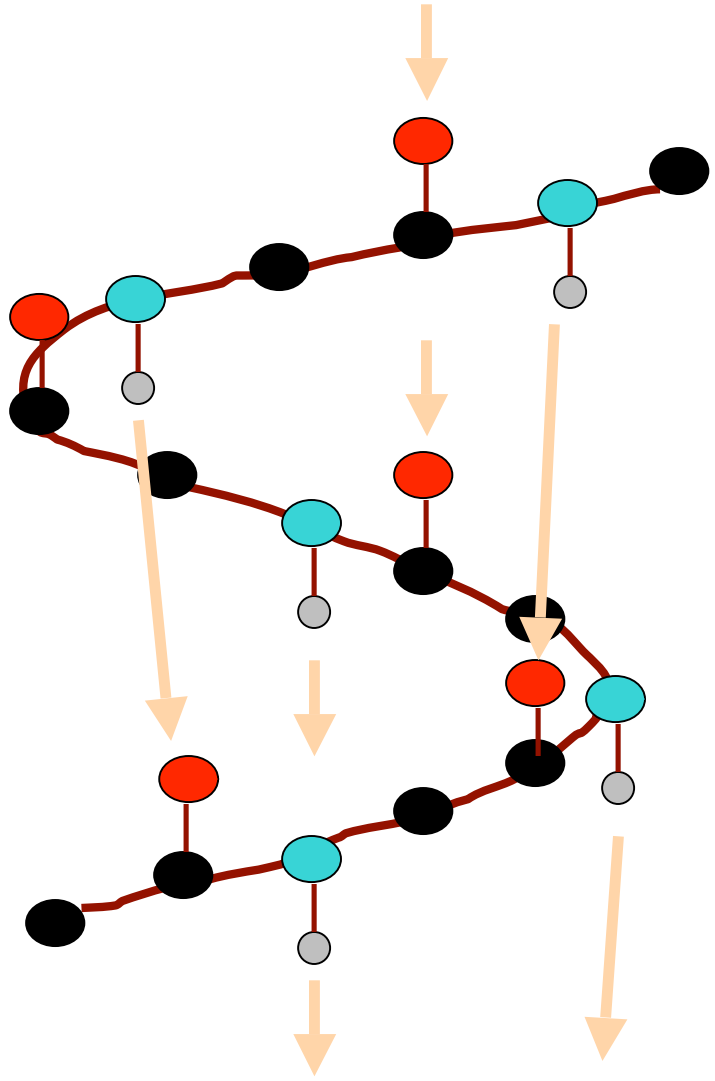


Proteins

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



α helix



Hydrogen bond

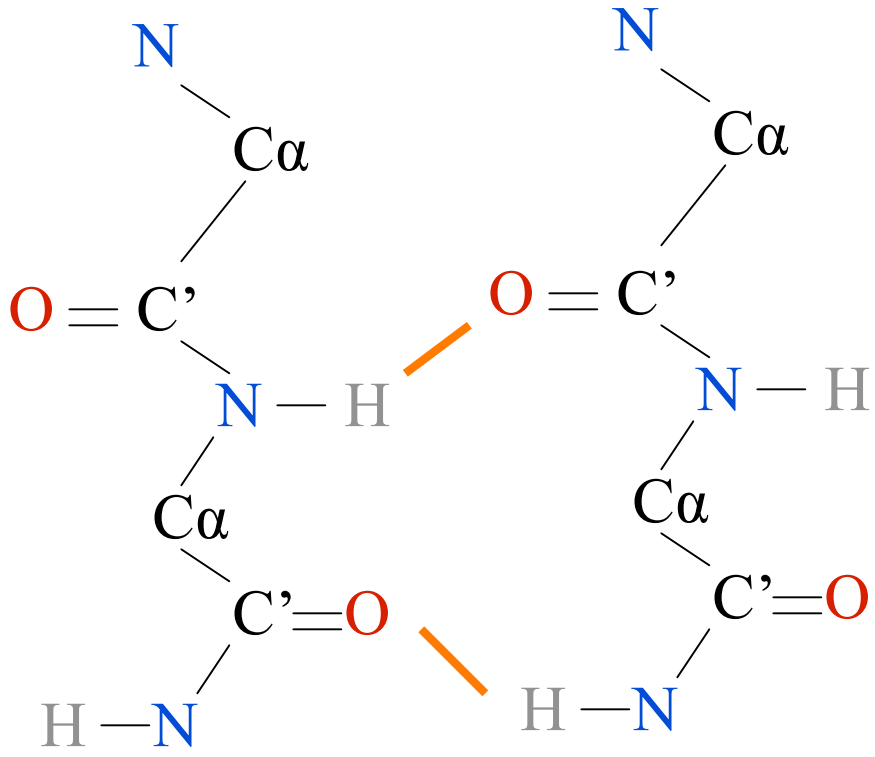
Usually 4-5 to 40 residues

Rise per residue = 1.5A

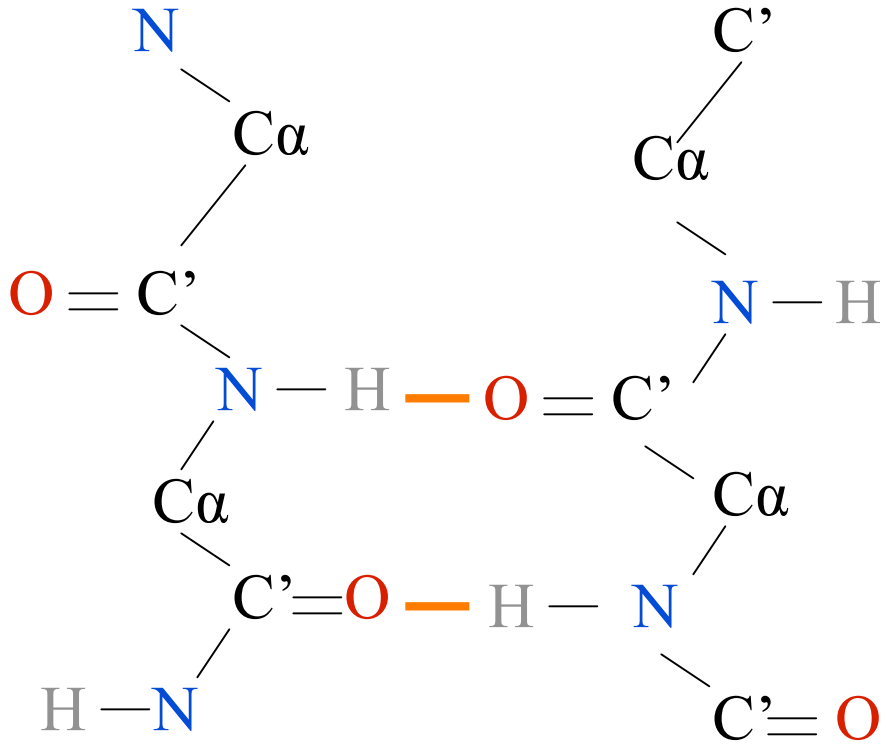
Residues project outwards from the axis

Hydrogen bonds cause a dipole moment, approx 0.5-0.7 unit charge on each end

Beta sheets



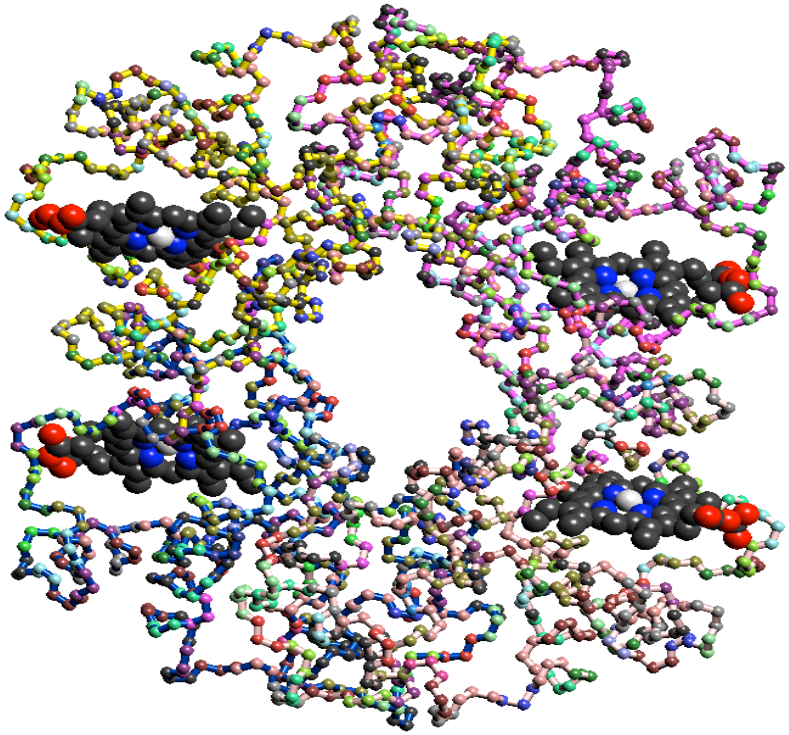
Parallel beta sheets

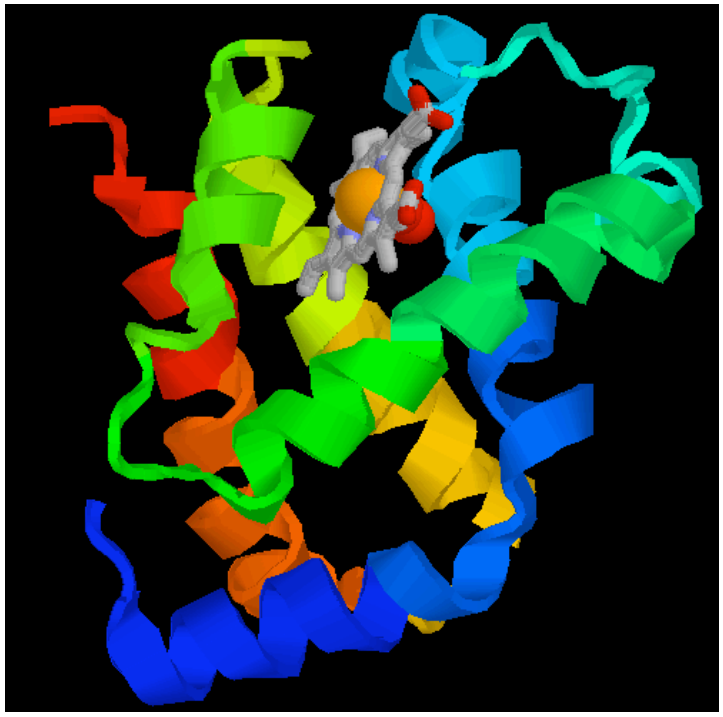


Anti Parallel beta sheets

Structure of Hemoglobin

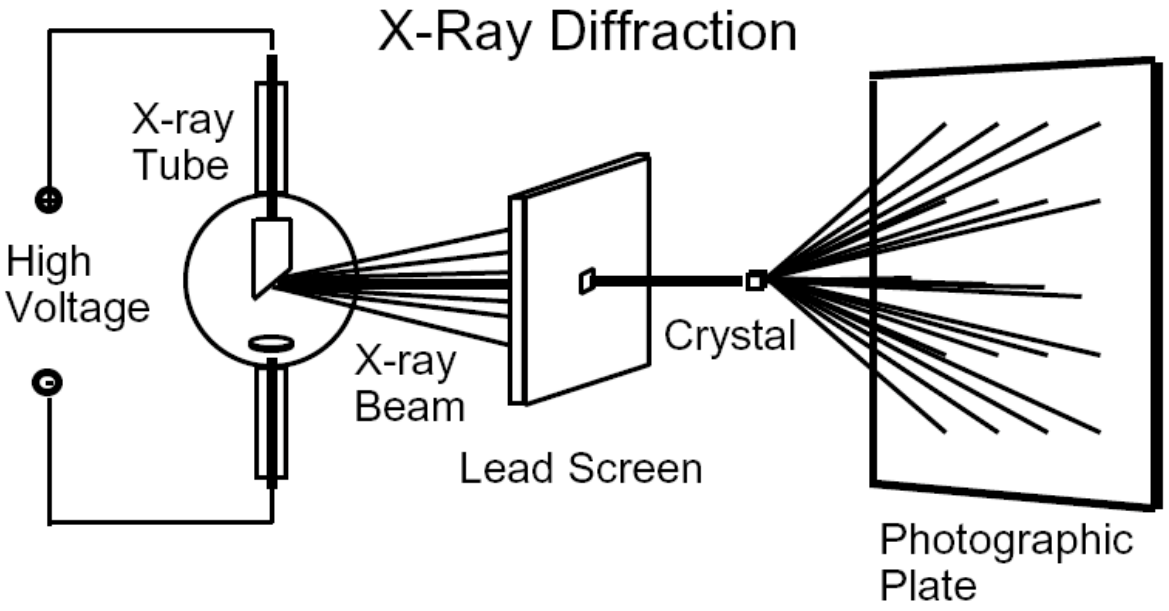
- secondary, tertiary, quaternary structure

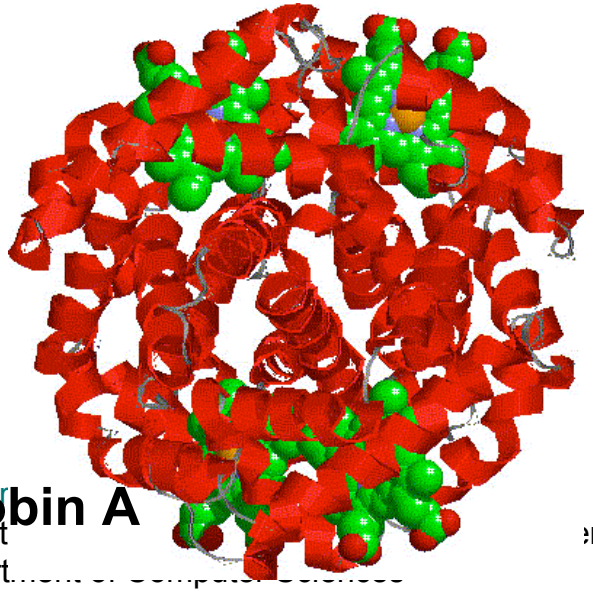




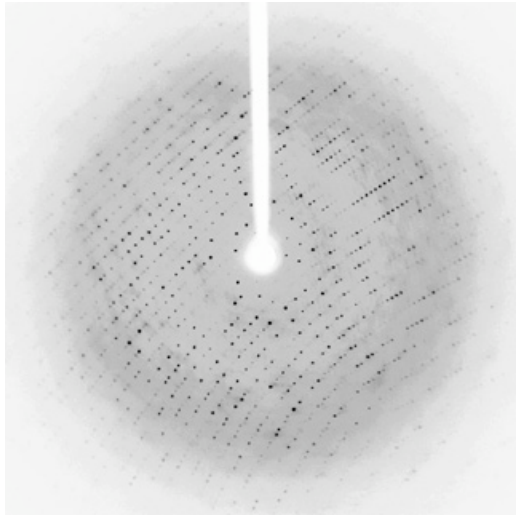
- One ` chain contains eight ä helices and no -sheets.

X-ray diffraction





human hemoglobin A



Hard-Sphere Model

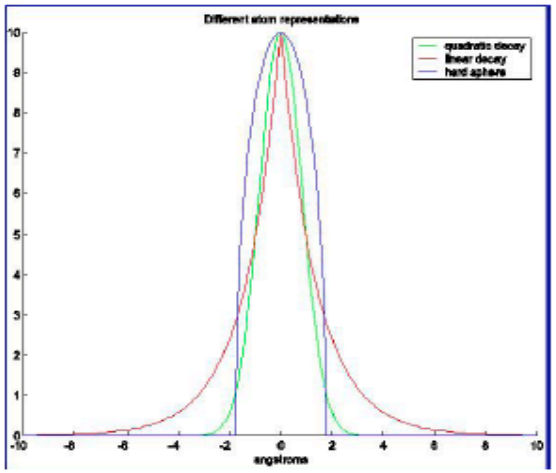
Atom: Electron Ball

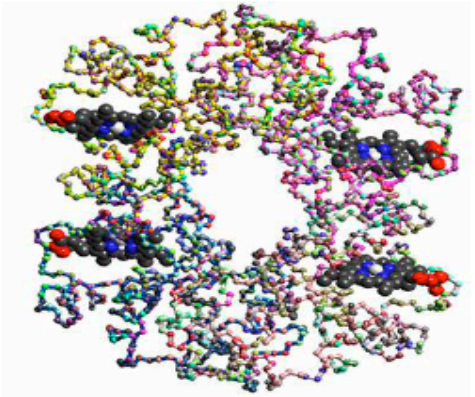
$$u^{hs}(\mathbf{x}) = \begin{cases} 1 & \text{if } |\mathbf{x} - \mathbf{x}_c| \leq r, \\ 0 & \text{otherwise;} \end{cases}$$

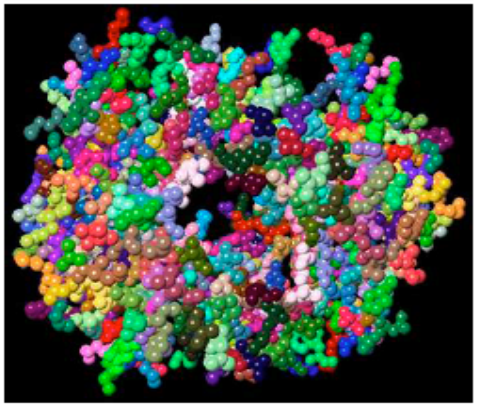
where, r is the atomic van der Waals or solvation radius, and $\mathbf{x}_c$ is the center of the atom.

Molecule: Union of Electron Balls

vdW Surface

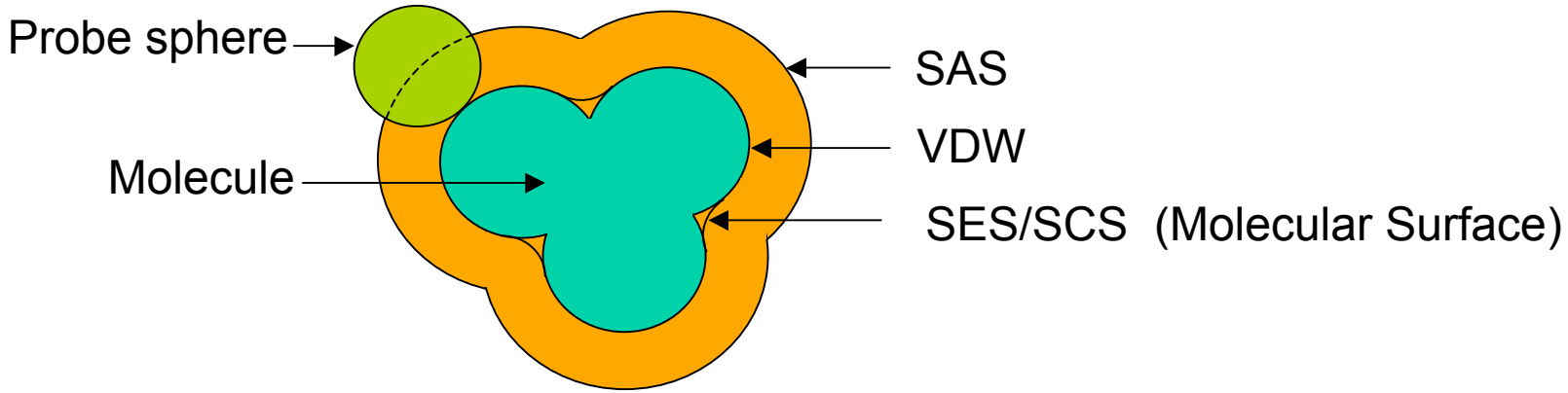






Molecules in Solvent (Implict Model)

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

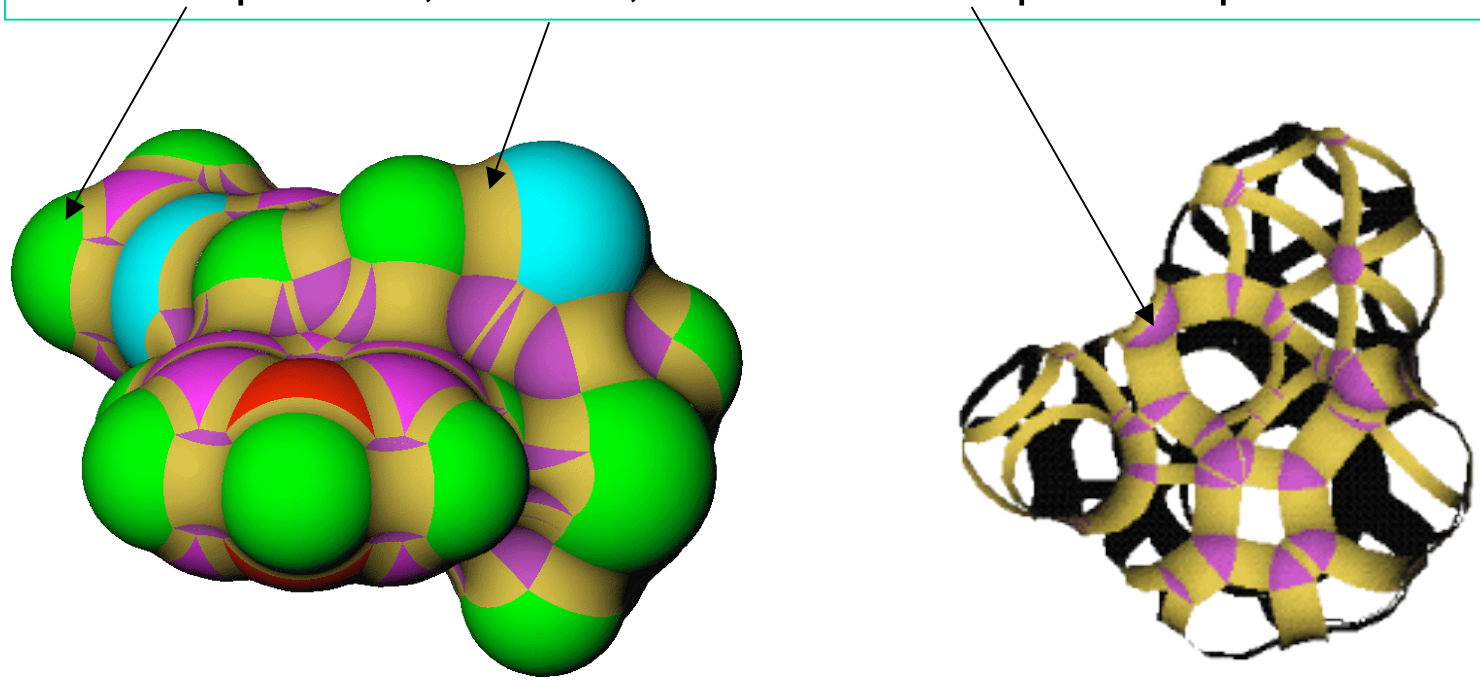


SAS: solvent accessible surface: locus of probe center
VDW: van der Waals surface: Union of spheres with VDW radii
SES/SCS: solvent excluded/contact surfaces

Implicit Solvation Surface for the Hard Sphere Model

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

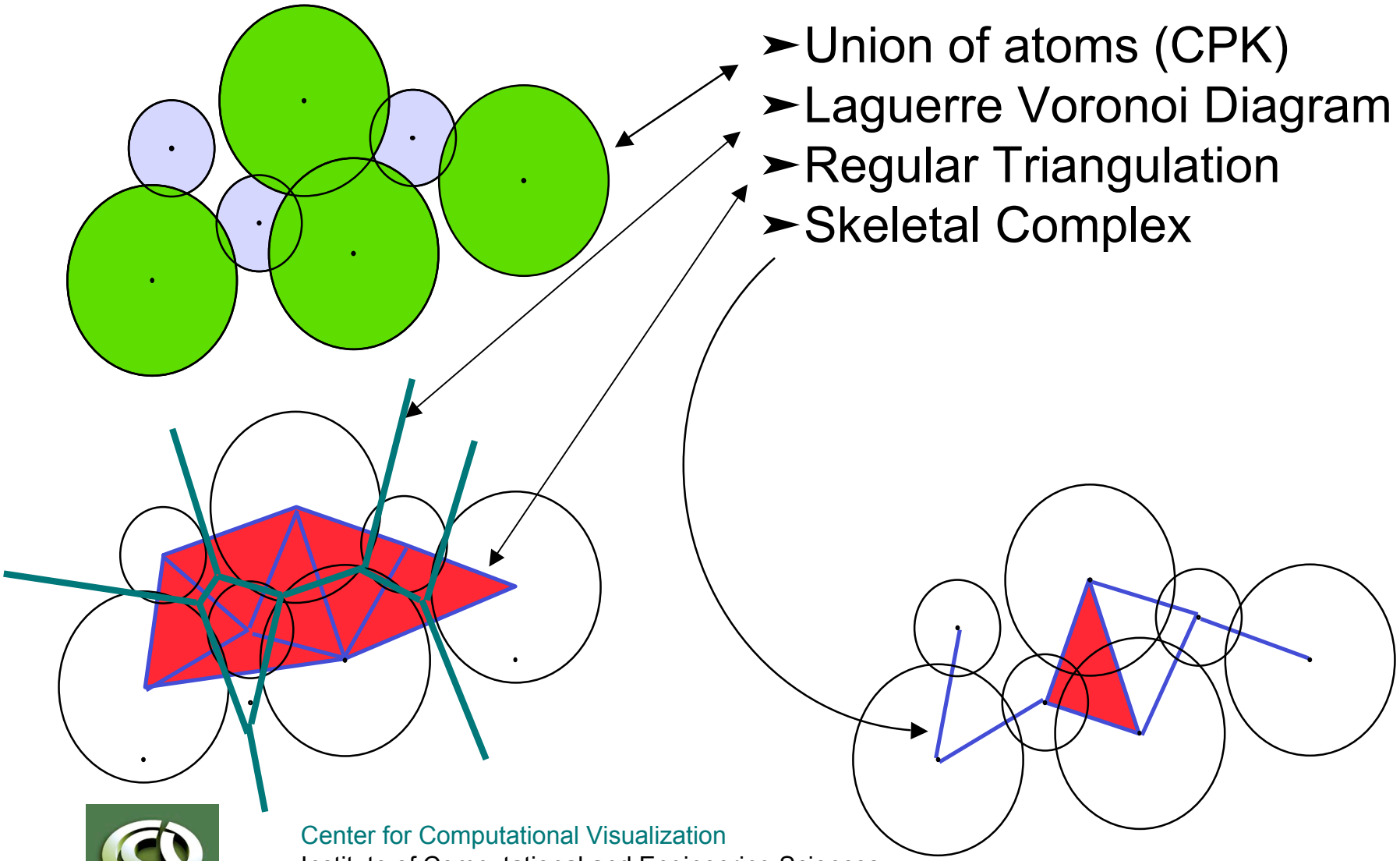
convex spherical, toroidal, and concave spherical patches



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

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

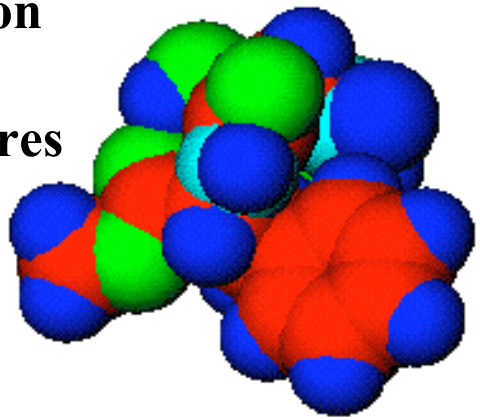
Molecules as Union of Hard Spheres

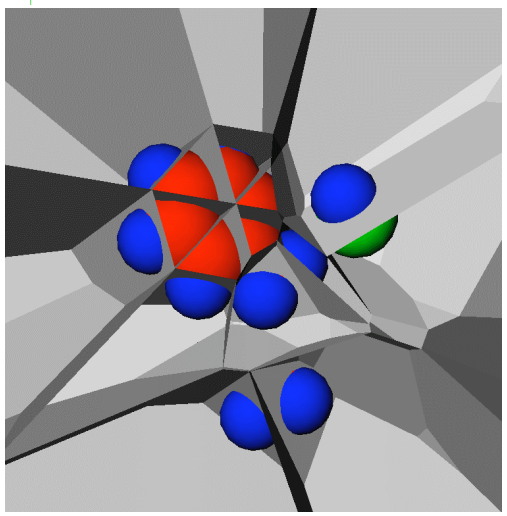


- Union of atoms (CPK)
- Laguerre Voronoi Diagram
- Regular Triangulation
- Skeletal Complex

Molecular Solvation (Nutrasweet)

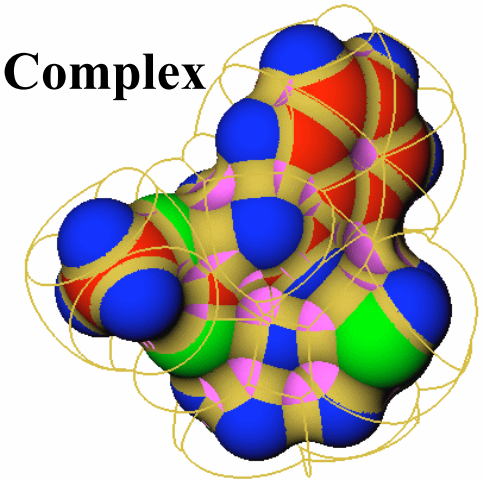
Union of Spheres

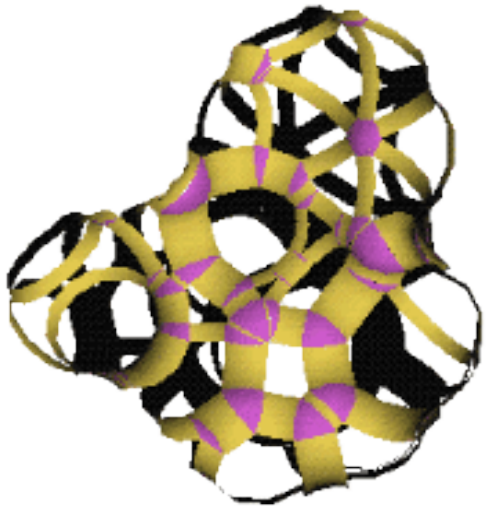




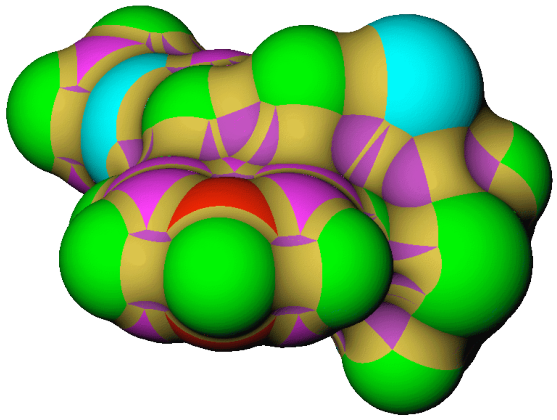
Laguerre Voronoi Diagram of Union

Solvent Offset Curves Complex



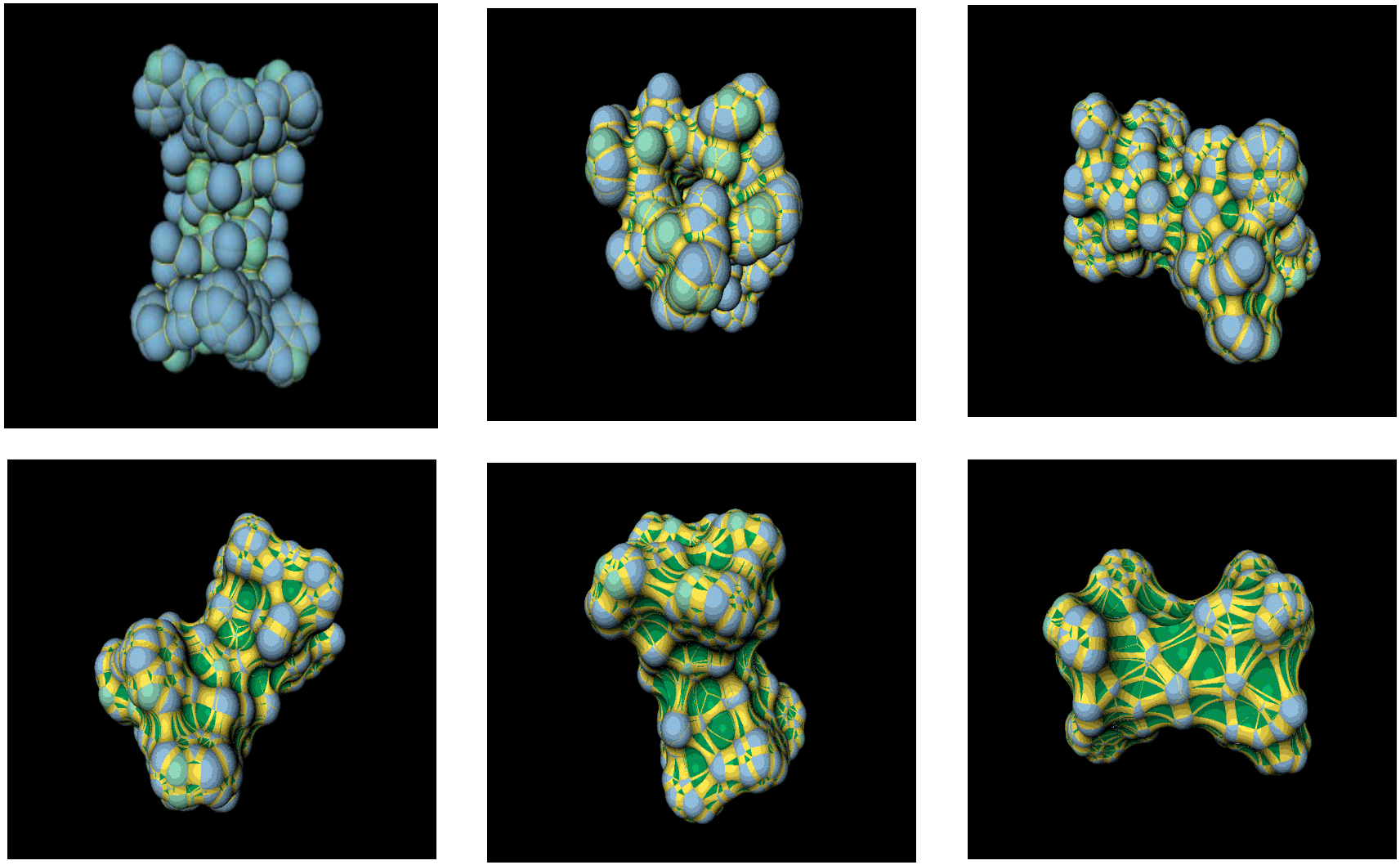


Solvent Contact Surface Patches

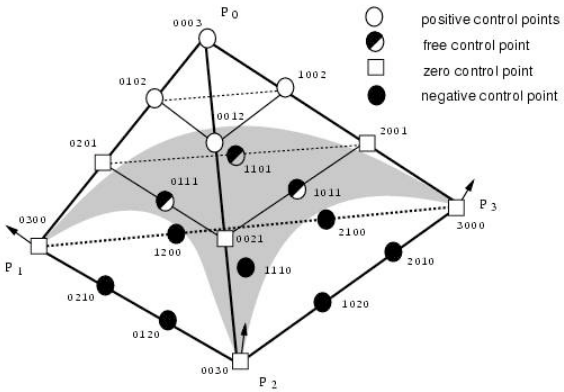


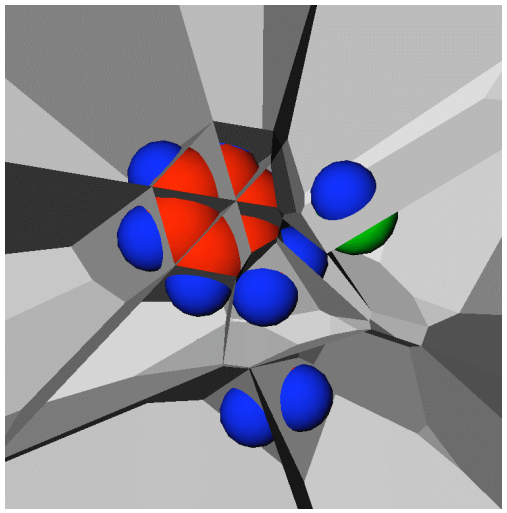
Solvent Excluded Surfaces

Molecular Surfaces for Varying Solvent Radii



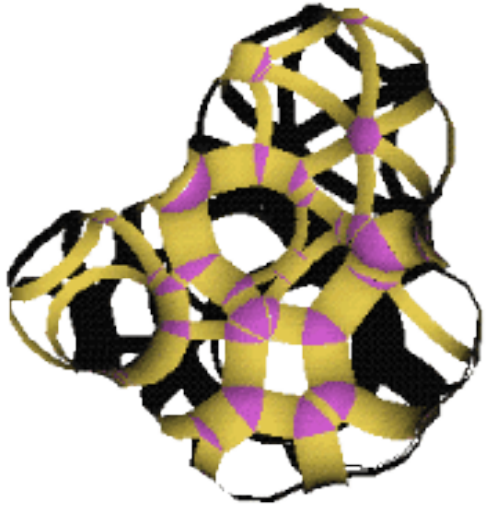
C^1 A-Patch Complex of the LR Molecular Surface



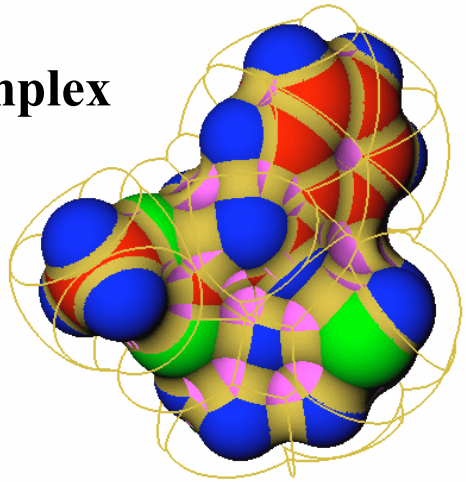


Laguerre Voronoi Diagram of Union

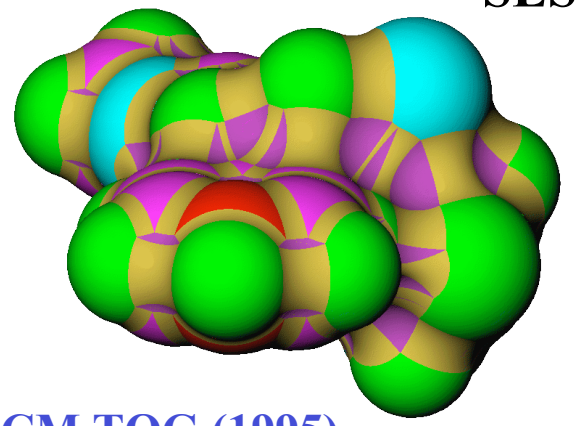
SES Curves Complex



Solvent Contact Patch Complex



SES



Bajaj, Chen, Xu, ACM TOG (1995)

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

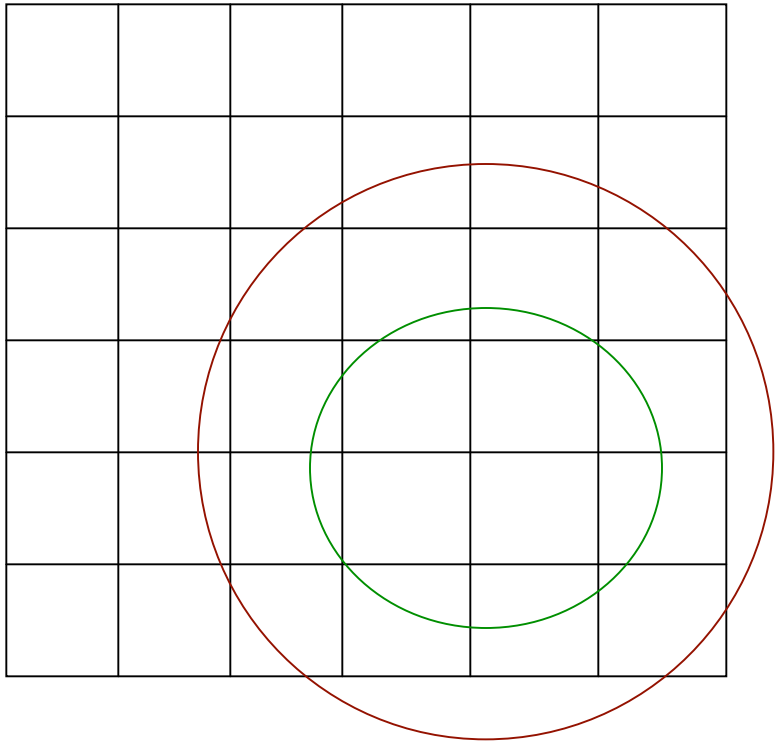
Adaptive Grid based Molecular Surface

- Main advantage:
  - Grids are commonly used in simulations
  - Grids permit many other operations including computing surface area etc.

- Sign distance function definition of SDF:
  - Let sdf(x) be the sign distance function of SAS surface.
  - Let the sign inside the surface be positive.
  - Surf={x:sdf(x)=probe radius}

Grid classification

N coarse cells
M atoms
Grid spacing chosen as ~ 0.5A,.



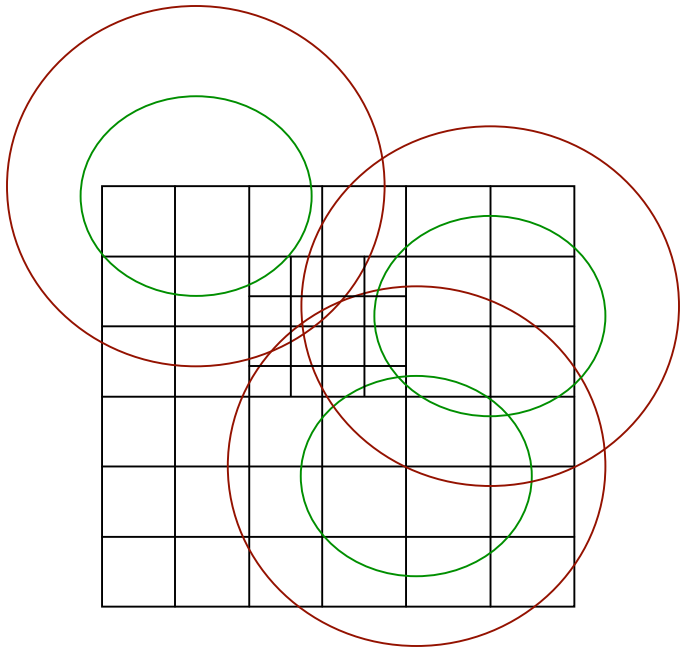
Classify:
Boundary SAS cell, vertices
Boundary VDW cells , vertices
Region in between SAS & VDW
Interior to VDW
Exterior to SAS

Each cell also contains atom intersections.

Atom cover at most h^3 cells. Atoms inserted in order into grid, updating classification.
Cost: O(Mh^3)

Gives us a patch complex defining the SES.
Intersecting atoms gives us location of
- Trimming curves
- Toroidal patches
- Spherical concave patches

SES by grid classification and SDF



Octree construction:
Subdivide cells with multiple SAS spheres,

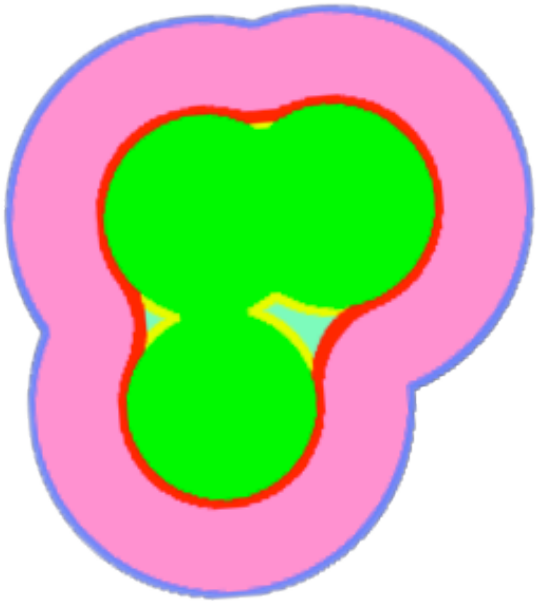
Keep max level L of subdivision

SDF computation:
For each point inside SAS,
    Search neighborhood to min dist to SAS
If cell contains too many intersecting cells at highest resolution, cell center is used to compute distance.
point - spherical patch distance implemented.

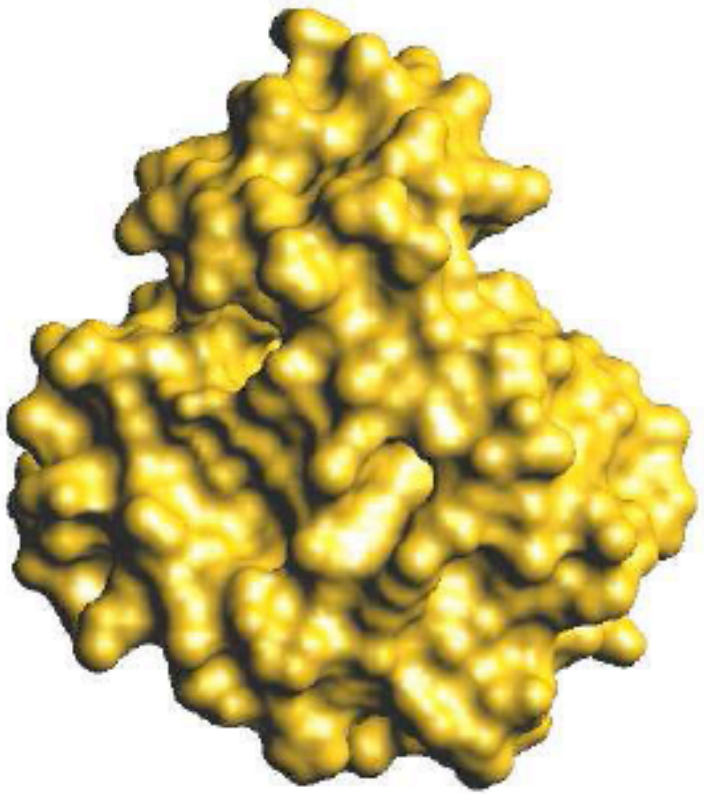
Isocontour approximation:
Isosurface with isovalue = r_p approximates SES.

Atom cover at most h^3 cells
Cost: Mh^6L
Propagation based SDF: Mh^3L

SDF classification results



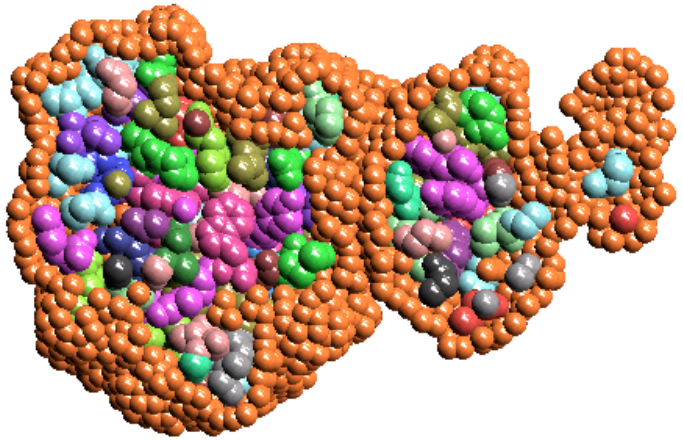
- Cut off of volume rendering
  - Dark blue: SAS surface
  - Pink: SAS volume
  - Red: SES surface
  - Yellow: part of VDW surface
  - Light blue: part of SES volume
  - Green: VDW volume



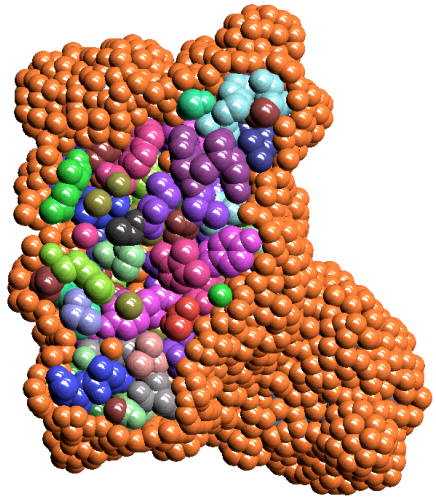
Myoglobin 101M.PDB
Time taken ~ 15s for base 64^3 grid

Surface atoms

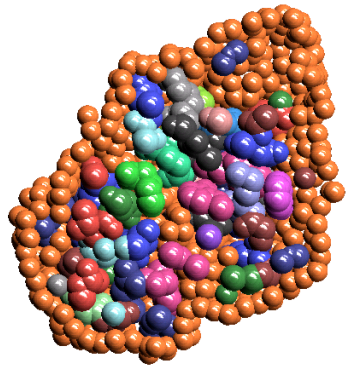
Grid points within atom approximates its distance to surface.
Cost: Once classification, SDF is complete, cost is at most linear in size of grid



2nd Protein of: 1IAI.PDB



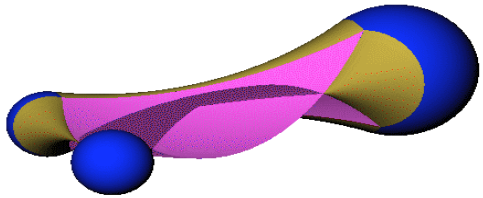
2VIR.PDB

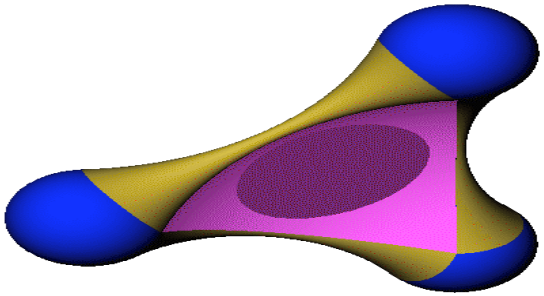


1BQL.PDB

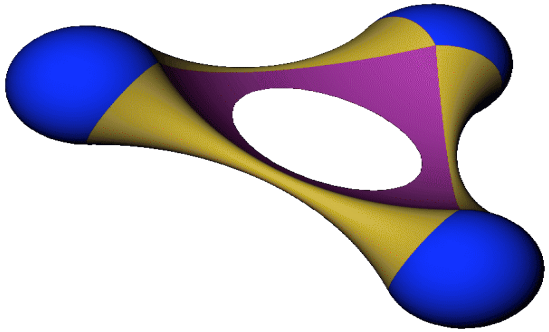
Interior atoms colored by residue type. Surface atoms in uniform orange color

Problem with LR: Singularities



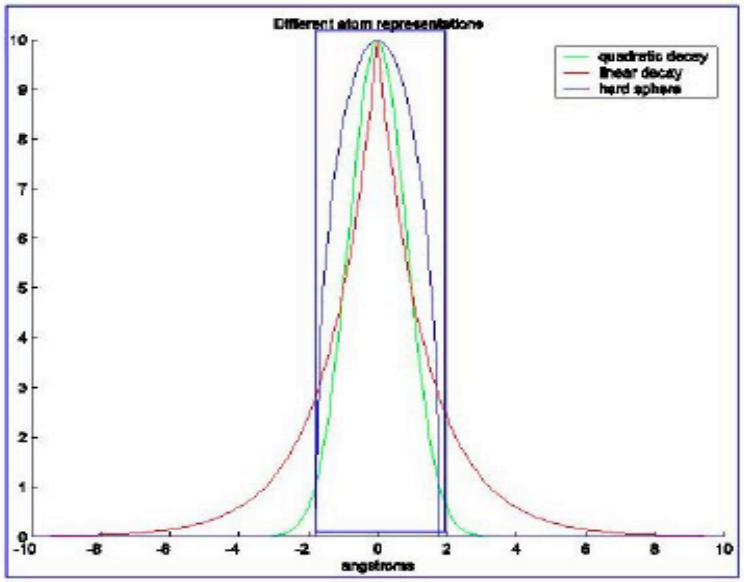


Probe at center touching 3 atoms



Atomic Shape Parameters

- Isotropic Quadratic Kernel $$G_i(\mathbf{x}) = e^{-\frac{\beta}{r_i^2}\left((\mathbf{x}-\mathbf{x}_i)^2 - r_i^2\right)}$$
- Isotropic Linear Kernel $$G_i(\mathbf{x}) = e^{-\beta\left(|\mathbf{x}-\mathbf{x}_i| - r_i\right)}$$
  - where
    - The decay β controls the shape of the Gaussian function.
    - The van der Waals radius is r_i.
    - The center of the atom is $\mathbf{x}_i$.
- Anisotropic Kernels

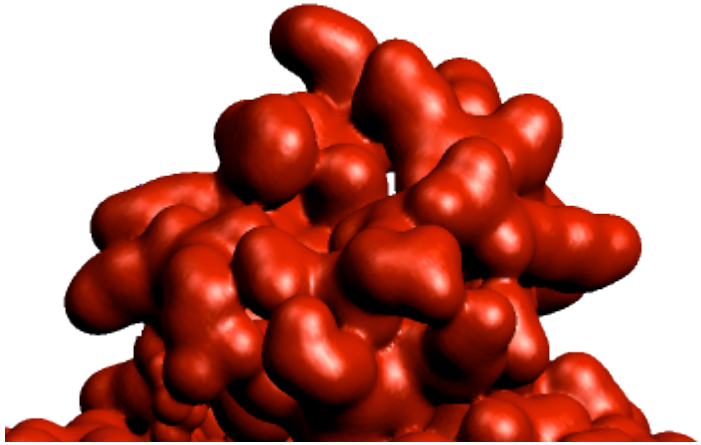


β = 2.3, r = 1.8 Å

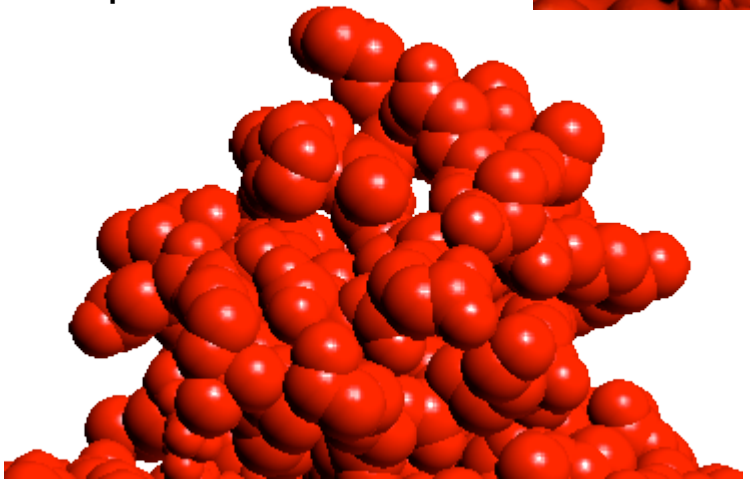
β value suggested in (Boys 50), (Grant Pickup 99)

Smooth Molecular Surfaces (Implicit Solvation) Models

Linear decay model



Hard sphere model



Quadratic decay model

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

Implicit Solvation Analytic Molecular Surfaces as Level Sets

- For a molecule with M atoms, we can define a synthetic electron density function as

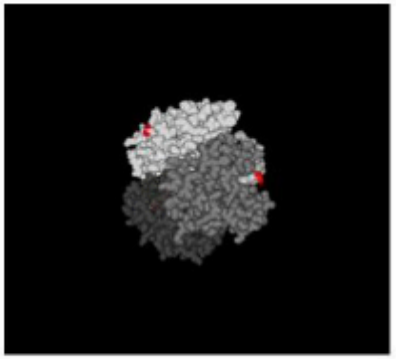
$$f_{elec\_dens}(\mathbf{x}) = \sum_{i=1}^{M} G_i(\mathbf{x}), \quad \mathbf{x} \in \Re^3$$

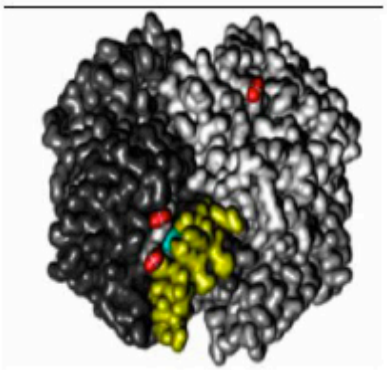
- Molecular surface for quadratic decay kernels, $A_i = e^{\beta}$:

$$f_{elec\_dens}(\mathbf{x}) = \sum_{i=1}^{M} A_i e^{-\frac{\beta}{r_i^2}(\mathbf{x}-\mathbf{x}_i)^2} \delta(\mathbf{x}-\mathbf{x}_i)$$

- Molecular Surface for linear decay kernels, $A_i = e^{\beta r_i}$:

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



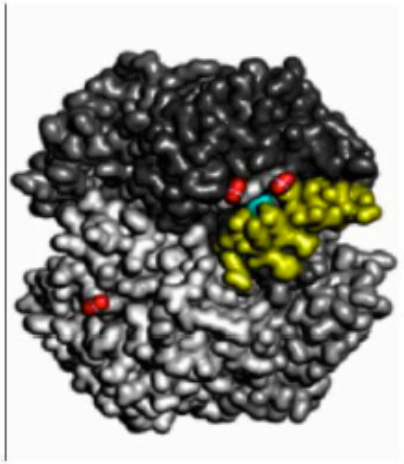


Hemoglobin Molecular surface

Fast Analytic Molecular Volume and Polarization Energy/Force Computations

- For smooth kernels G:

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



- The convolution theorem.

Convolution in spatial = multiplication in frequency

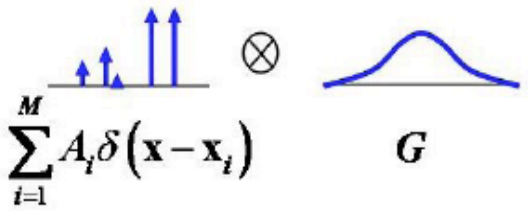
$$F_{elec\_dens} = FFT^{-1}\left(FFT(kernel) \times FFT(atom\ centers)\right)$$

For N cubature samples of M atom molecules

$O(NM) \rightarrow$ naiive

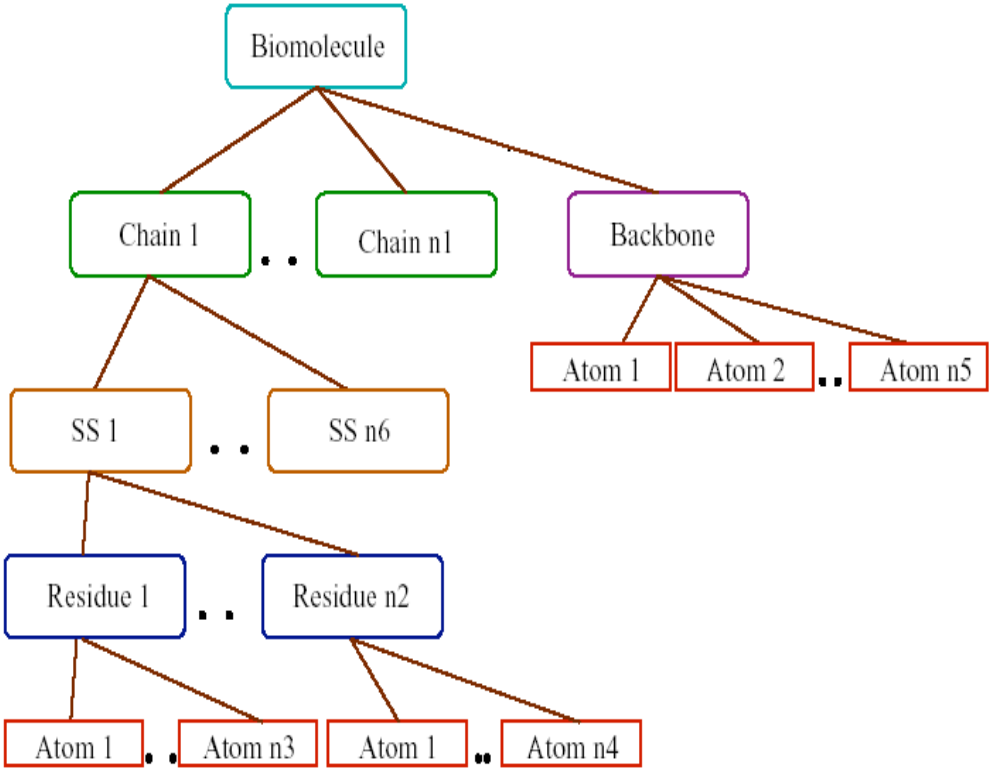
$O(M \log M + N \log N) \rightarrow$ irregular FFT

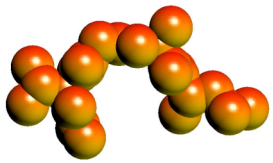
Bajaj, Siddahanavalli, (2005)

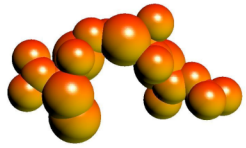


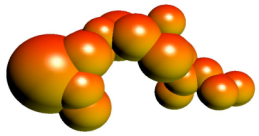
FCC Cluster Hierarchy

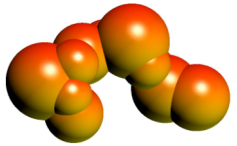
- Clustering of atoms based on biochemical units as well preserve molecular shape features

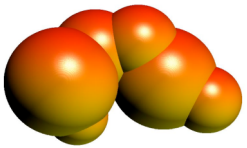


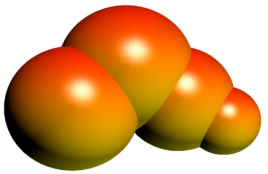




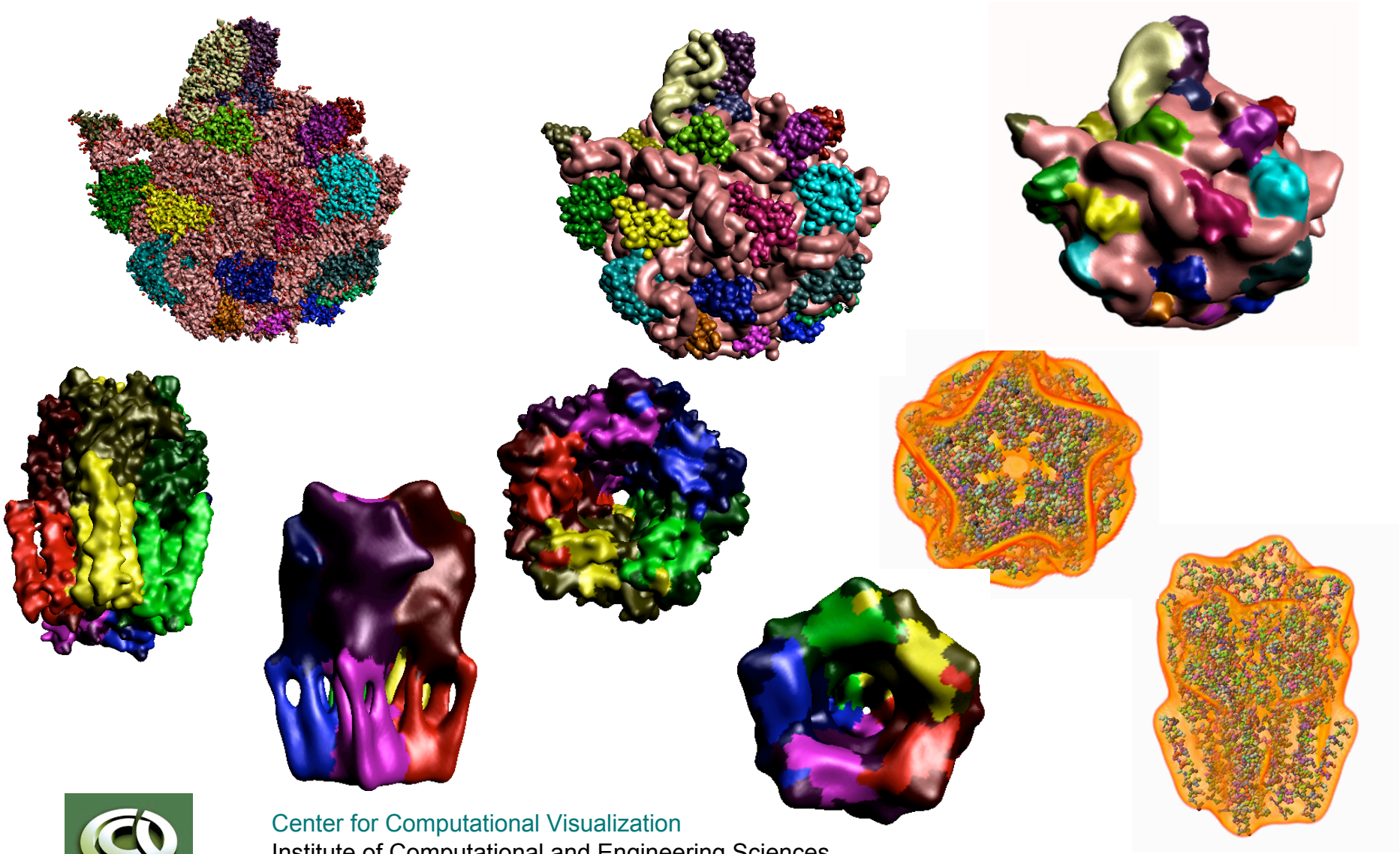






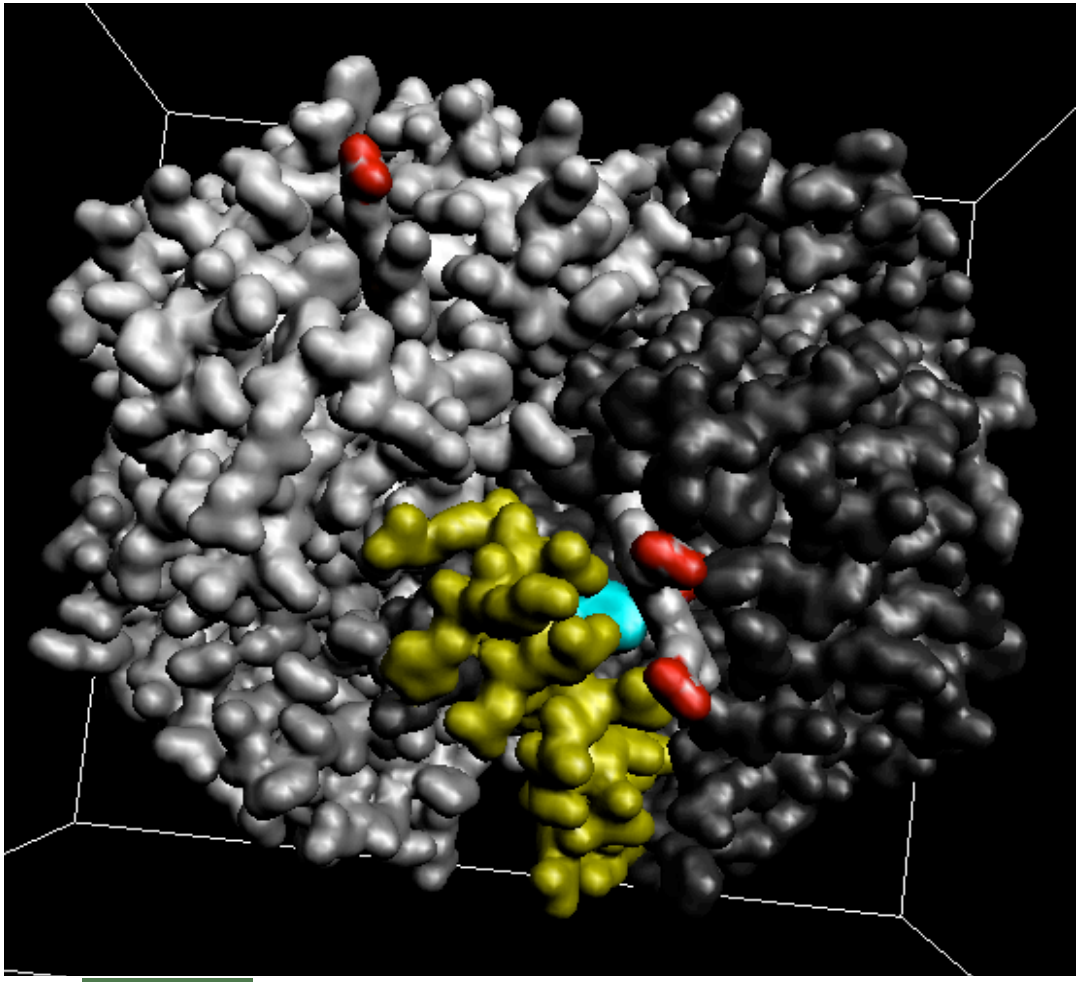


FCC Multi-Resolution Models



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

Molecular Surface Segmentation



Subunit A
Subunit B
Subunit C
Subunit D

Within Subunit A

F helix
Histidine Ligand(HIS87)
O_2

• Oxy process : O2 binds to the Fe2+ ion on the opposite side of the histidine ligand. F helix shifts position through the oxy-deoxy cycle.

Molecular Surfaces Properties

- Curvatures

Let f (x,y,z) = 0 represent an implicit function in $\mathbf{R}^3$.

The Mean curvature H and Gaussian curvature K

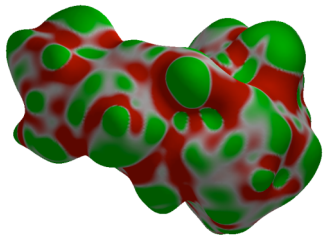
$$H = (C (f_x^2 (f_{yy} + f_{zz})) - 2 * C (f_x f_y f_{xy})) / (2 * (C (f_x^2))^{1.5})$$

and

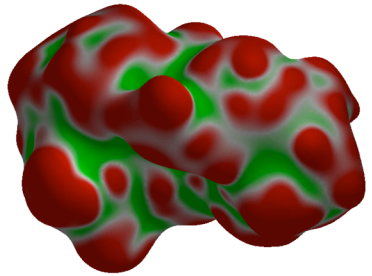
$$K = (2 * C (f_x f_y (f_{xz} f_{yz} - f_{xy} f_{zz}))) / ((C (f_x^2))^2)$$

Where C represents a cyclic summation over x, y and z, and the subscripts denote partial differentiation with respect to those variables.

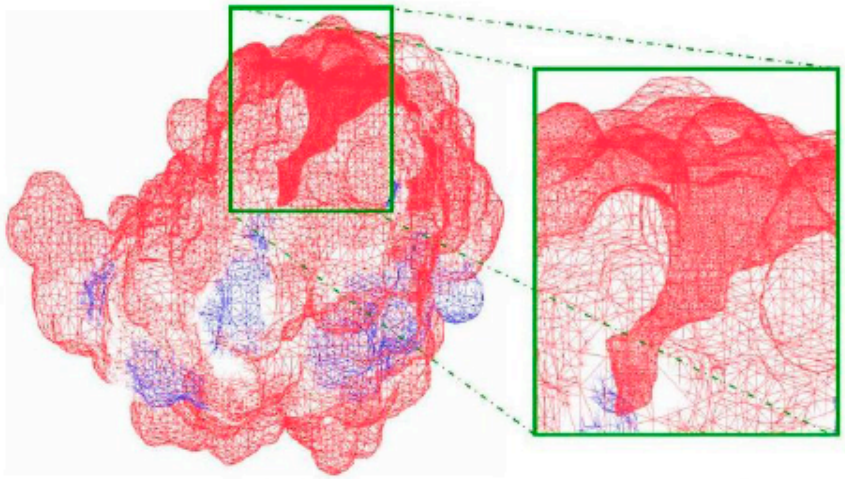
Gaussian curvature
Protein kinase (1A06.pdb)



Mean curvature
Protein kinase (1A06.pdb)

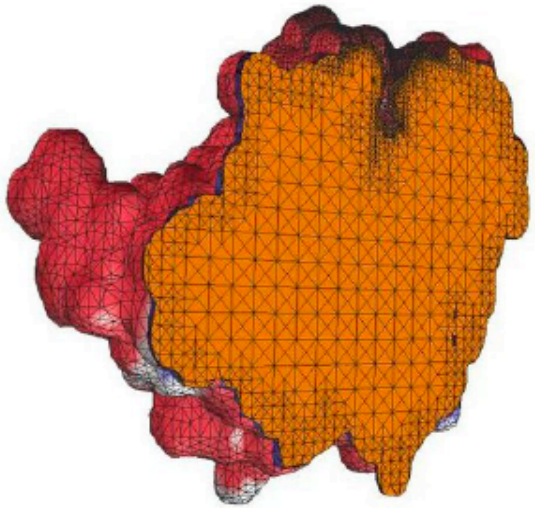


Meshing

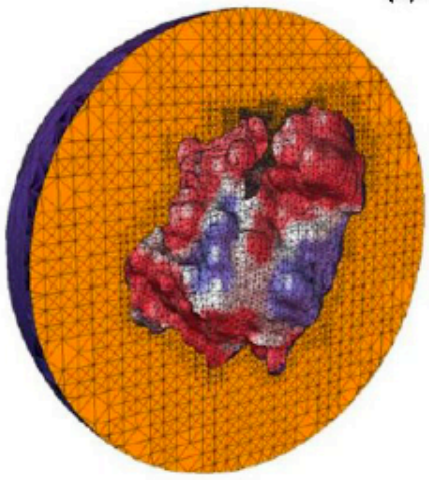


(a) monomer mAChE

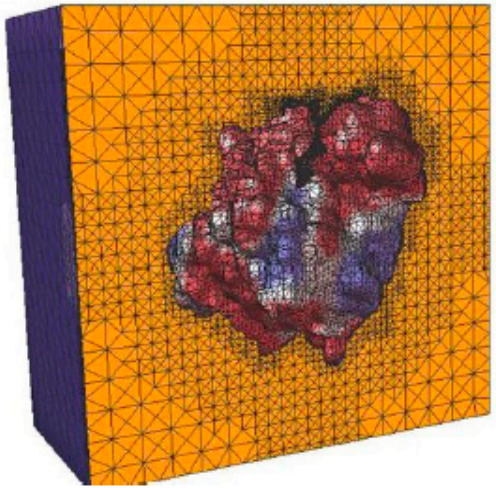
(b) cavity



(c) interior mesh



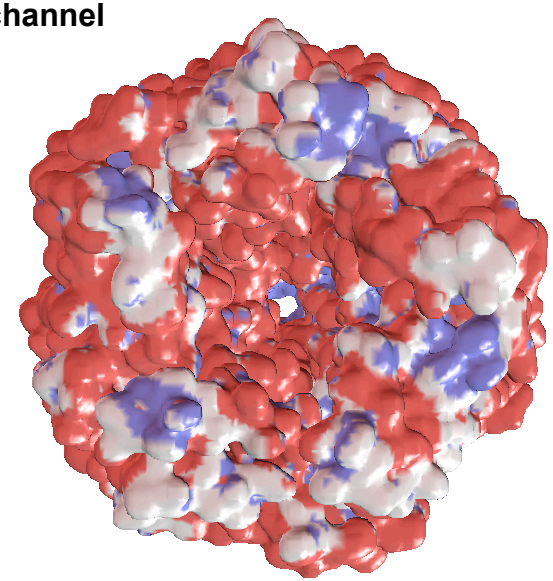
(d) exterior meshes



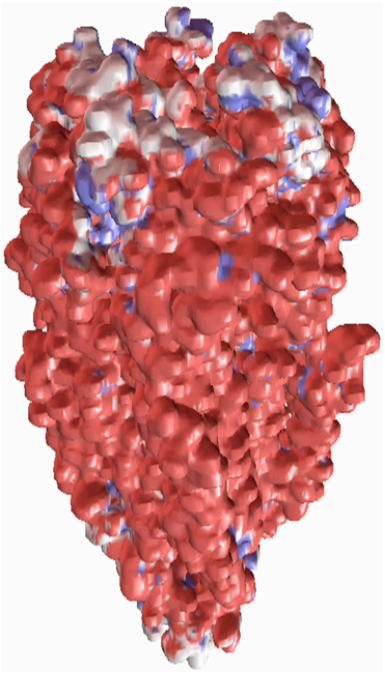
•Y. Zhang, C. Bajaj, B. Sohn, *Special issue of Computer Methods in Applied Mechanics and Engineering (CMAME) on Unstructured Mesh Generation, 2004.*

•Y. Zhang, C. Bajaj, *Proc. of Meshing Roundtable 2005.*

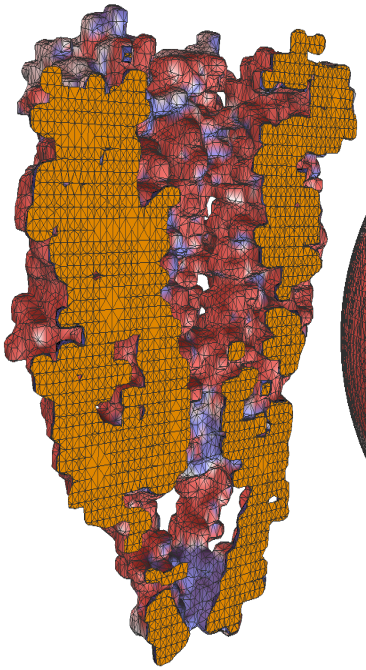
Open channel



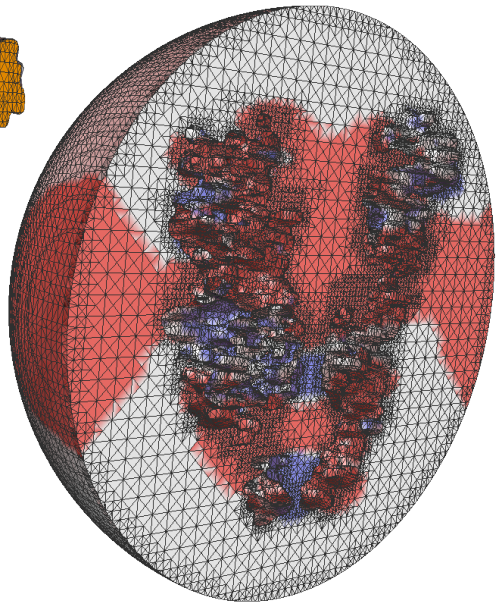
Top view



side view

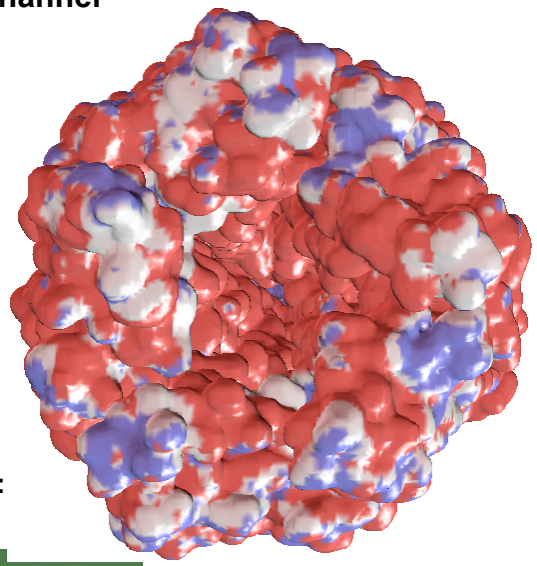


the interior mesh

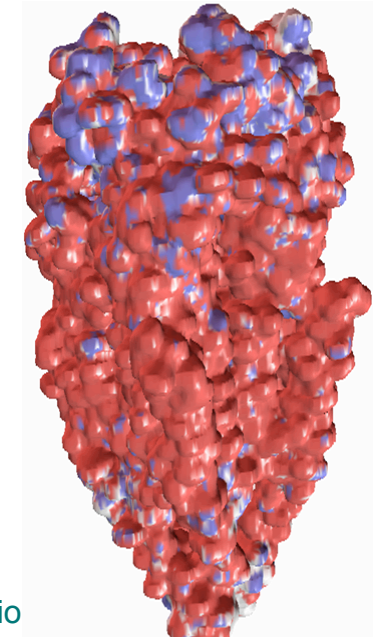


The exterior mesh

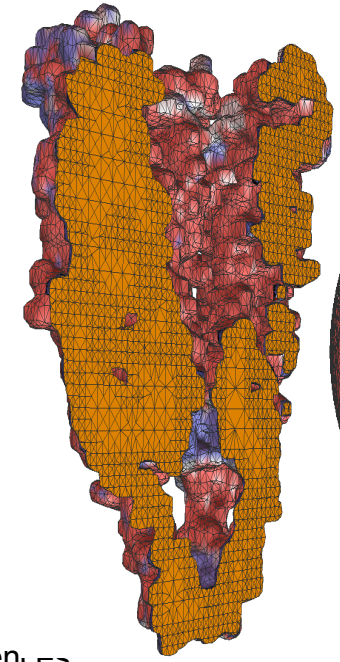
Closed channel



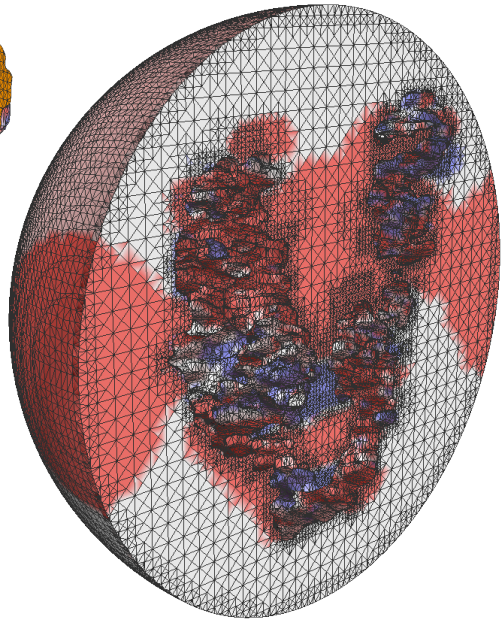
Top view



side view



the interior mesh



The exterior mesh

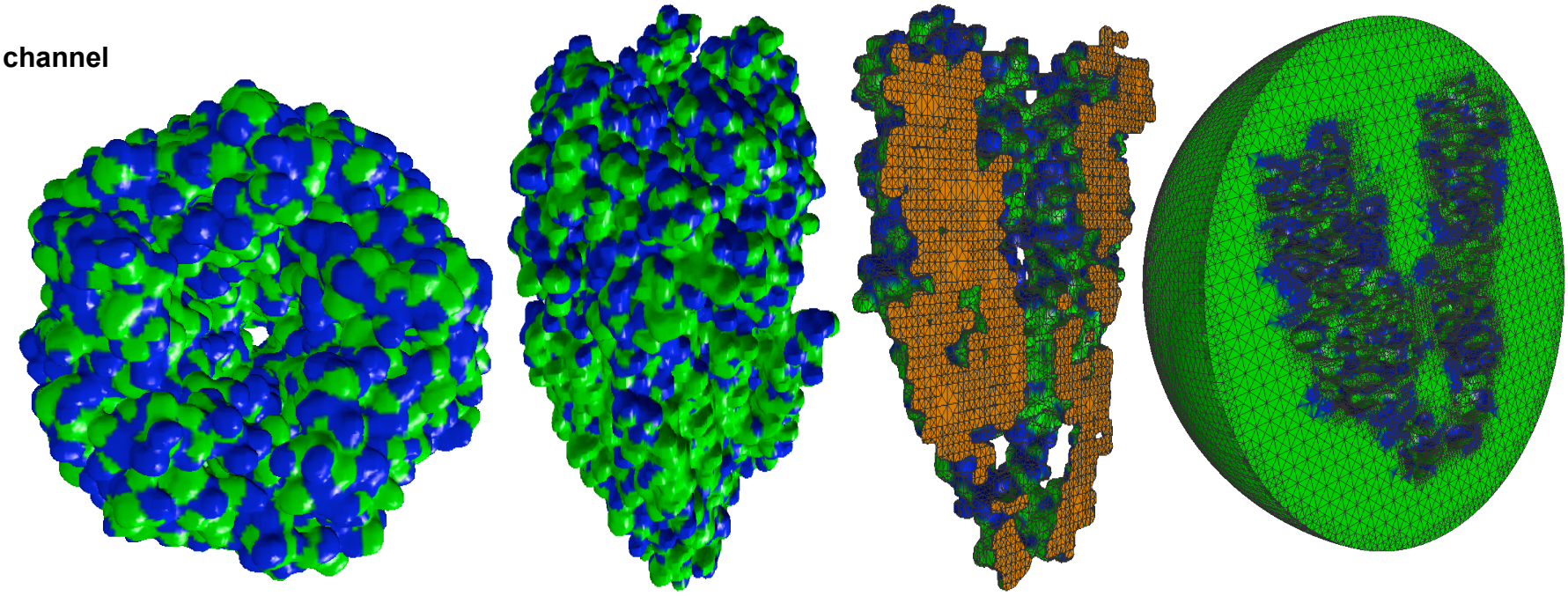
Color map:

< -1 – red

(-1, 1) – white

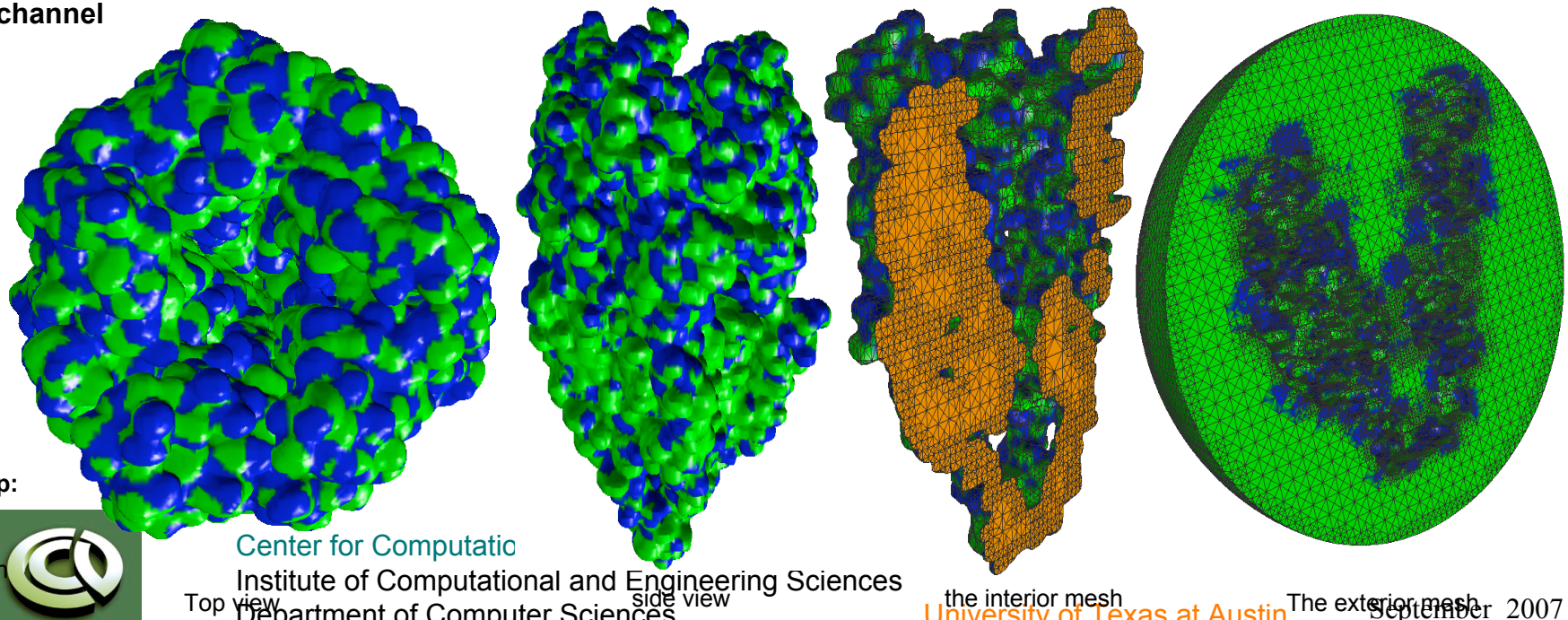
> 1 - blue

Open channel



Top view | side view | the interior mesh | The exterior mesh

Closed channel



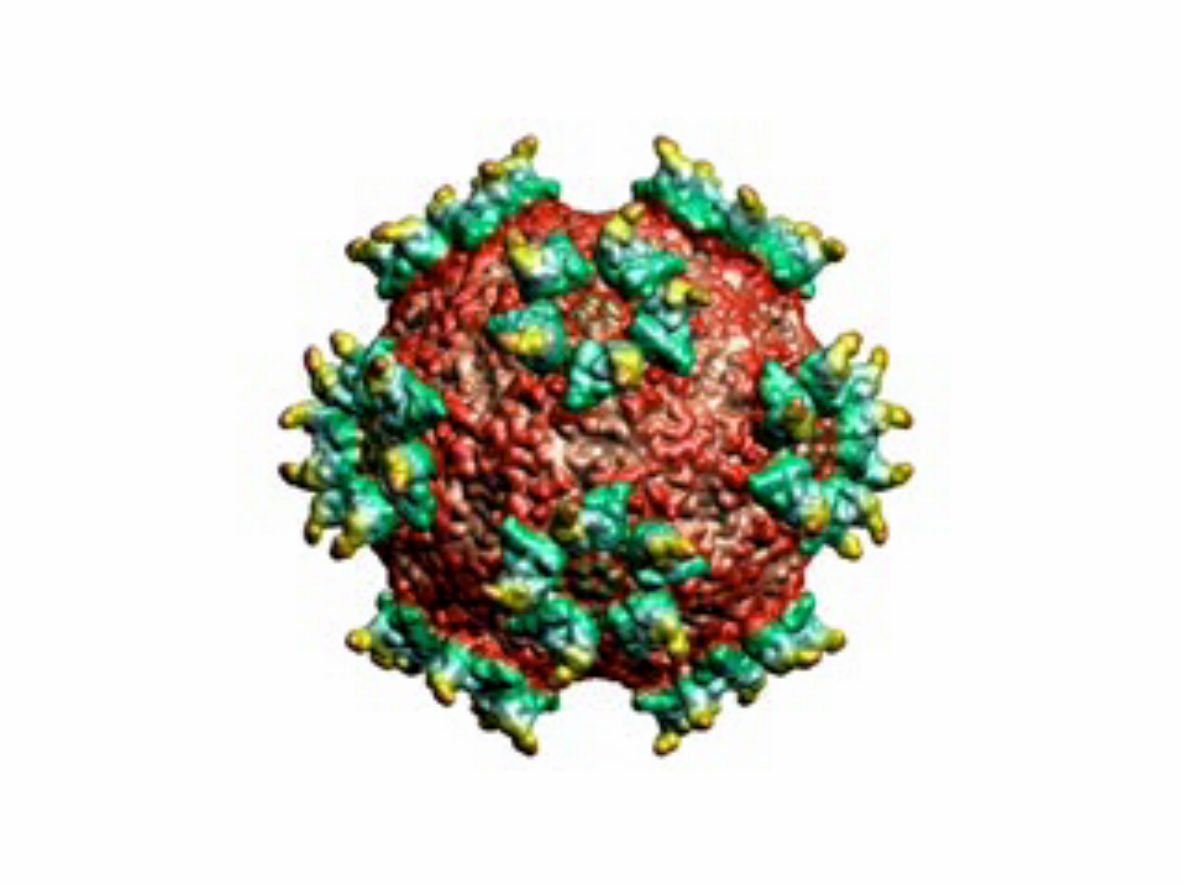
Top view | side view | the interior mesh | The exterior mesh

Color map:

> 0 – blue

< 0 - green

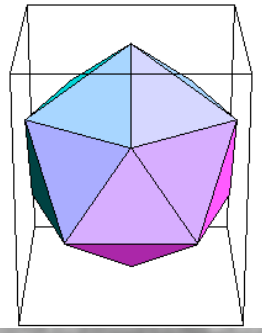
Human Rhinovirus



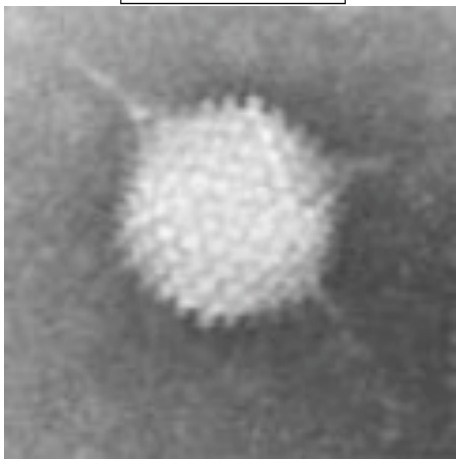
The Human Rhinovirus 14 (1RVF.pdb) complexed (docked) with Immunoglobulins

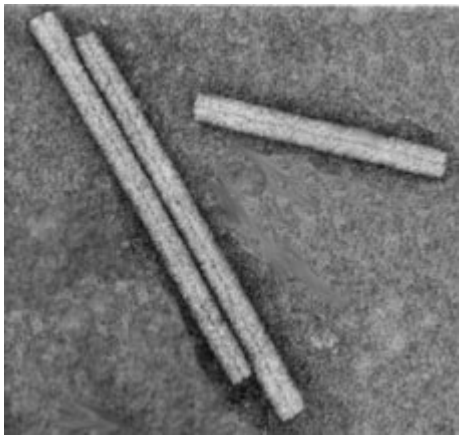
Center for Computational Visualization
Institute of Computational and Engineering Sciences
Department of Computer Sciences
University of Texas at Austin

Types of Viruses

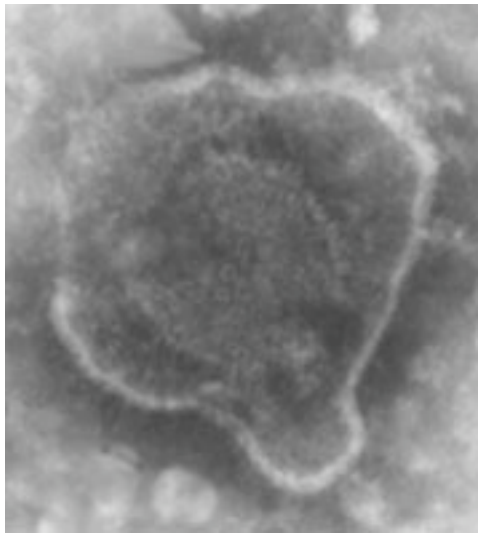


Icosahedra

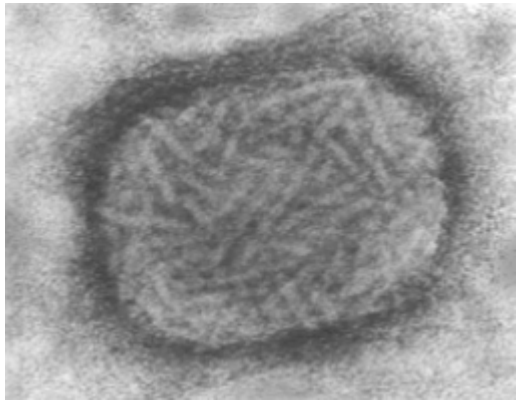




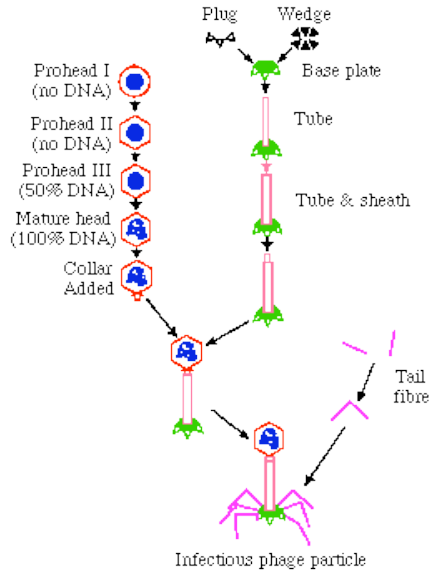
Helical (tobacco mosaic virus)



Enveloped

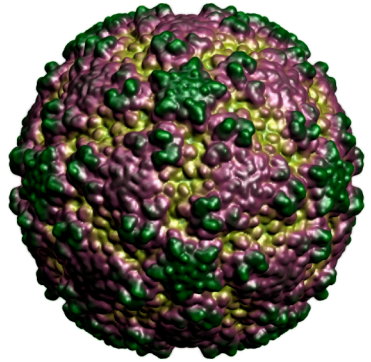


Complex (Poxvirus family)



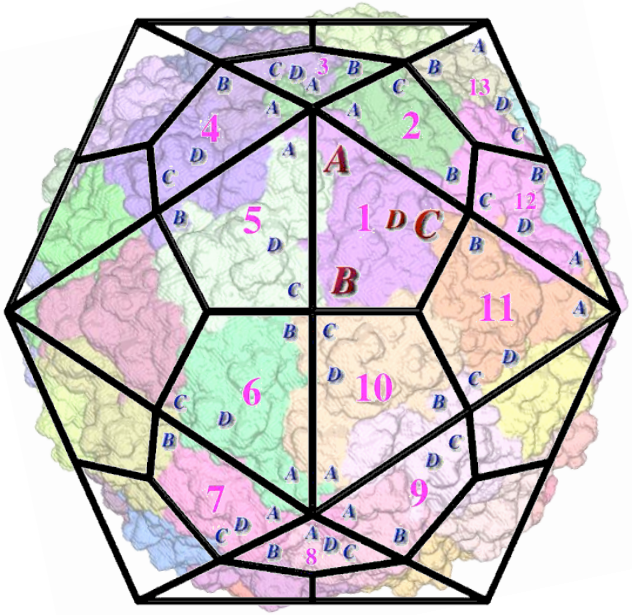
Complex (tailed virus)

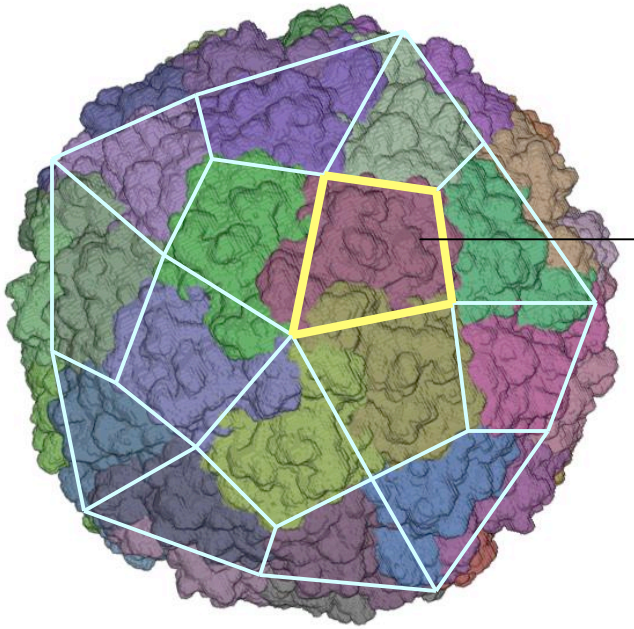
Human Rhinovirus Serotype 2



- Subunit PDB id: 1FPN
- Number of subunits: 60
- Number of atoms per subunit: 6316
- Resolution (Å): 2.6
- Dimension (Å): 308.68 x 352.98 x 380.48
- Symmetry: icosahedral, T=1

The Capsid: Human Rhinovirus (1FPN)





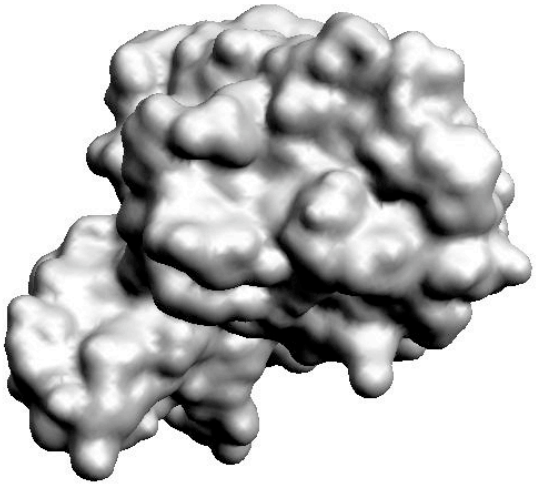
A: Coat Protein VP1

B: Coat Protein VP3

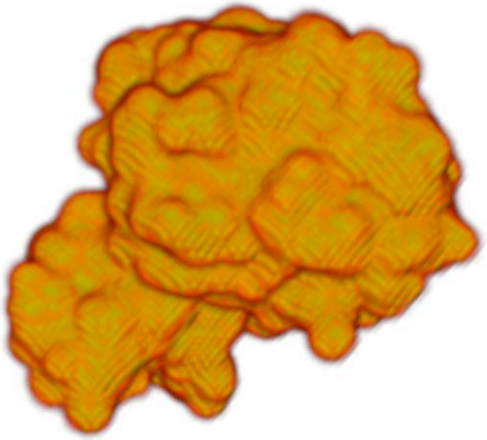
C: Coat Protein VP2

D: Coat Protein VP4

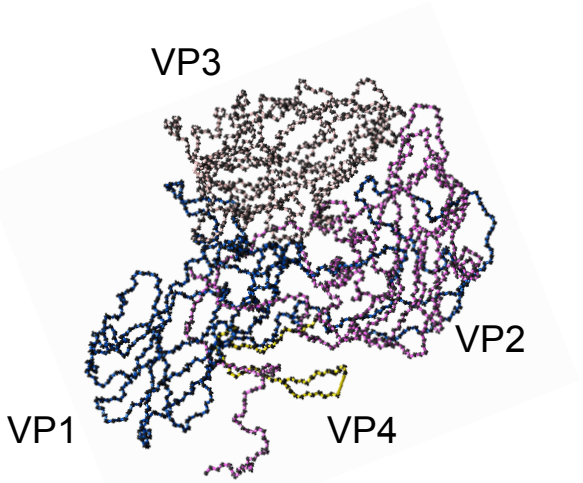
Capsomeres: (1FPN)

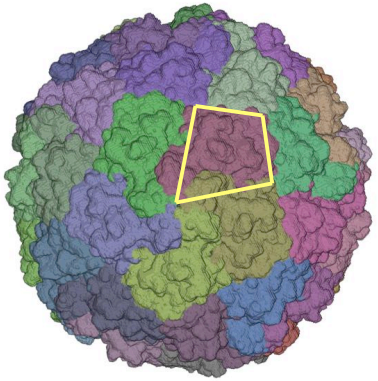


Iso-surface rendering of a capsomere



Volume rendering of a capsomere





5-Fold Symmetry

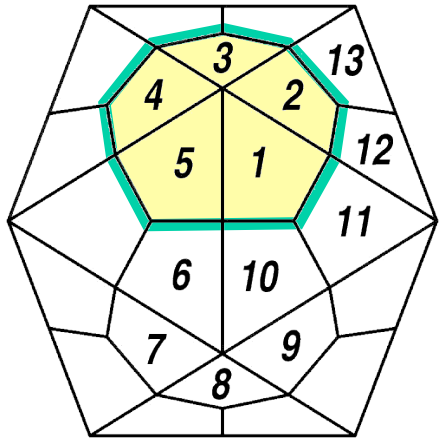
1) 1.00000 0.00000 0.00000
0.00000 1.00000 0.00000
0.00000 0.00000 1.00000

2) 0.30902 -0.80902 0.50000
0.80902 0.50000 0.30902
-0.50000 0.30902 0.80902

3) -0.80902 -0.50000 0.30902
0.50000 -0.30902 0.80902
-0.30902 0.80902 0.50000

4) -0.80902 0.50000 -0.30902
-0.50000 -0.30902 0.80902
0.30902 0.80902 0.50000

5) 0.30902 0.80902 -0.50000
-0.80902 0.50000 0.30902
0.50000 0.30902 0.80902

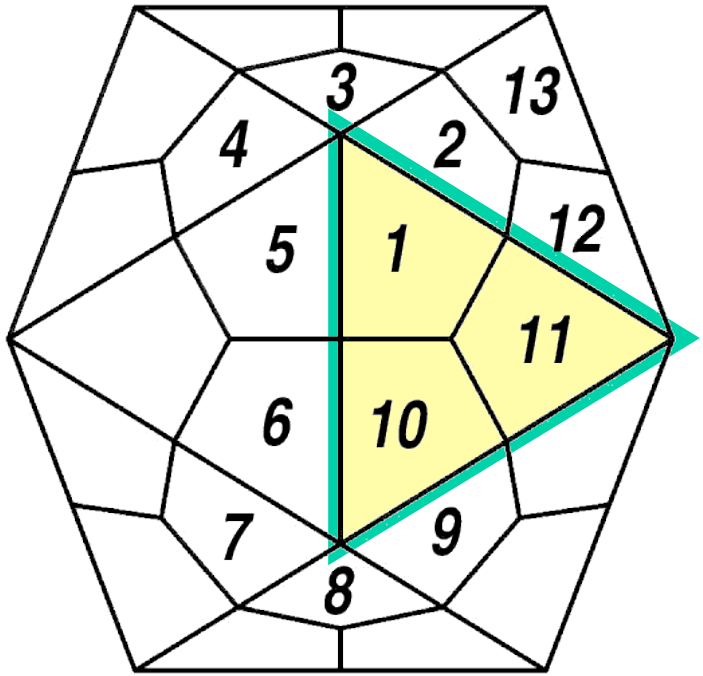


3-Fold Symmetry

1) 1.00000 0.00000 0.00000
0.00000 1.00000 0.00000
0.00000 0.00000 1.00000

10) -0.30902 -0.80902 0.50000
0.80902 -0.50000 -0.30902
0.50000 0.30902 0.80902

11) -0.30902 0.80902 0.50000
-0.80902 -0.50000 0.30902
0.50000 -0.30902 0.80902



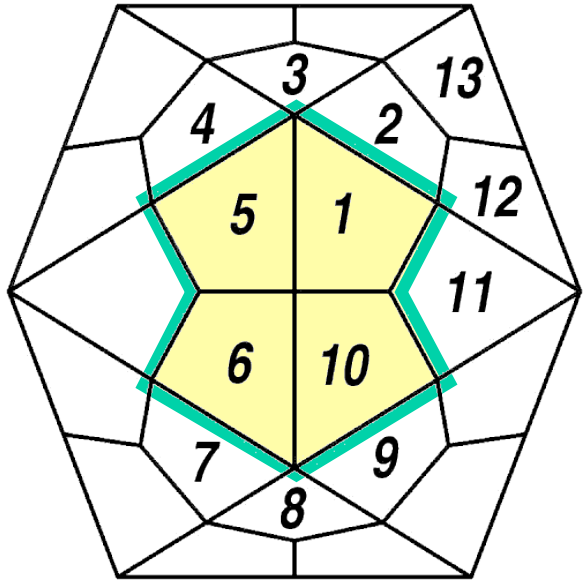
2-Fold Symmetry

1) 1.00000 0.00000 0.00000
0.00000 1.00000 0.00000
0.00000 0.00000 1.00000

5) 0.30902 0.80902 -0.50000
-0.80902 0.50000 0.30902
0.50000 0.30902 0.80902

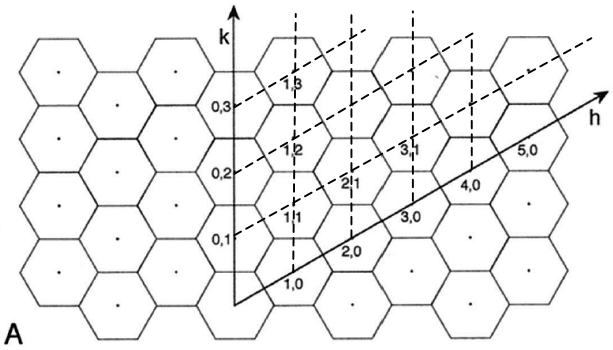
6) -1.00000 0.00000 0.00000
0.00000 -1.00000 0.00000
0.00000 0.00000 1.00000

10) -0.30902 -0.80902 0.50000
0.80902 -0.50000 -0.30902
0.50000 0.30902 0.80902

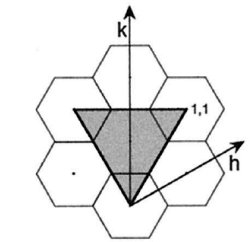


Icosahedral Symmetry: Triangulation Numbers

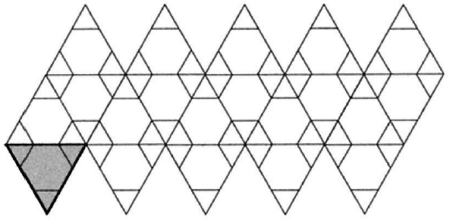
- Icosahedral symmetry overview (Caspar & Klug 1962; Baker et al. 1999)

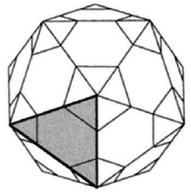


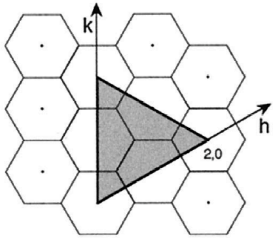
A



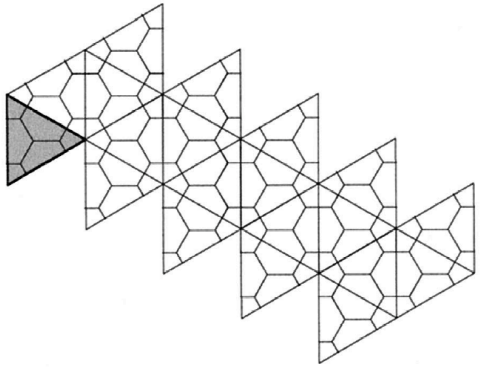
B T=3 (h,k=1,1)

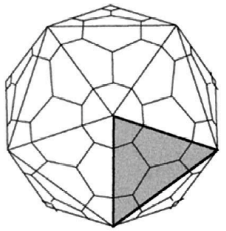






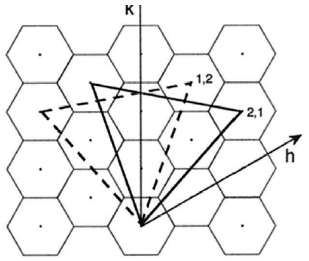
C T=4 (h,k=2,0)



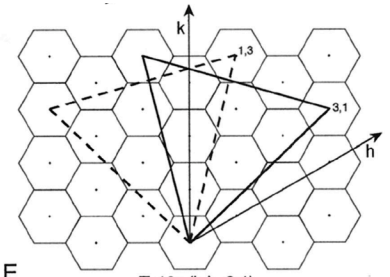


h k	T	Example
1 0	1	bacteriophage ϕX174
1 1	3	tomato bushy stunt virus
2 0	4	Sindbis virus
1 2	7*d*	polyoma virus
3 1	13*l*	reovirus
1 3	13*d*	infectious bursal disease virus
4 0	16	herpesvirus
5 0	25	adenovirus

Notations: *d* = *dextro* (right handed)
l = *laevo* (left handed)

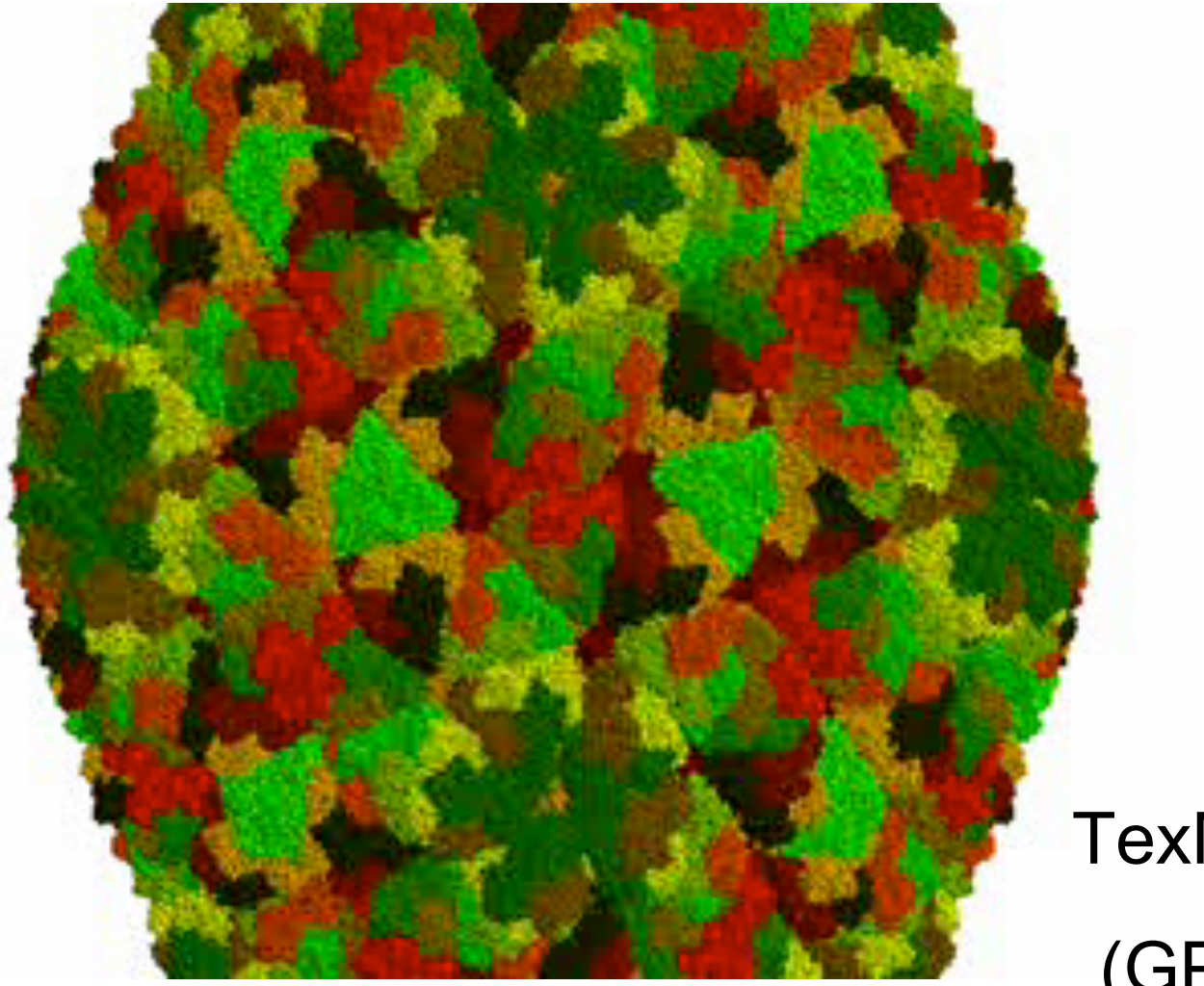


D T=7*l* (h,k=2,1)
T=7*d* (h,k=1,2)



E T=13*l* (h,k=3,1)
T=13*d* (h,k=1,3)

Rice Dwarf Virus (High Resolution)



3.6M atoms

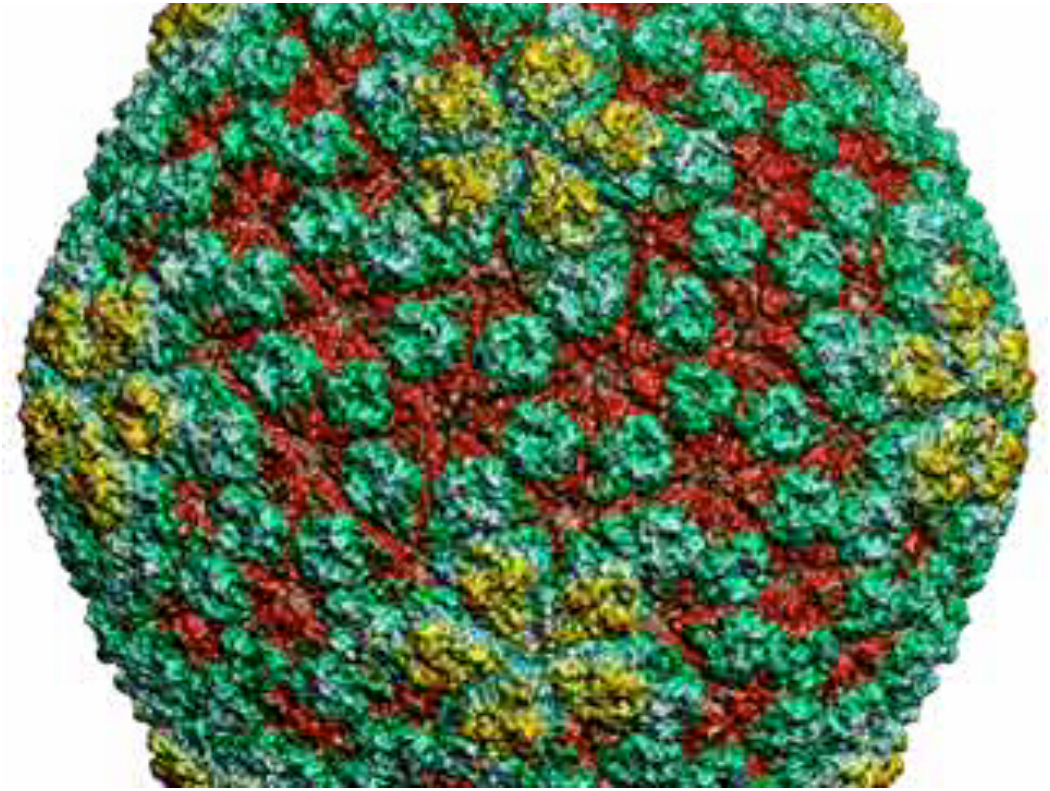
Texture Sphere Rendering

TexMol

(GPU accel)

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

Rice Dwarf Virus (medium resolution)



Fast Isocontour rendering from Blurred Maps

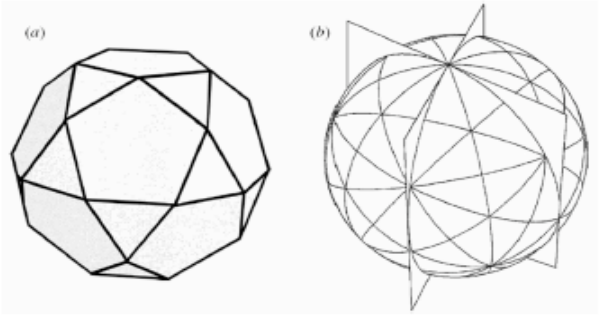
UT TexMol

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

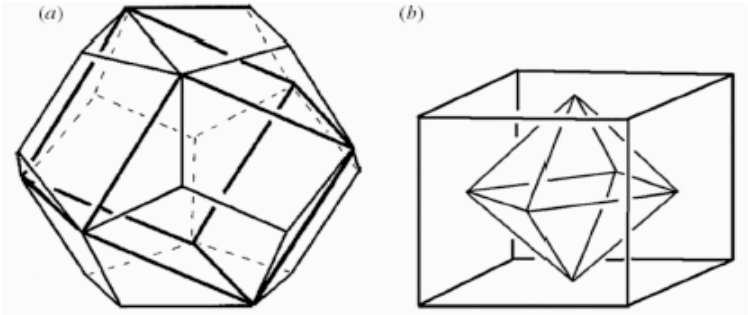
September 2007

Tiling Theory - Filling some Gaps (Twarock'04)

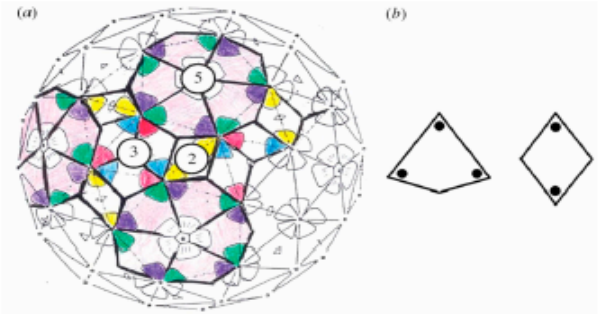
- Casper-Klug theory is incomplete; e.g., it cannot account for the structures of:
  - *Papovaviridae* family, which contain cancer-causing viruses
  - icosahedral viruses with pentamers, such as *polyomavirus*
  - *Sericestis* and *Tipula* iridescent viruses (follow Goldberg polyhedral structure)
- Viral Tiling theory closes the gap. It describes locations of protein subunits and inter-subunit bonds based on mathematical theory of quasicrystals.



(*a*) root polytope, (*b*) some reflection planes encoded by root vectors



(*a*) cube inscribed in dodecahedron, (*b*) tetrahedron in a cube



tilings of the virus capsids of polyomavirus and simian virus 40

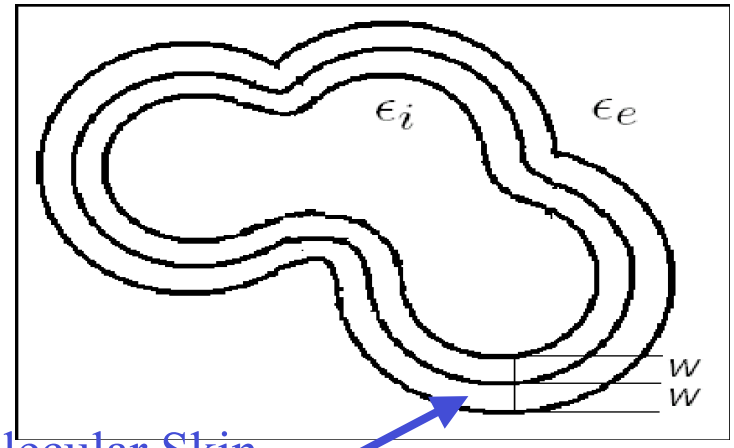
Molecular Skins (or Shells)

For atom i, define the volume density as

$$\rho_i(x) = \begin{cases} 1 & x \le a_i - w \\ \frac{1}{4w^3}(x-(a_i-w))^3 - \frac{3}{4w^2}(x-(a_i-w))^2 + 1 & a_i - w < x < a_i + w \\ 0 & x \ge a_i + w \end{cases}$$

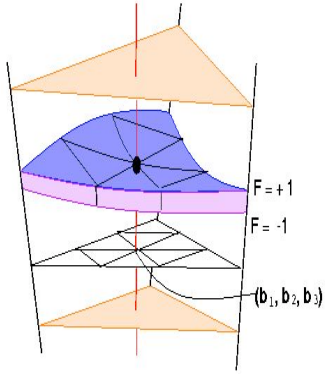
where $x = \|\mathbf{r} - \mathbf{x}_i\|$

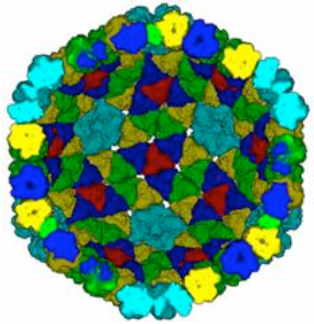
$$\rho(\mathbf{r}) = \sum_i \rho_i - \sum_{i<j} \rho_i \rho_j + \sum_{i<j<k} \rho_i \rho_j \rho_k - \sum_{i<j<k<l} \rho_i \rho_j \rho_k \rho_l$$ inclusion-exclusion

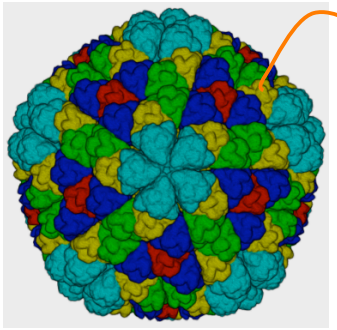


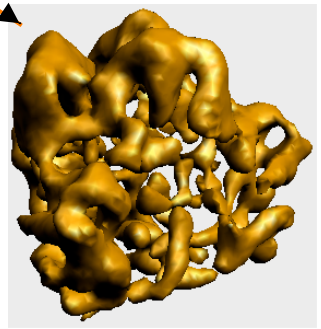
Molecular Skin

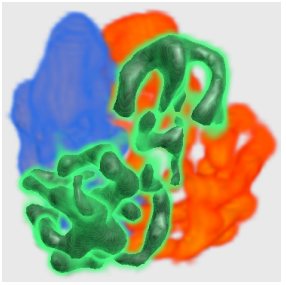
C^1 A-Shell Molecular Skin Models

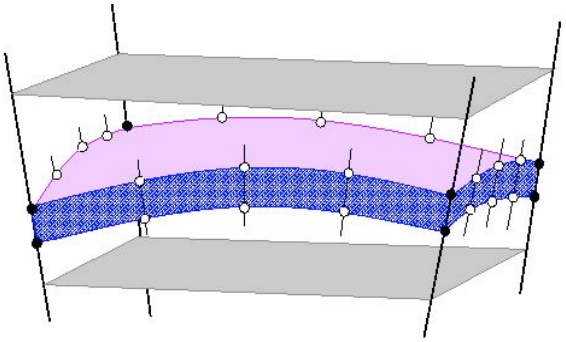






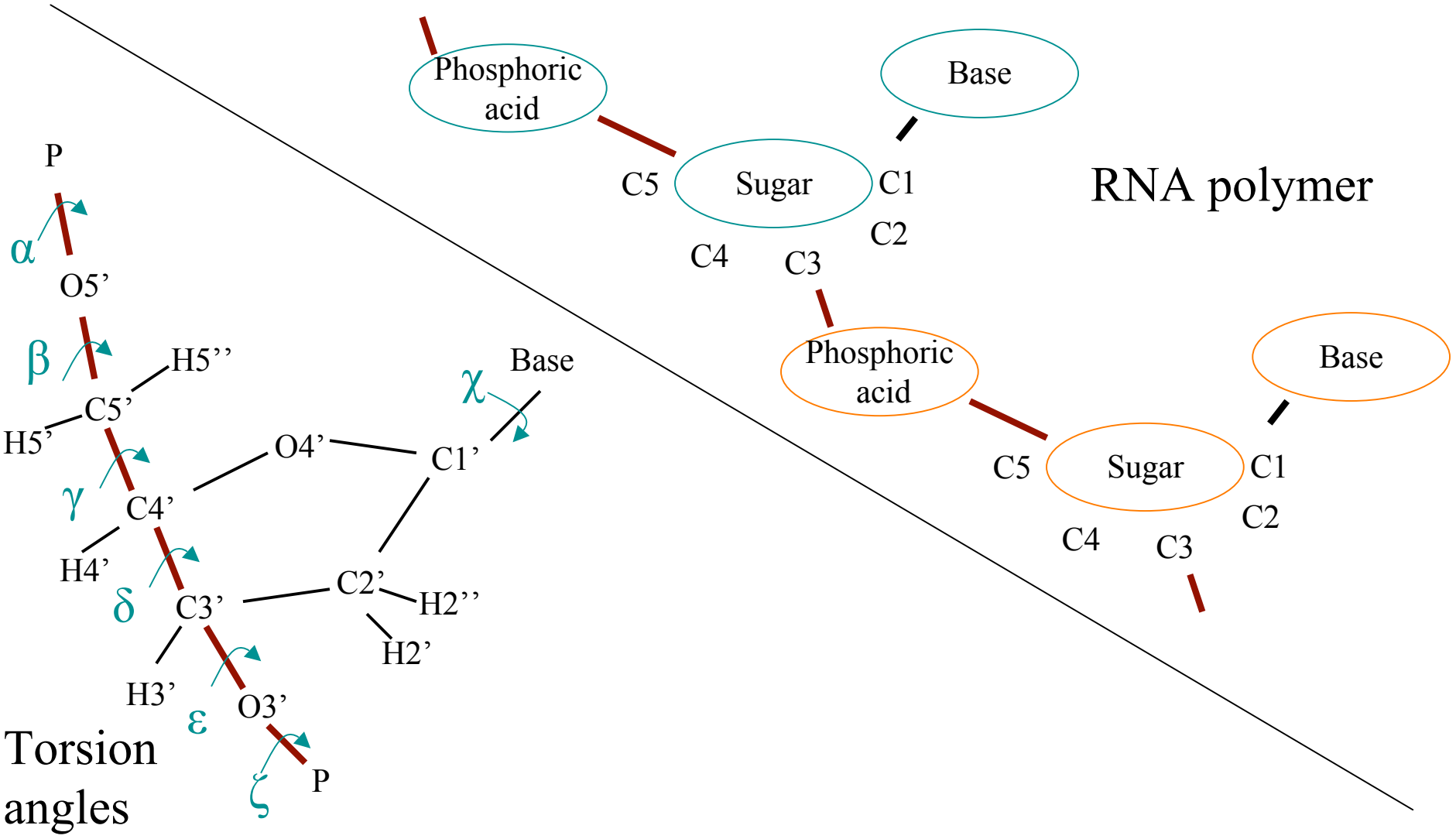






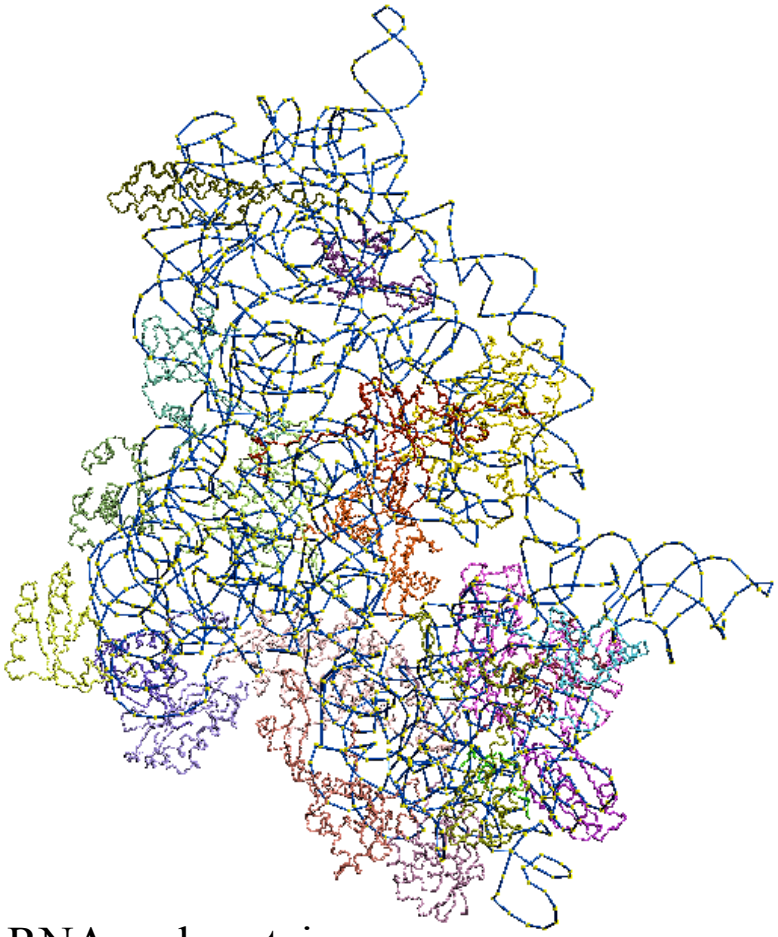
Bajaj, Geometric Mod. Computing, (2001)

Geometric properties: Flexibility in RNA

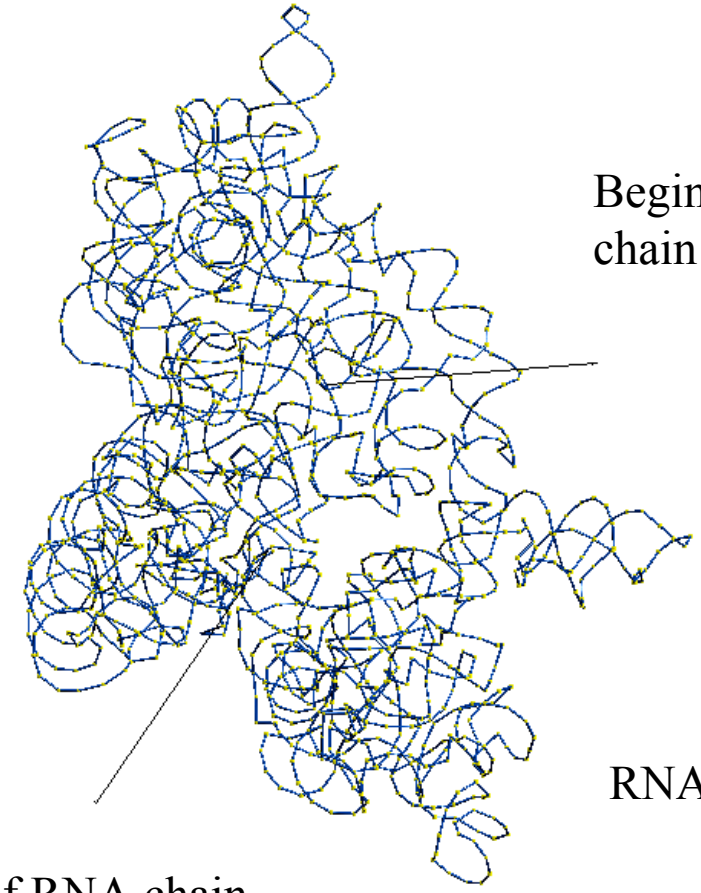


Can also specify ribose dihedral angles and puckering phase, amplitude

Small subunit of ribosome



RNA and protein backbone



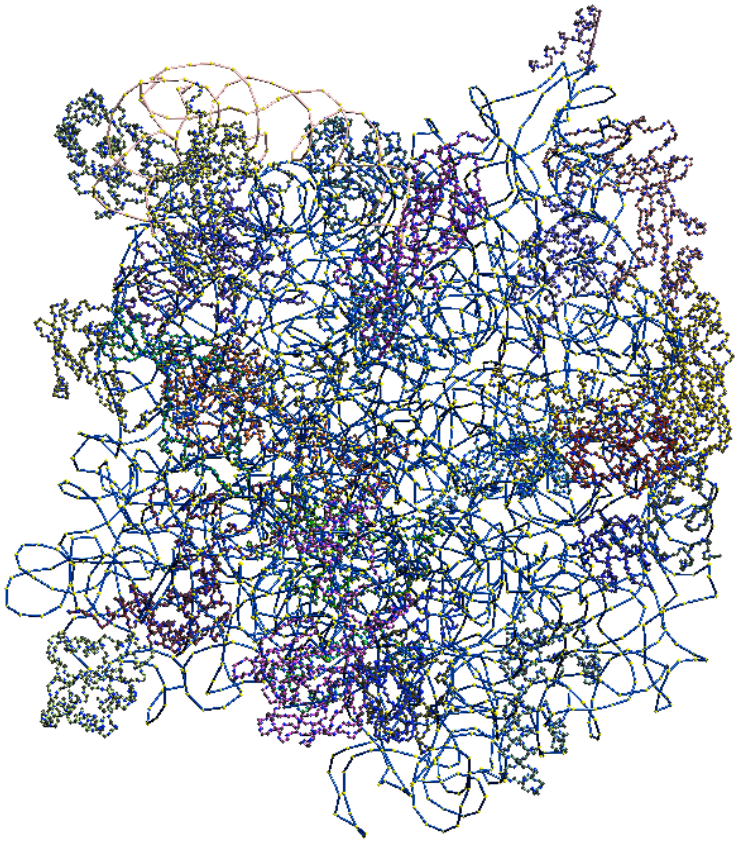
Beginning of RNA chain

RNA backbone

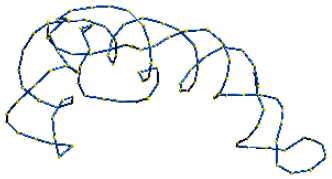
End of RNA chain

A 30S ribosome molecule (1J5E.pdb) 21 chains including a single RNA chain

Large subunit of ribosome

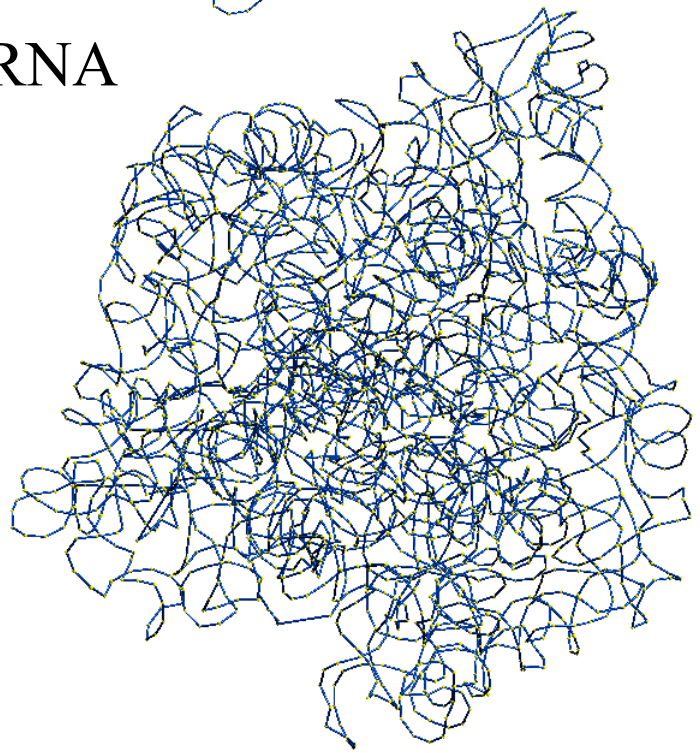


Entire backbone



5S RRNA

2 RNA chains



23S RRNA

A 50S ribosome molecule (1JJ2.pdb) with 2 RNA chains

Ribosome Active Sites