Tools for the Analysis of Macromolecular Motions & Geometry

Mark Gerstein

Analysis of Macromolecular Geometry & Motions

W Krebs, C Chothia,
J Tsai, Y Harpaz, R Taylor

bioinfo.mbb.yale.edu/MolMovDB

1 Database of Macromolecular Motions

Interface packing, shear v. hinge, web DB, 125+4000 motions

2 Morph Movie Server

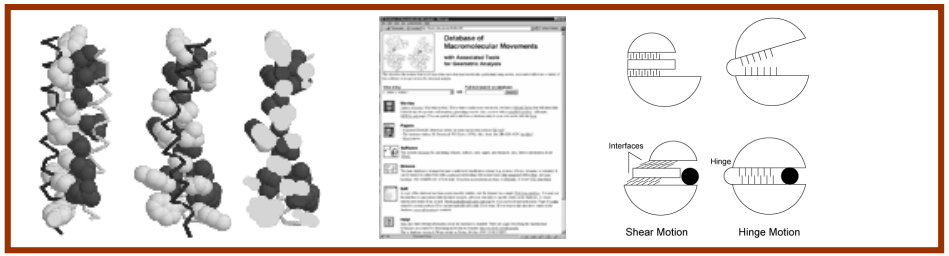
Restrained Interpolation betw. submitted endpoints, std. stats based on 2-core+hinge model, hinge finder

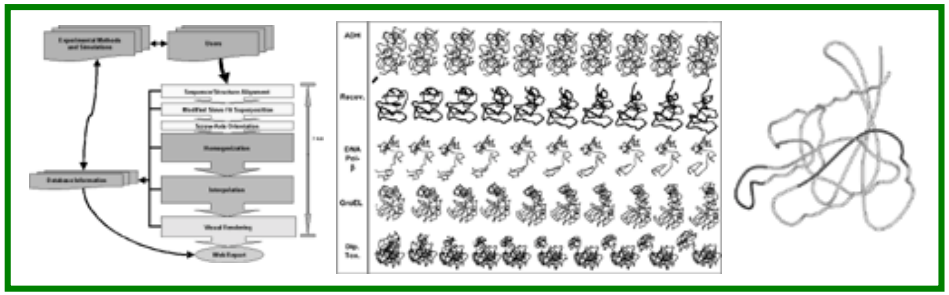
3 Geometry Software (Packing)

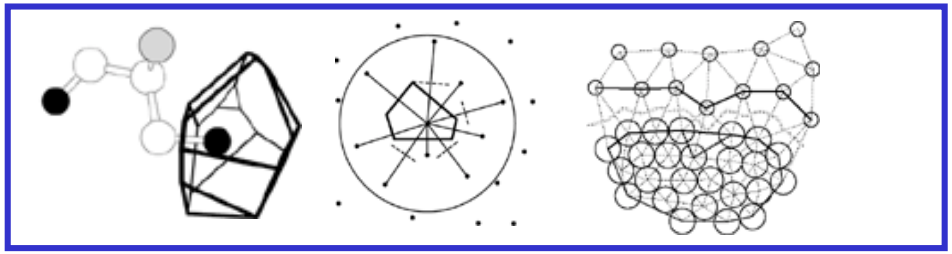
Voronoi volumes, relation to surface, radii problem, packing measurement

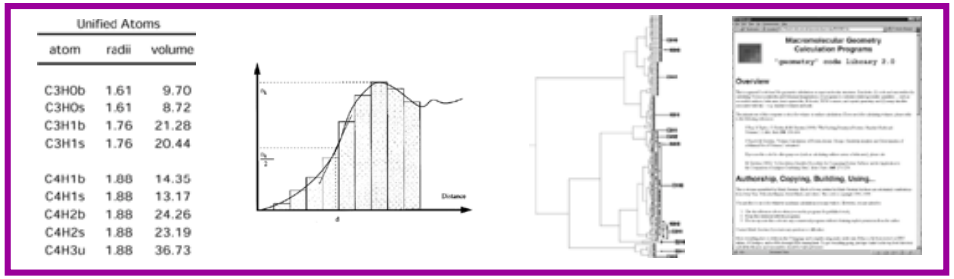
4 Standard Radii & Volumes

Needed for volume calculation, 173 raw to 18 basic types, Sensitivity Analysis, radii set, atom selection criteria, structure set



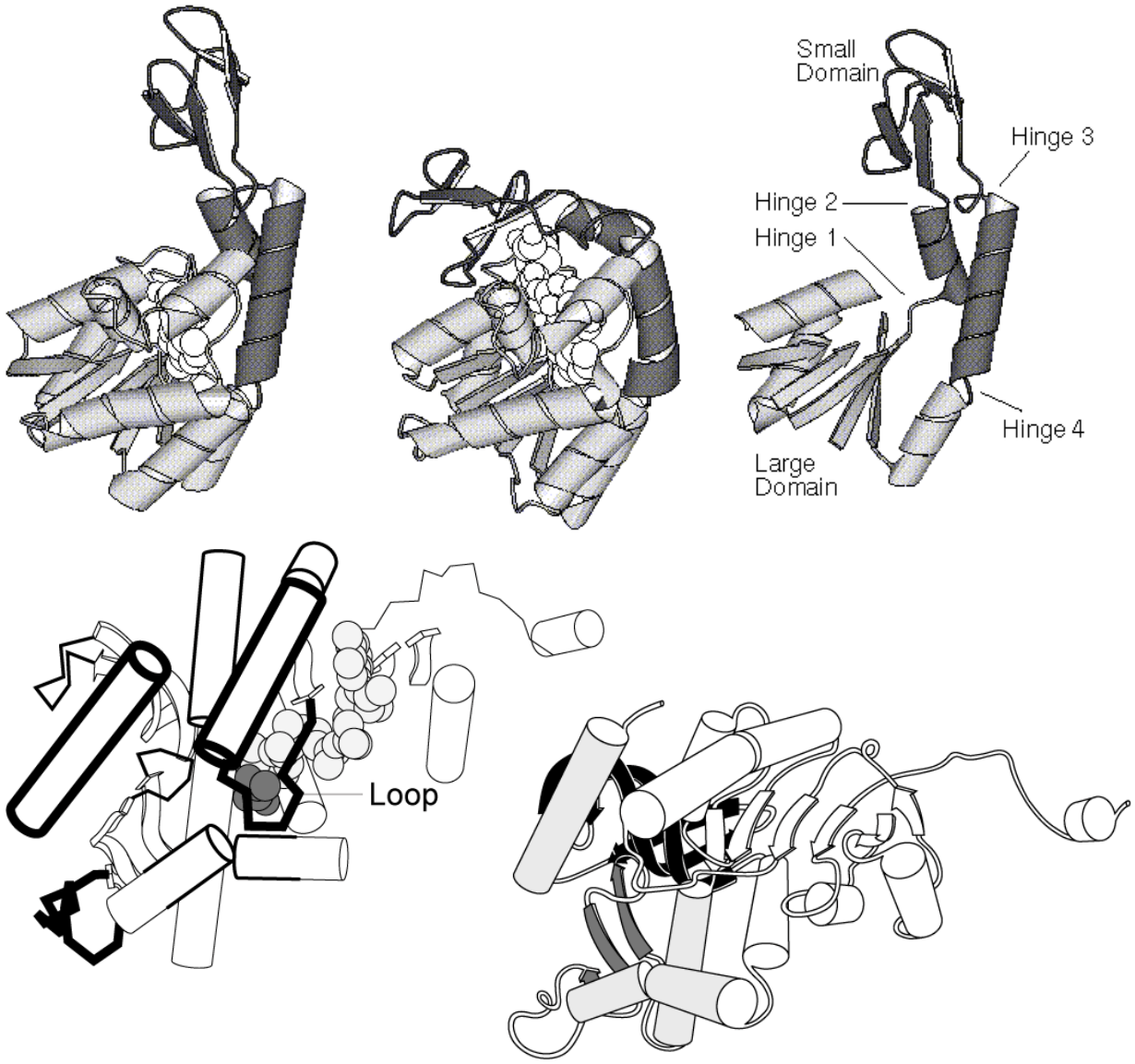


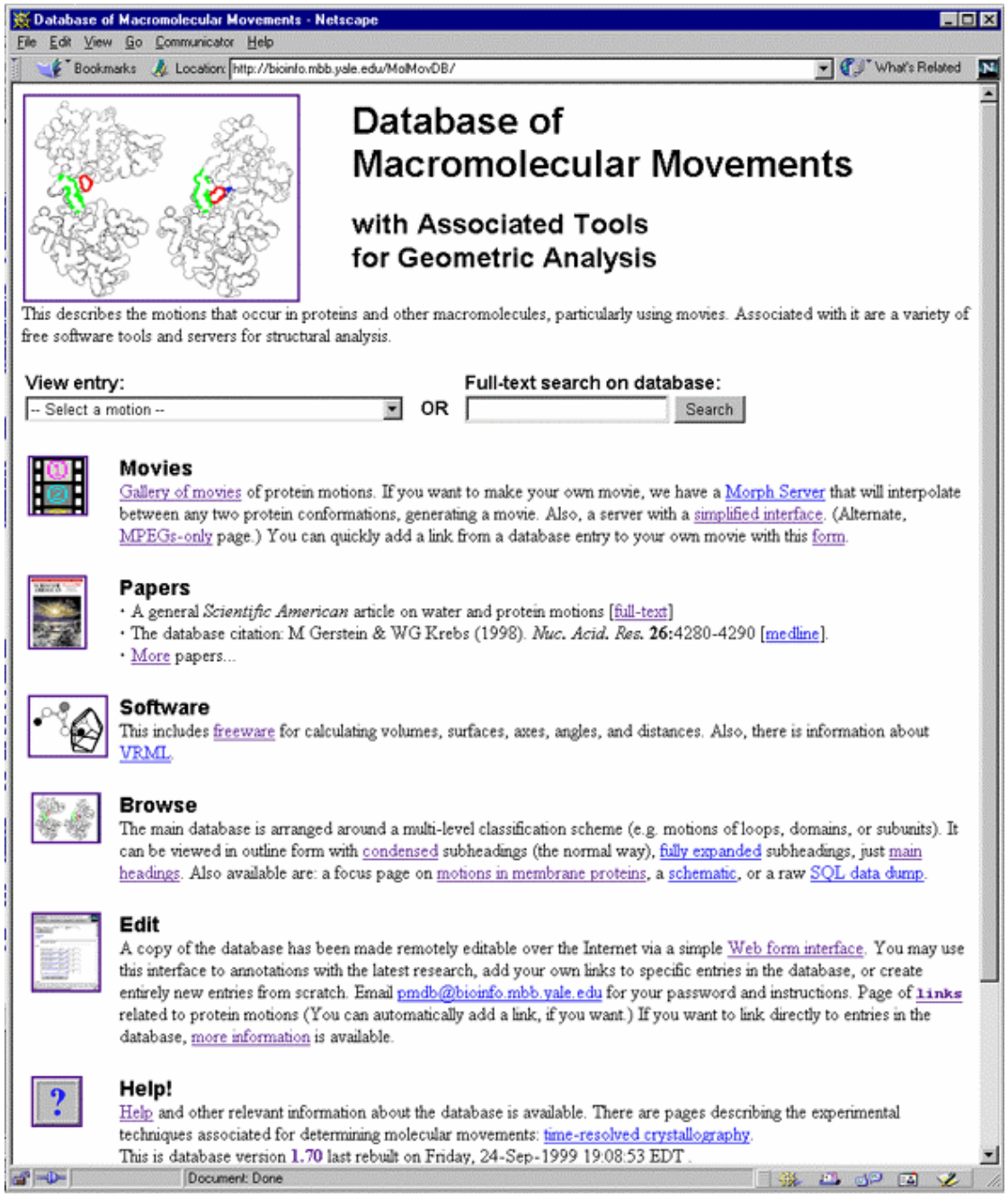


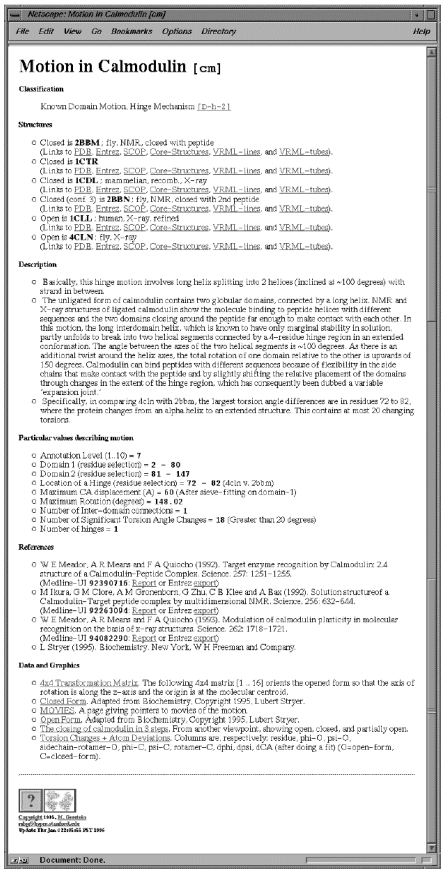


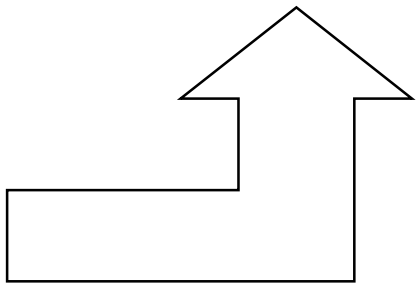
Macromolecular Motions

- What are they?
  - ◊ Proteins, Nucleic Acids (Hammerhead)
  - ◊ Sidechains (trivial), Loops (LDH), Domains (ADK), Subunits (Hb)
  - ◊ When a Ligand Binds: Open, Closed
- Essential link between structure and function
  - ◊ catalysis, regulation, transport, formation of assemblies, and cellular locomotion
- A complicated biological phenomena that can be studied in quantitative detail
  - ◊ changes in thousands of atomic coordinates









Query with Calmodulin

What's in the DB?

Basic

Motion in Calmodulin [cm]

Classification

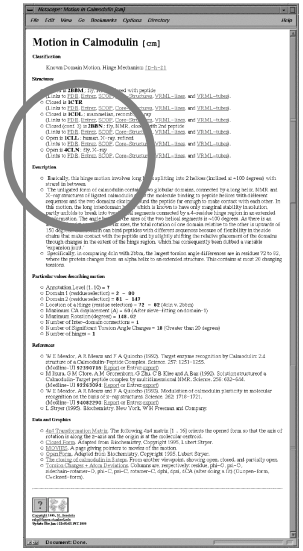
Known Domain Motion, Hinge Mechanism [D-h-2]

Structures

- Closed is **2BBM**; fly, NMR, closed with peptide
  (Links to PDB, Entrez, SCOP, Core-Structures, VRML-lines, and VRML-tubes).
- Closed is **1CTR**
  (Links to PDB, Entrez, SCOP, Core-Structures, VRML-lines, and VRML-tubes).
- Closed is **1CDL**; mammelian, recomb., X-ray
  (Links to PDB, Entrez, SCOP, Core-Structures, VRML-lines, and VRML-tubes).
- Closed (conf. 3) is **2BBN**; fly, NMR, closed with 2nd peptide
  (Links to PDB, Entrez, SCOP, Core-Structures, VRML-lines, and VRML-tubes).
- Open is **1CLN**; human, X-ray, refined
  (Links to PDB, Entrez, SCOP, Core-Structures, VRML-lines, and VRML-tubes).
- Open is **4CLN**; fly, X-ray
  (Links to PDB, Entrez, SCOP, Core-Structures, VRML-lines, and VRML-tubes).

Description

- Basically, this hinge motion involves long helix splitting into 2 helices (inclined at ~100 degrees) with strand in between.
- The unligated form of calmodulin contains two globular domains, connected by a long helix. NMR and X-ray structures of ligated calmodulin show the molecule binding to peptide helices with different sequences and the two domains closing around the peptide far enough to make contact with each other. In this motion, the long interdomain helix, which is known to have only marginal stability in solution, partly unfolds to break into two helical segments connected by a 4-residue hinge region in an extended conformation. The angle between the axes of the two helical segments is ~100 degrees. As there is an additional twist around the helix axes, the total rotation of one domain relative to the other is upwards of 150 degrees. Calmodulin can bind peptides with different sequences because of flexibility in the side



Manual Gold-standard Motions

~150 Different Motion Ids, ~250 PDB identifiers

Submitted + Auto-generated Motions

>5000

PDB id acts as Foreign Key into other DBs

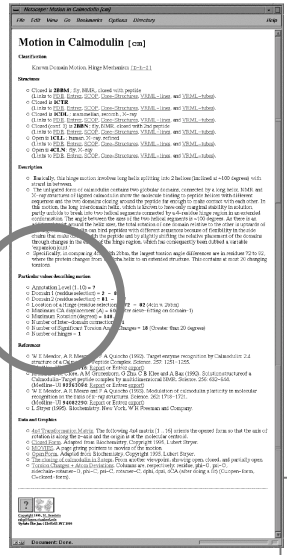
Text Blurb, Literature refs...

~35 Relational Tables

Standardized Terminology

Particular values describing motion

- Annotation Level (1..10) = **7**
- Domain 1 (residue selection) = **2 – 80**
- Domain 2 (residue selection) = **81 – 147**
- Location of a Hinge (residue selection) = **72 – 82** (4cln v. 2bbm)
- Maximum CA displacement (A) = **60** (After sieve-fitting on domain-1)
- Maximum Rotation (degrees) = **148.02**
- Number of Inter-domain connections = **1**
- Number of Significant Torsion Angle Changes = **18** (Greater than 20 degrees)
- Number of hinges = **1**



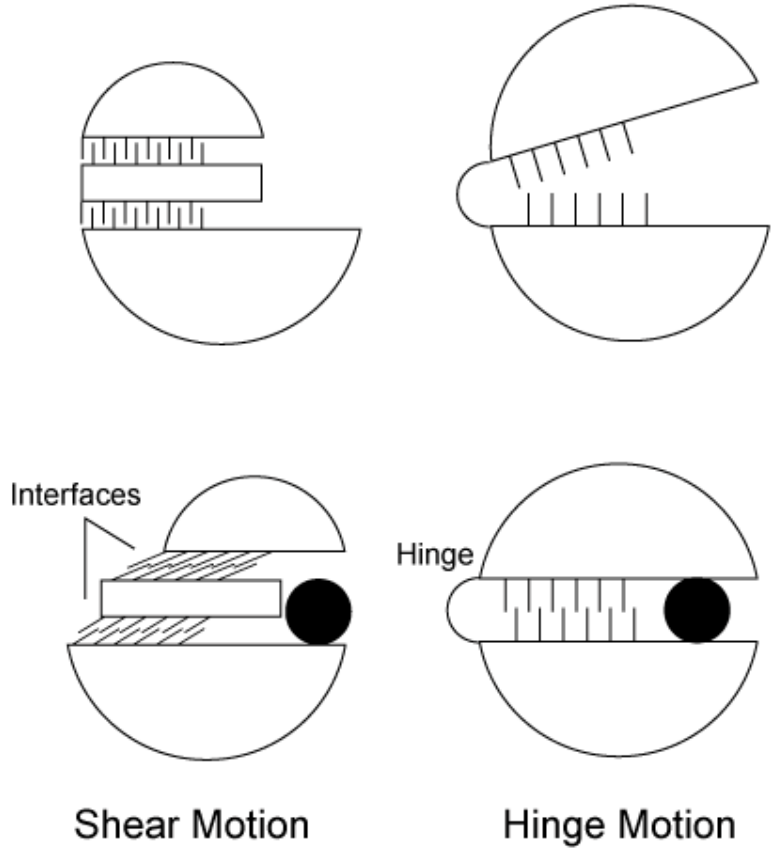
- Standard statistics
  - ◊ torsion angles, max CA disp., &c
- Relations betw. motions
  - ◊ "sim-to", "contains," "Share-characteristics"
- Inferred Motions
  - ◊ 1 structure but "sim-to" another with 2

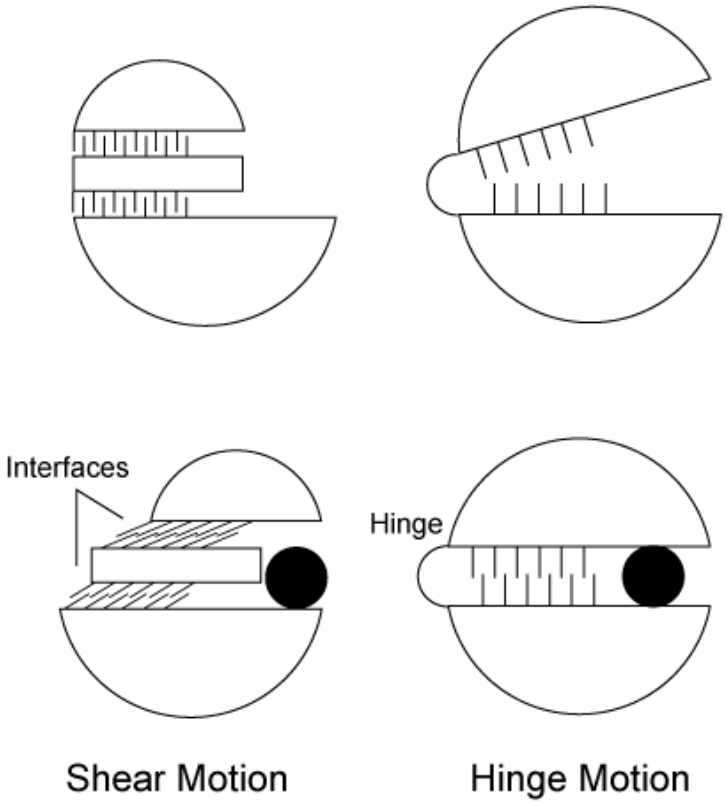
Name	Mean of max	Min of max	Max of max
Maximum Alpha Change	140º	16º	180º
Maximum Phi Change	180º	140º	180º
Maximum Psi Change	150º	23º	180º

		Hand-gathered statistics				Automatically collected motion statistics				
Value		Min	Max	Mean		Min	Max	Mean	Median	Stdev
Maximum Cα displacement (Å)		1.5	60	12		0.90	81	23	17	19
Maximum hinge rotation (º)		5	148	24		0.0	150	35	9.5	46

Information, Size, then Packing Based Classification

Number Known Forms	Size of Motion	Mechanism of Motion	Examples	#
Motion				
2 forms	Fragment	**Hinge**	TIM, LDH, TGL	14
		Shear	Insulin	3
		Unclassifiable	MS2 Coat	3
	Domain	**Hinge**	LF, ADK, CM	16
		Shear	CS, TrpR, AAT	8
		Refold	Serpin, RT	3
		Special	Ig elbow	1
		Unclassifiable	TBP, EF-tu	3
	Subunit	Allosteric	PFK, Hb, GP	4
		Non-allosteric	Ig VL-VH	2
		Unclassifiable		
1 form	Fragment	**Hinge**		
		Shear		
		Unclassifiable	bR	1
	Domain	Refold		
		Hinge	LF~TF,SBP	10
		Shear	HK~PGK,HSP	4
		Special		
		Unclassifiable	Myosin	4
	Subunit	Allosteric		
		Non-allosteric		
		Unclassifiable	PCNA, GroEL	3



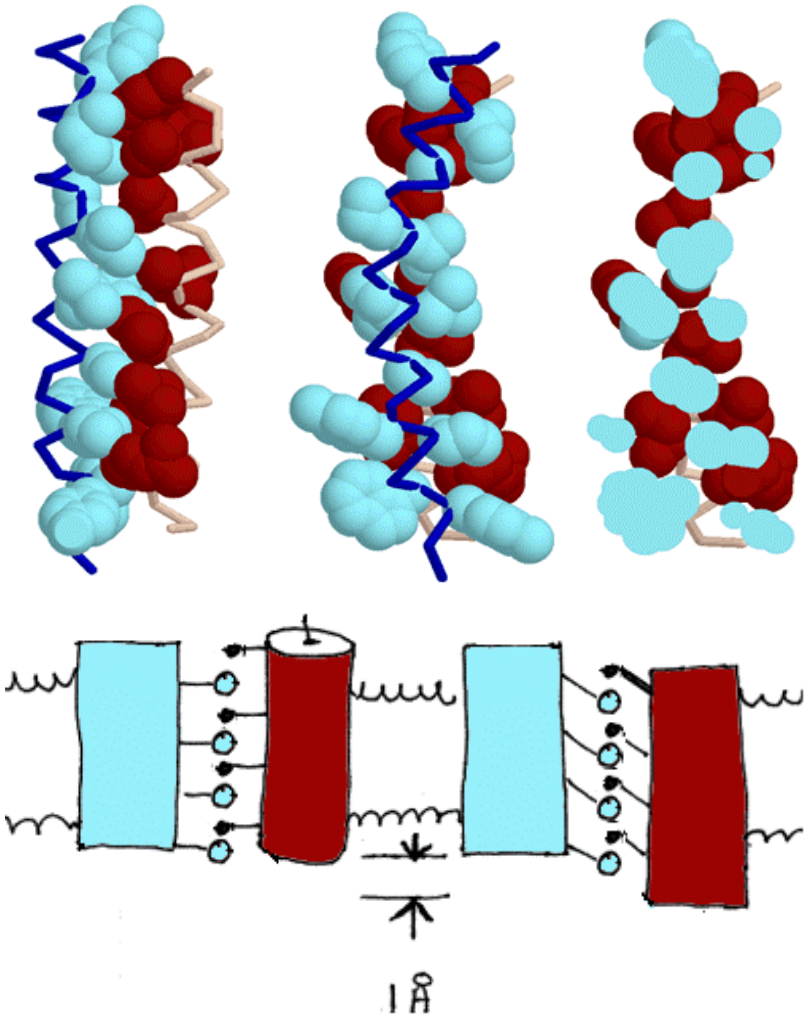


Shear/Hinge Paradigm Is there a maintained Interface?

	Shear Mechanism	**Hinged Mechanism**
Well-Packed Interfaces	**MAINTAINED,** throughout motion	**NOT MAINTAINED;** Rather created, burying surface
Mainchain Packing	Constrained by close packing	Free to kink at hinge
Mainchain Torsions	Many small changes	A few large changes
Motion Overall	Concatenation of small local motions	Identical to twisting at hinge
Motion at Interface	Parallel to plane of interface (shear)	Perpendicular to interface
Sidechain Packing	Same packing in both forms	New contacts; Packing at base of hinge crucial.
Sidechain Torsions	Mostly small changes	Some large changes
Simple Example	Trp Repressor, Insulin	Lactoferrin, Calmodulin

Interface Packing and Motions

- Intercalcating Interface, Knobs into Holes
- Packing is a strong constraint on motions
  - ◊ Domain or loop motions have to be fast (~10 ps – 100 ns)
  - ◊ Can't cross big energy barriers involved in repacking an interface
- Not applicable to allosteric motions, which are much slower (~1 ms) and do involve repacking interfaces

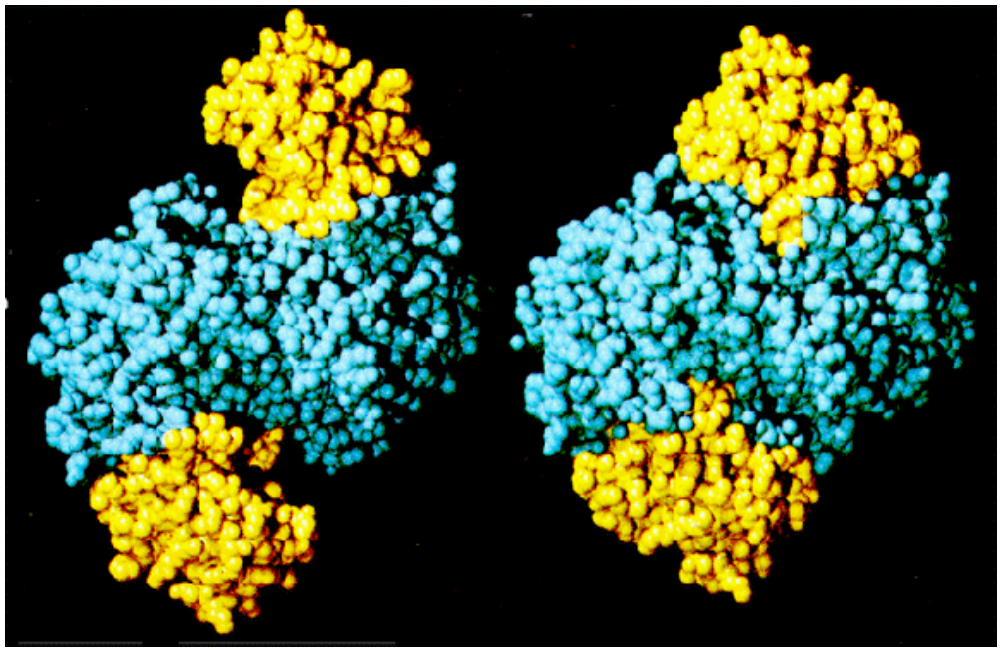


Current State of Gold Standard Manually Curated Motions

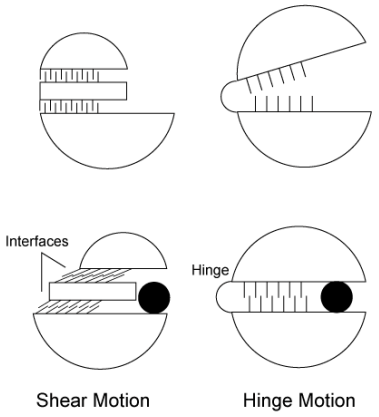
Experimental Technique	Entries studied by this technique	Fraction of database
All Techniques	122	100%
Traditional X-ray crystallography	116	95%
NMR	9	7%
Molecular Dynamics Simulations	4	3%
Time-resolved crystallography	3	2%
Circular Dichroism (CD)	2	2%
Fourier Transform Infrared Spectroscopy (FTIR)	1	<1%
Molecular Biology Studies of Motion	1	<1%

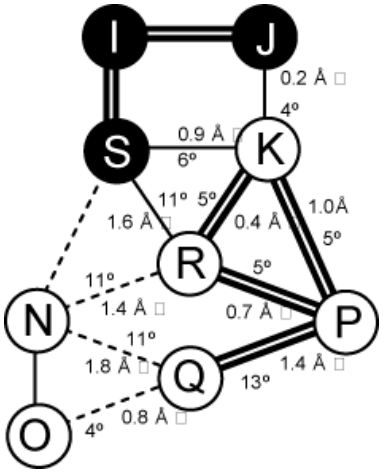
Size / Mechanism	Domain	Fragment	Subunit	Complex	Total
Hinge	38 51%	16 59%			54 44%
Shear	14 19%	3 11%			17 14%
Partial Refolding	5 7%				5 4%
Allosteric			8 57%		8 7%
Other/Non-Allosteric	2 3%	1 4%	6 43%		9 7%
Unclassifiable	15 20%	7 26%		3 50%	25 20%
Notably Motionless					1 1%
Nucleic Acid				3 50%	3 2%
Known** / %category	53 72%	25 93%	11 79%	5 83%	94 77%
Suspected / %category	21 28%	2 7%	3 21%	1 17%	28 23%
Totals / %DB	74 61%	27 22%	14 11%	6 5%	122

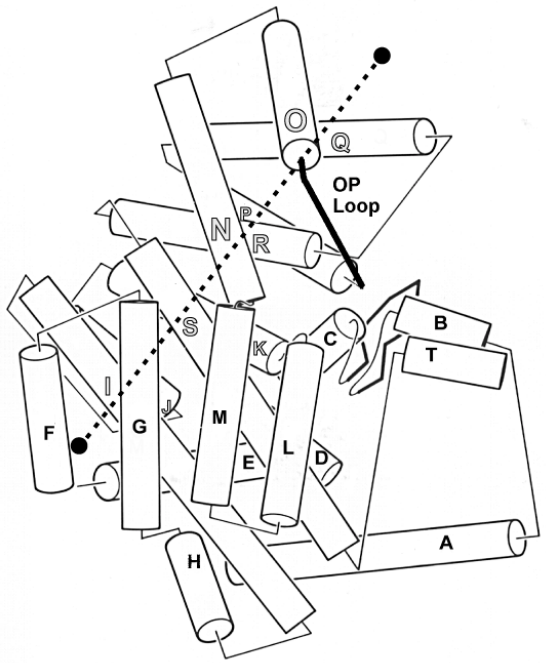
Motion in Citrate Synthase

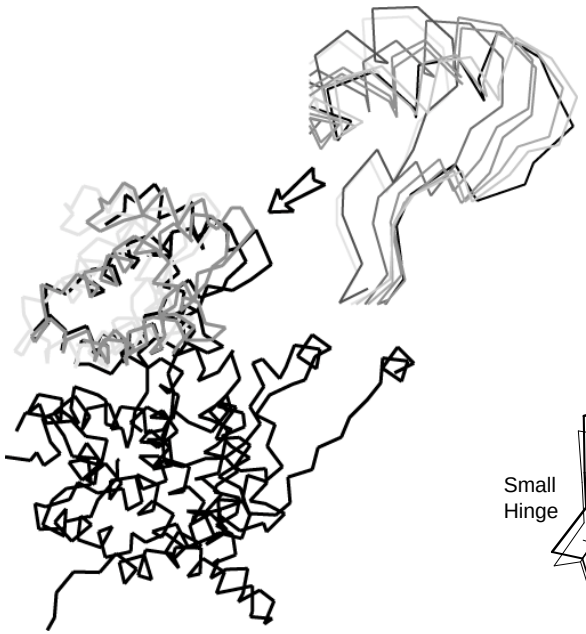


Packing Based Classification: Hinge v ***Shear***

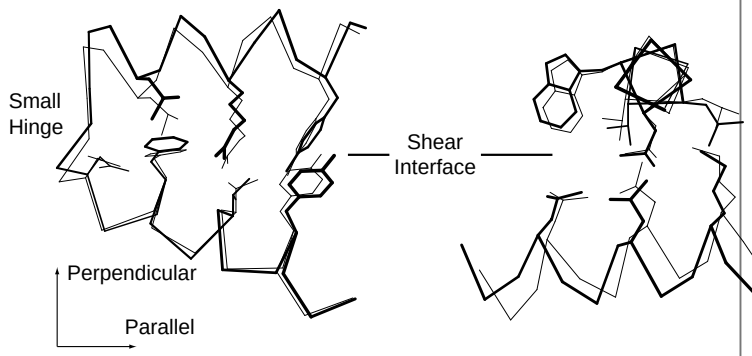




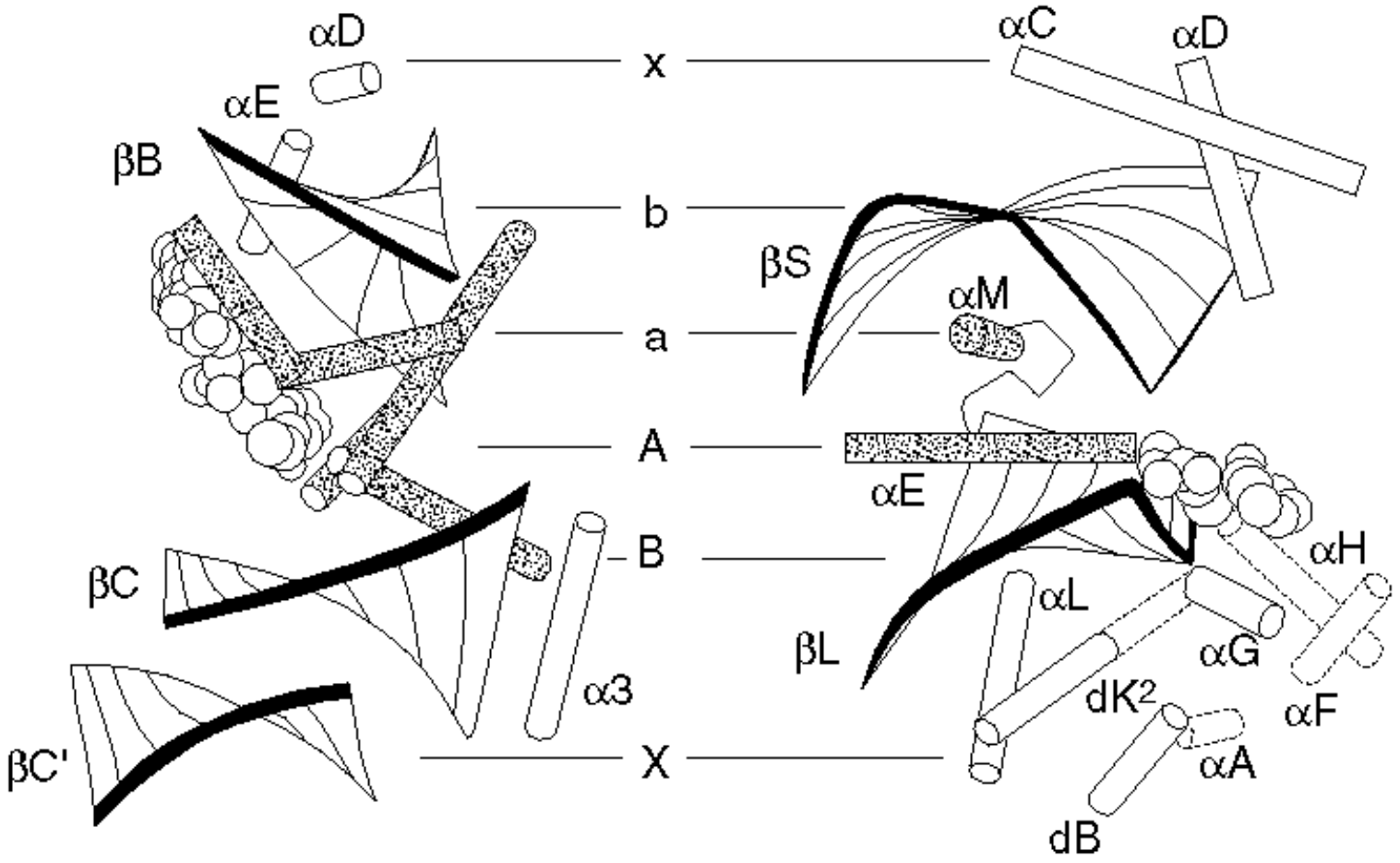




Shear Mechanism Involves Many Small Motions across a Continuously Maintained Interface



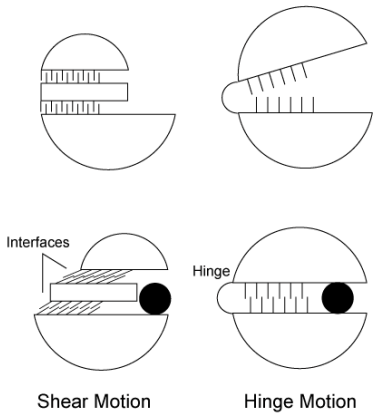
Proteins With Shear Motions are Often Divided into Layers



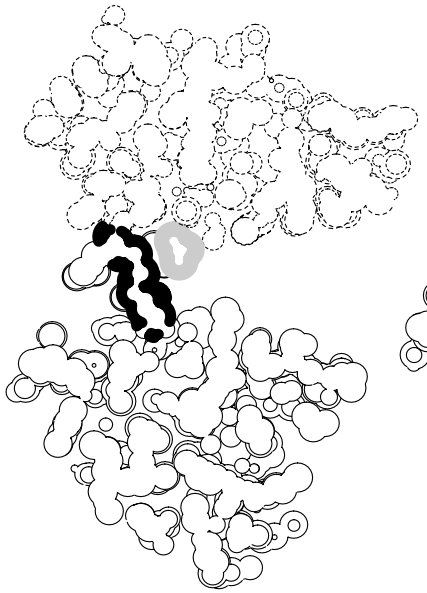
GAPDH

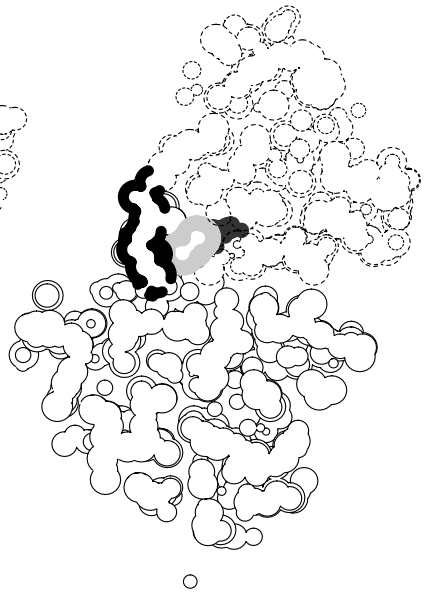
Hexokinase

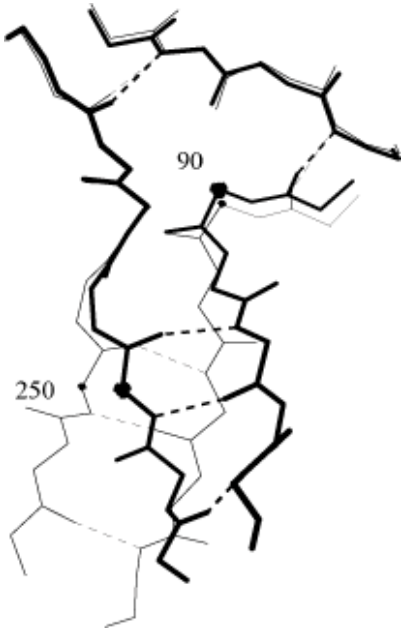
Packing Based Classification: ***Hinge*** v Shear



Hinge Mechanism involves absence of steric constraints (continuously maintained interface), esp. at hinge

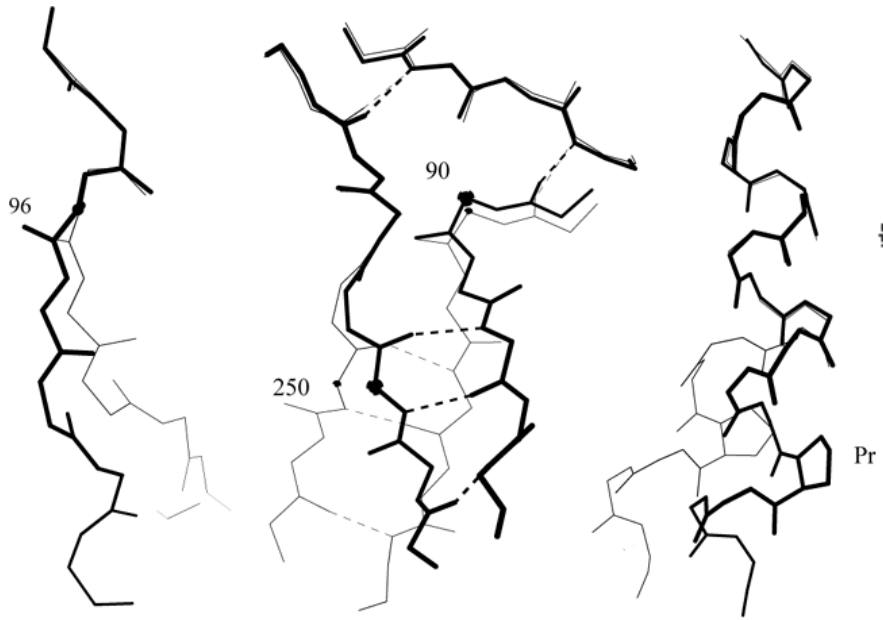


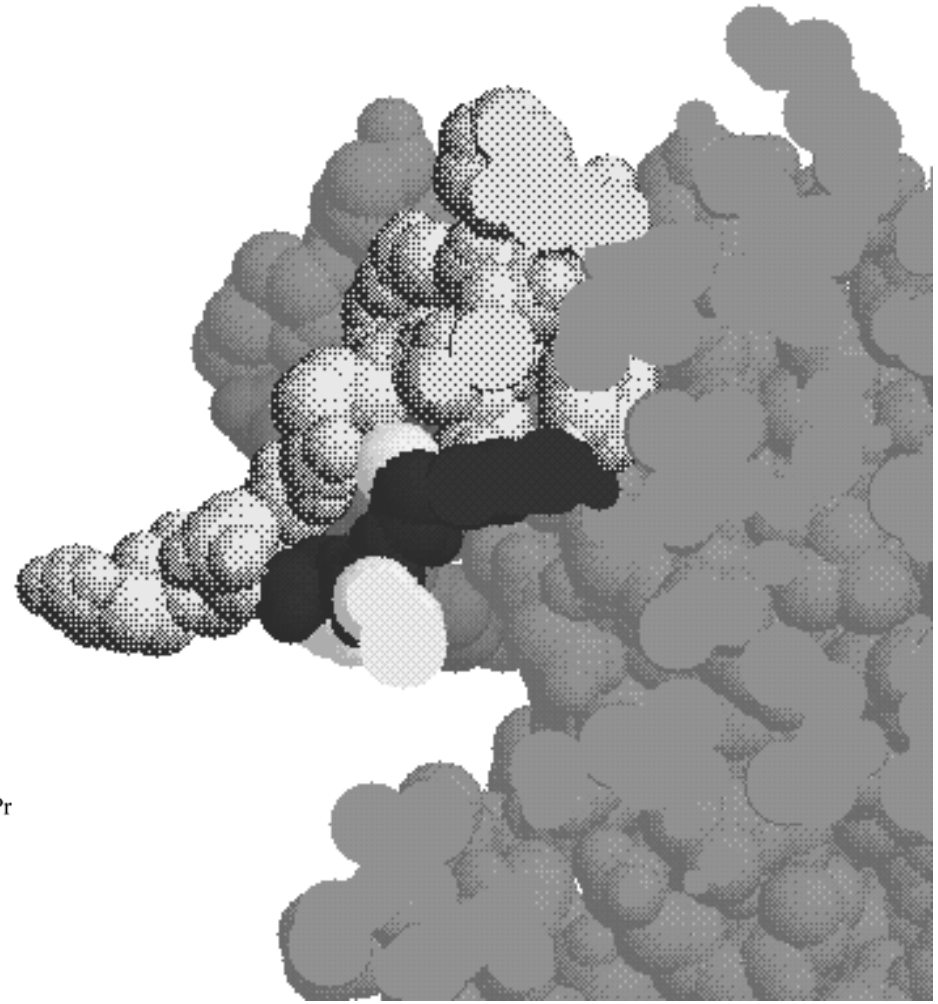




Absence of Tight Packing at Hinge

Chain Topology is not important





Analysis of Macromolecular Geometry & Motions

W Krebs, C Chothia,
J Tsai, Y Harpaz, R Taylor

bioinfo.mbb.yale.edu/MolMovDB

1 Database of Macromolecular Motions

Interface packing, shear v. hinge, web DB

2 Morph Movie Server

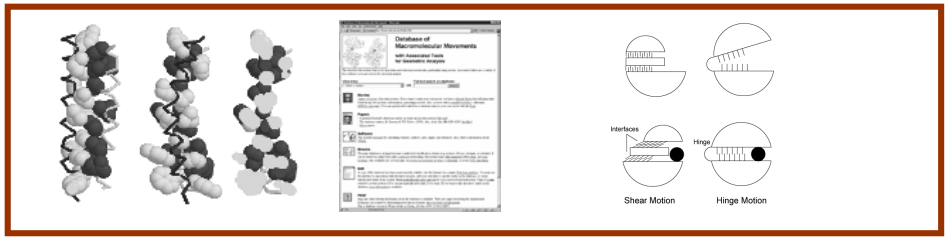
Restrained Interpolation between submitted endpoint conformations, standardized stats, hinge finder

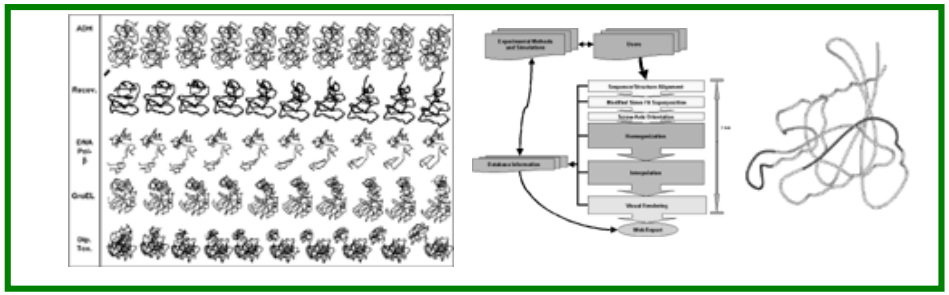
3 Geometry Software (Volumes)

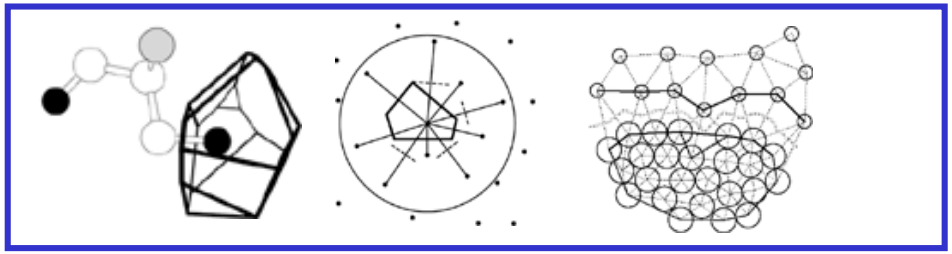
Voronoi volumes, relation to surface, radii problem, packing measurement

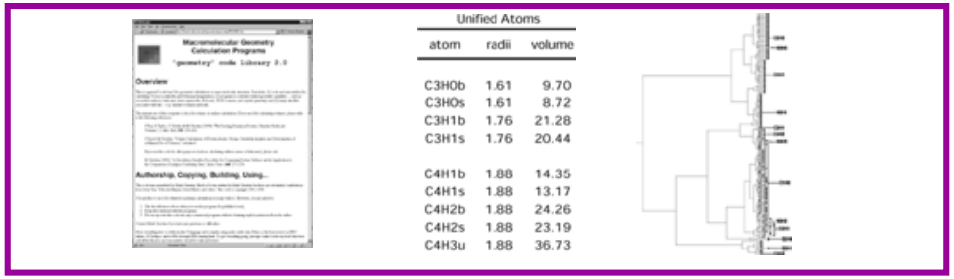
4 Standard Radii & Volumes

Needed for volume calculation, 18 basic types, Sensitivity Analysis, radii set, atom selection criteria, structure set

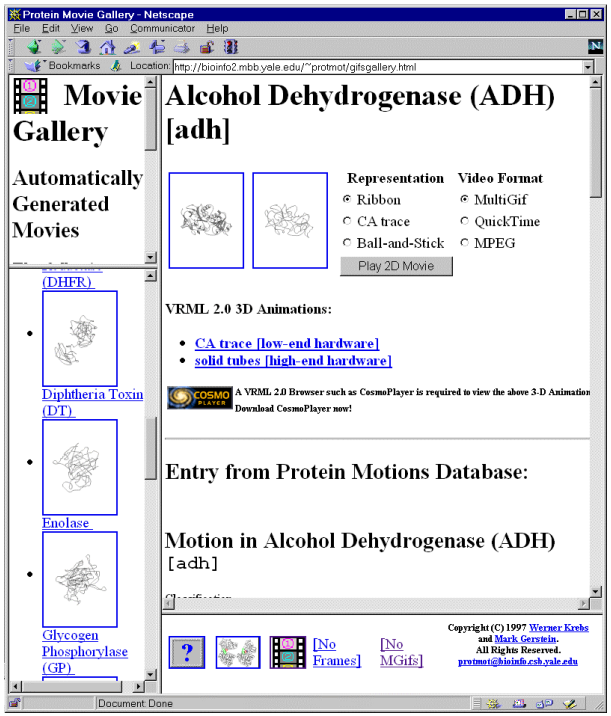


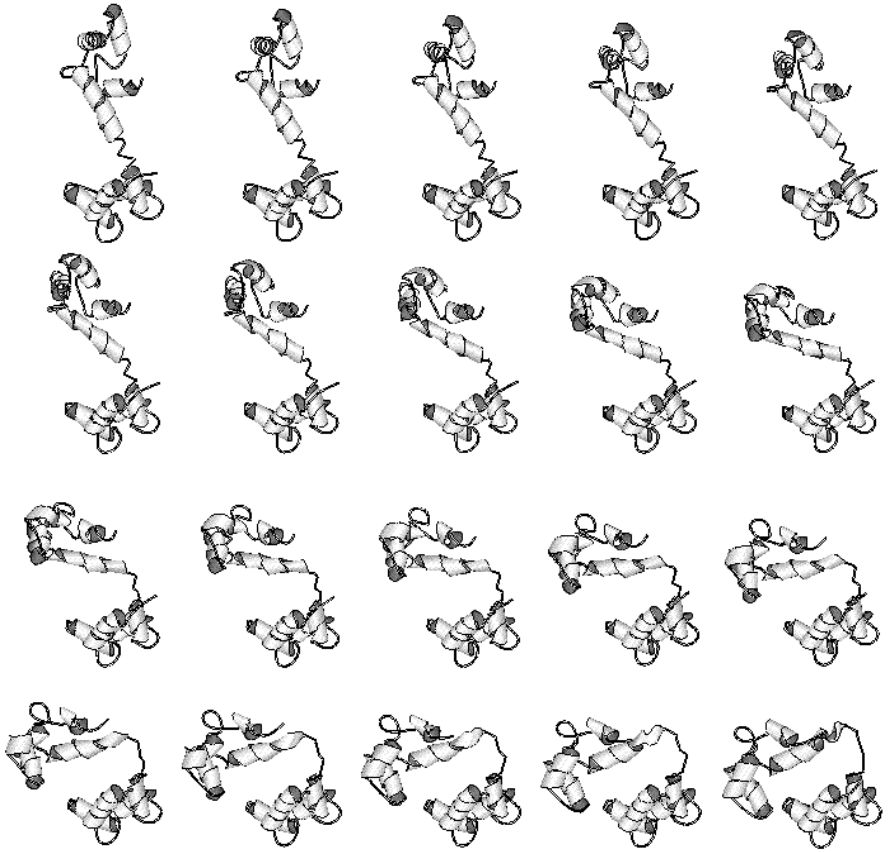






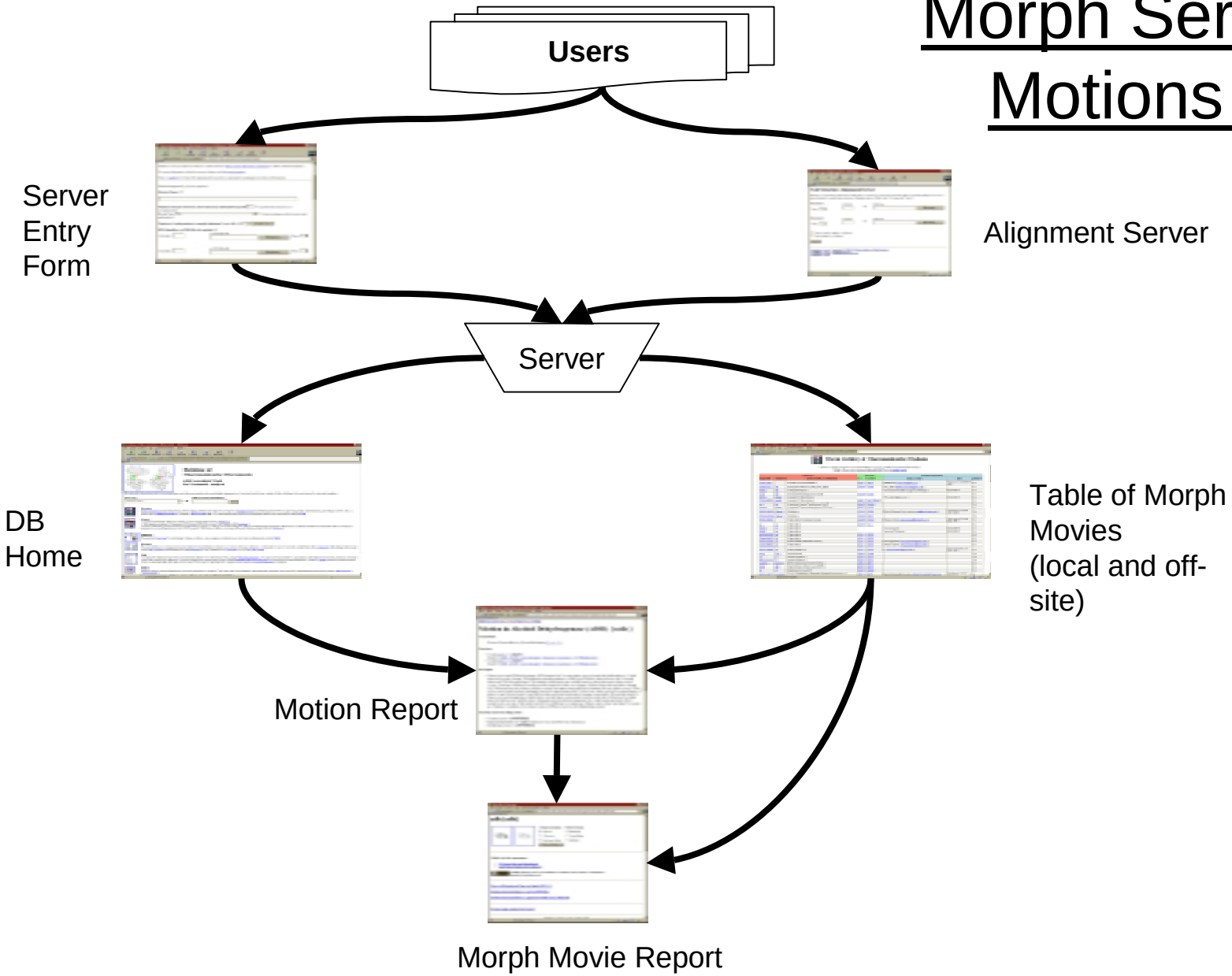
"Morph" Movies of Protein Motions



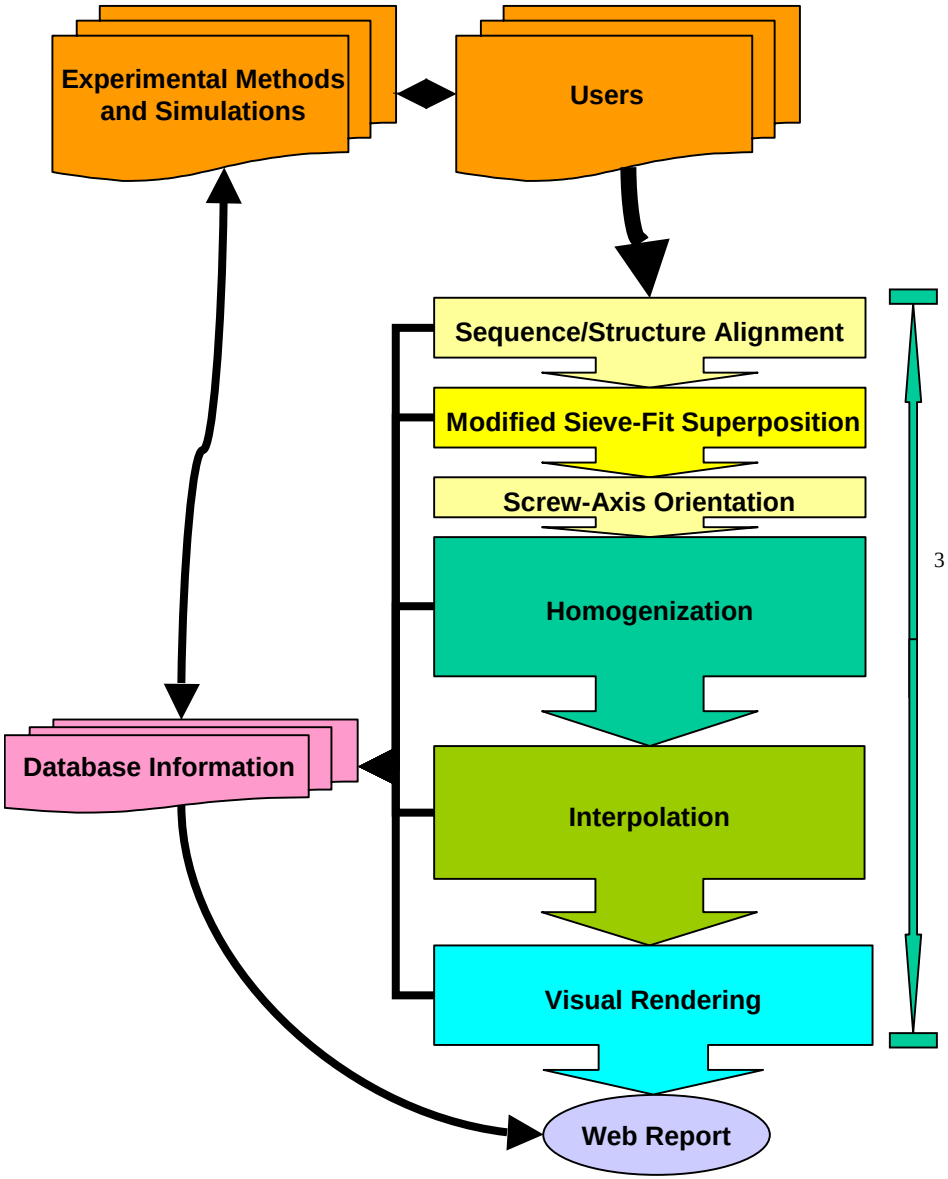


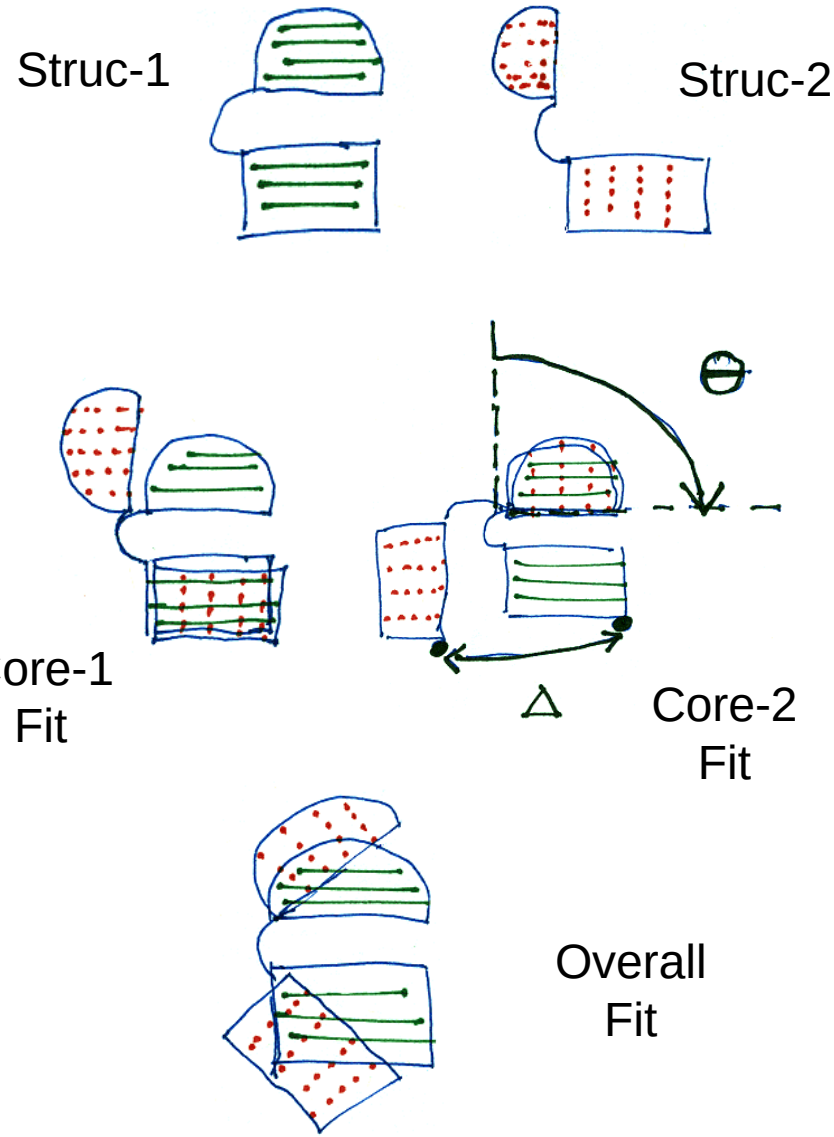
Server Produces Semi-realistic Minimized Interpolation (as MPEG, VRML, &c) between Any 2 Aligned Conformations, Analyzes the Motion

Relationship of Morph Server to Motions DB

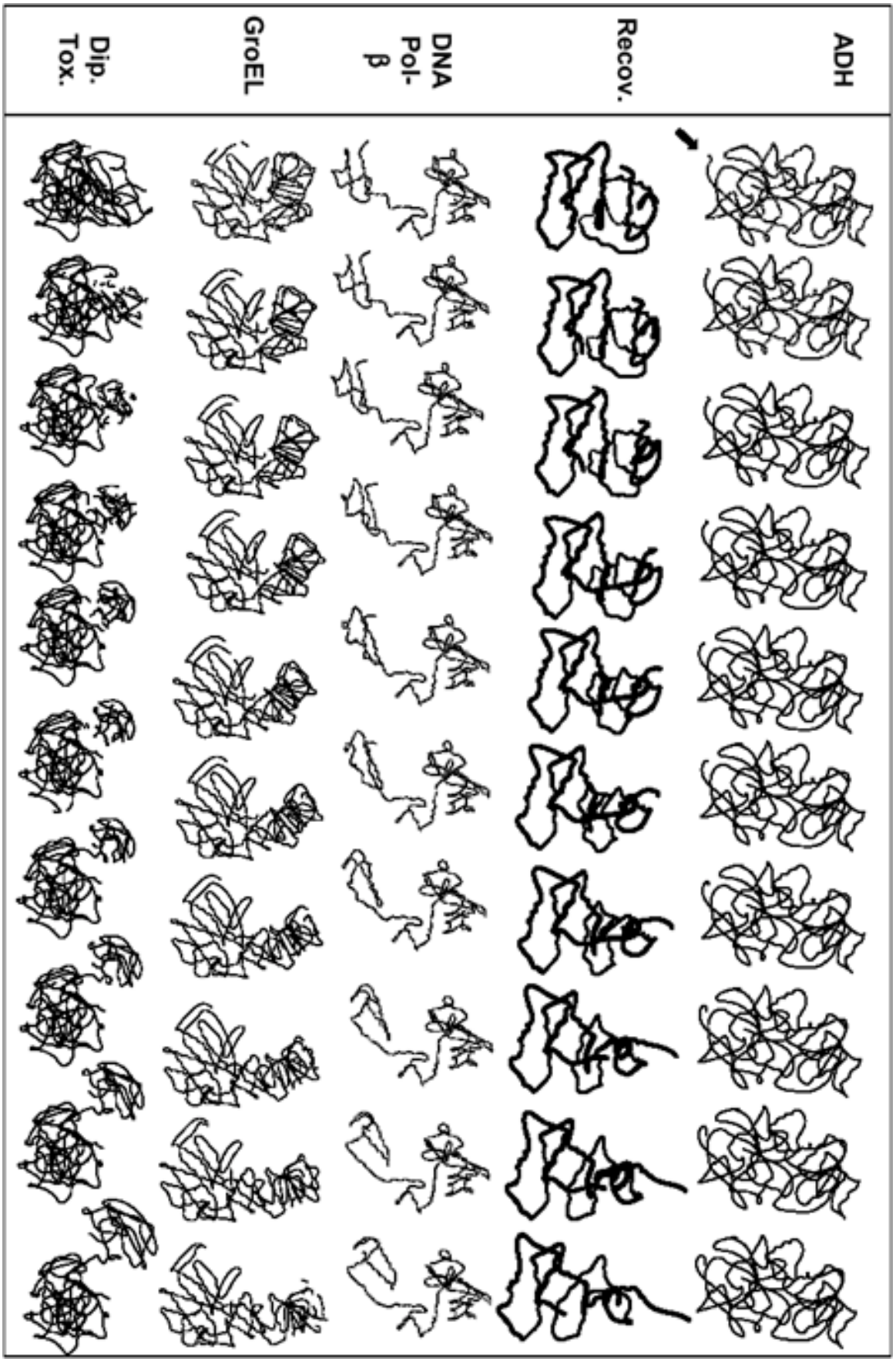


Server Flowchart: Steps in Interpolation, 2 Core Model



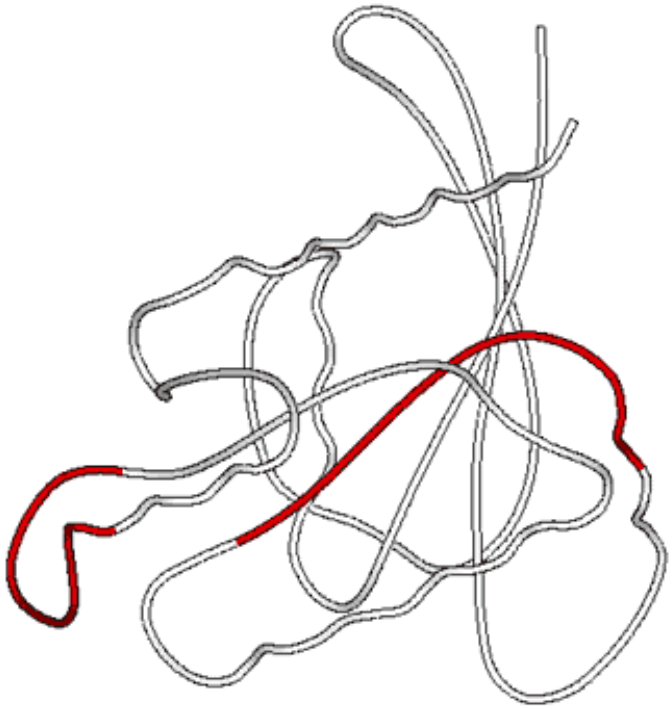


5 Movies Generated by the Server



Results of the Server, Statistics & Hinges

	Statistic	[Code]	Easy	Typical		Large	Impossible
			ADH	Recoverin	DNA Pol-Beta	GroEL	Diptheria Toxin
Input Structures	Motion ID	[ID]	adh	recvin	polbeta	groel	dt
	1st input frame	[inputframe0]	8ADH	1IKU	1BPD	1GRL	1DDT
	2nd input frame	[inputframe1]	6ADH	1JSA	2BPF	1AON	1MDT
	Size (Å) (in terms of window for rendering)	[max_x_or_y]	36	41	52	55	39
	Number of atoms	[natoms]	2887	1639	2697	3993	4110
	Number of residues	[nresidues]	374	201	335	548	535
Overall Motion	Overall RMS between first and last frames	[RMSoverall]	2.0	13	8.6	16	20
	Rotation (degrees)	[kappa]	4.9°	73°	9.9°	62°	62°
	Overall translation of centroid (Å)	[translation]	2.1	13	6.1	47	66
	X translation (Å)	[TransX]	1.1	-0.24	0.94	45	-45
	Y " .	[TransY]	-0.95	-9.14	4.1	-2.1	-0.54
	Z " .	[TransZ]	1.5	-9.78	-4.4	-10	48
1st Core	Number Cα's in 1st core	[AlignedCoreCAs]	187	95	160	259	262
	RMS of 1st core (Å)	[AlignedCoreRMS]	0.40	3.0	0.92	1.4	0.37
	Max Cα displacement in 1st Core (Å)	[MaxCore Deviation]	0.66	7.6	1.7	4.2	0.60
2nd Core	Num. Cα's in 2nd core	[2ndCoreCAs]	190	94	160	260	260
	RMS of 2nd core (Å)	[2ndCoreRMS]	2.9	18	12	23	29
	Max Cα displacement in 2nd core (Å)	[Max2ndCore Deviation]	7.1	38	28	49	60
	RMS of 2nd core (Å) after fitting on 1st core	[2ndCoreRMS postrefitting]	1.6	11	11	10	18
Hinge	Number of putative hinges detected	[NHinges]	0	0	0	1	1
	X position of 1st hinge (Å) rel. to centroid	[Hinge000X]	-	-	-	-4.7	-7.2
	Y position "	[Hinge000Y]	-	-	-	11	-0.91
	Z position "	[Hinge000Z]	-	-	-	3.3	-3.0
	1st Hinge Residue Selection	[Hinge000res]	-	-	-	380:403	352:375
	Sequence of 1st putative hinge	[Hinge000seq]	-	-	-	EVEMKE KKARVE DALHAT RAAVEE	INLFQV VHNSYN RPAYSP GHKTQP
Screw Axis	Distance betw. screw-axis (x0) & centroid (Å)	[x0ToCentroid Distance]	21	8.4	23	30	39
	X displacement centroid from screw axis (Å)	[x0X]	-0.16	-0.5	-2.5	17	-20
	Y "	[x0Y]	-5.0	-6.2	-5.2	-16	-24
	Z "	[x0Z]	-20	5.7	-22	19	-24
	Distance between screw axis and 1st hinge (Å)	[Hinge000x0dist]	-	-	-	26	45
Torsion Angles	Max phi change (Max of Abs. degrees, 0°-180°)	[MaxPhi]	180°	180°	180°	180°	180°
	Max psi change	[MaxPsi]	180°	180°	180°	180°	170°
	Max alpha change	[MaxAlpha]	150°	180°	180°	180°	170°



Analysis of Macromolecular Geometry & Motions

W Krebs, C Chothia,
J Tsai, Y Harpaz, R Taylor

bioinfo.mbb.yale.edu/MolMovDB

1 Database of Macromolecular Motions

Interface packing, shear v. hinge, web DB

2 Morph Movie Server

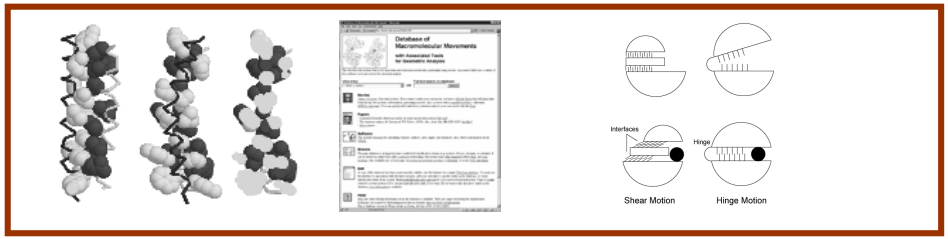
Restrained Interpolation between submitted endpoint conformations, standardized stats, hinge finder

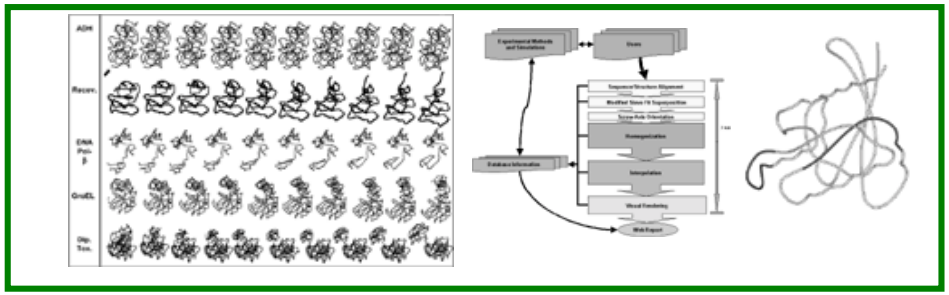
3 Geometry Software (Volumes)

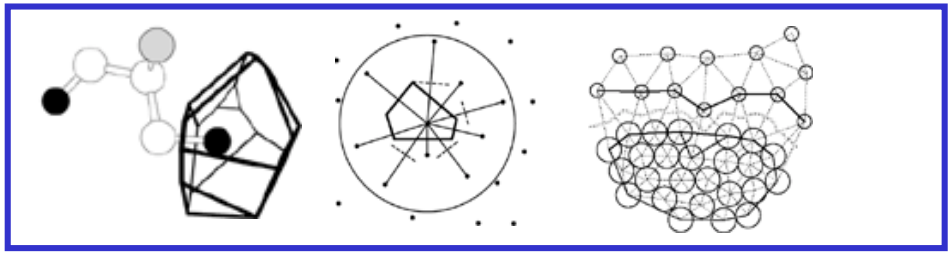
Voronoi volumes, relation to surface, radii problem, packing measurement

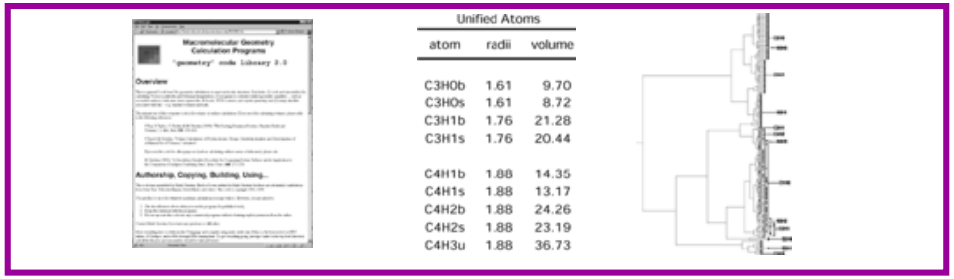
4 Standard Radii & Volumes

Needed for volume calculation, 18 basic types, Sensitivity Analysis, radii set, atom selection criteria, structure set



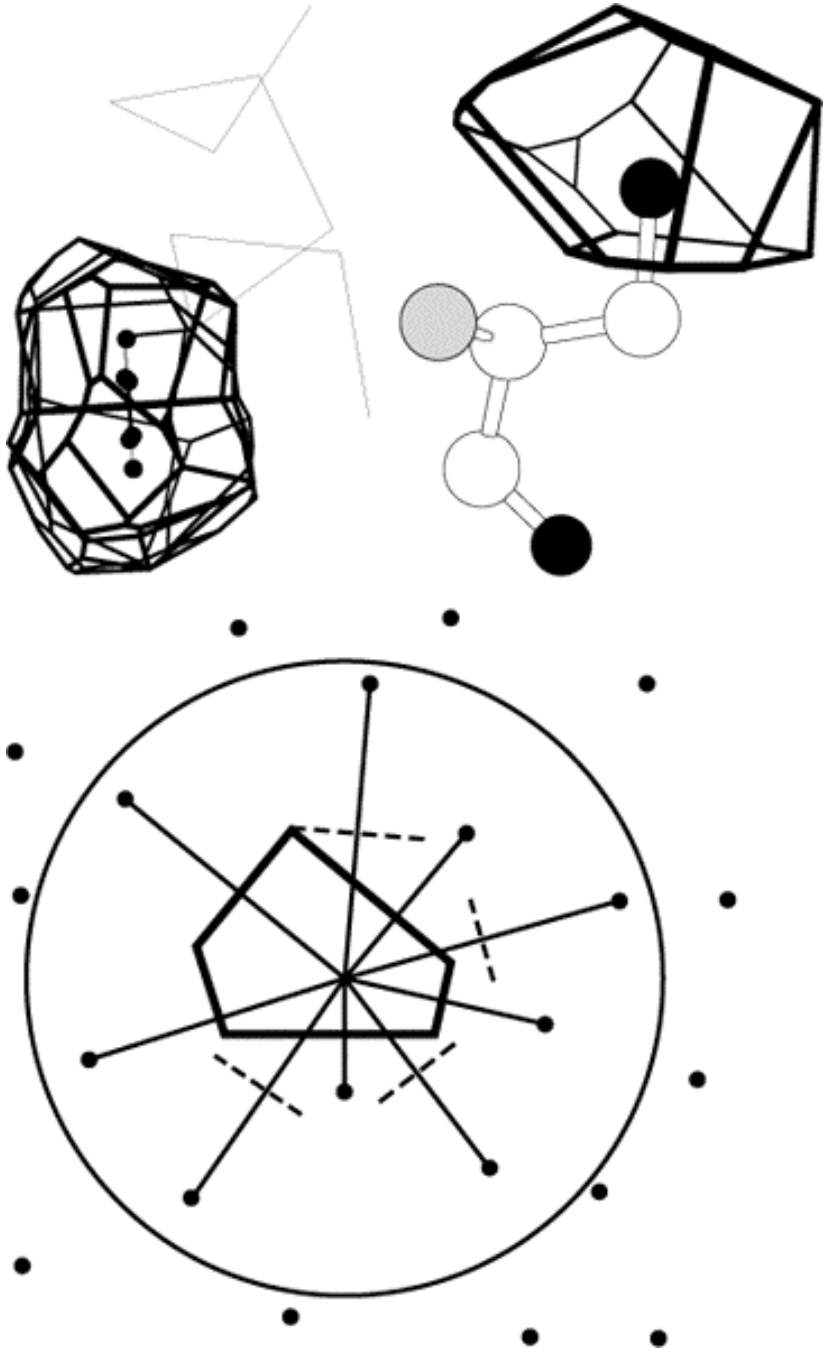






Quantify Packing and Contacts with Voronoi Polyhedra

- Each atom surrounded by a single convex polyhedron and allocated space within it
  - ◊ Allocation of all space (large V implies cavities)
- 2 methods of determination
  - ◊ Find planes separating atoms, intersection of these is polyhedron
  - ◊ Locate vertices, which are equidistant from 4 atoms



Calculating Volumes with Voronoi polyhedra

- In 1908 Voronoi found a way of partitioning all space amongst a collection of points using specially constructed polyhedra. Here we refer to a collection of "atom centers" rather than "points."
- In 3D, each atom is surrounded by a unique limiting polyhedron such that all points within an atom's polyhedron are closer to this atom than all other atoms.
- Likewise, points equidistant from 2 atoms form planes (lines in 2D). Those equidistant from 3 atoms form lines, and those equidistant form 4 centers form vertices.

Voronoi Volumes, the Natural Way to Measure Packing

Packing Efficiency

= Volume-of-Object

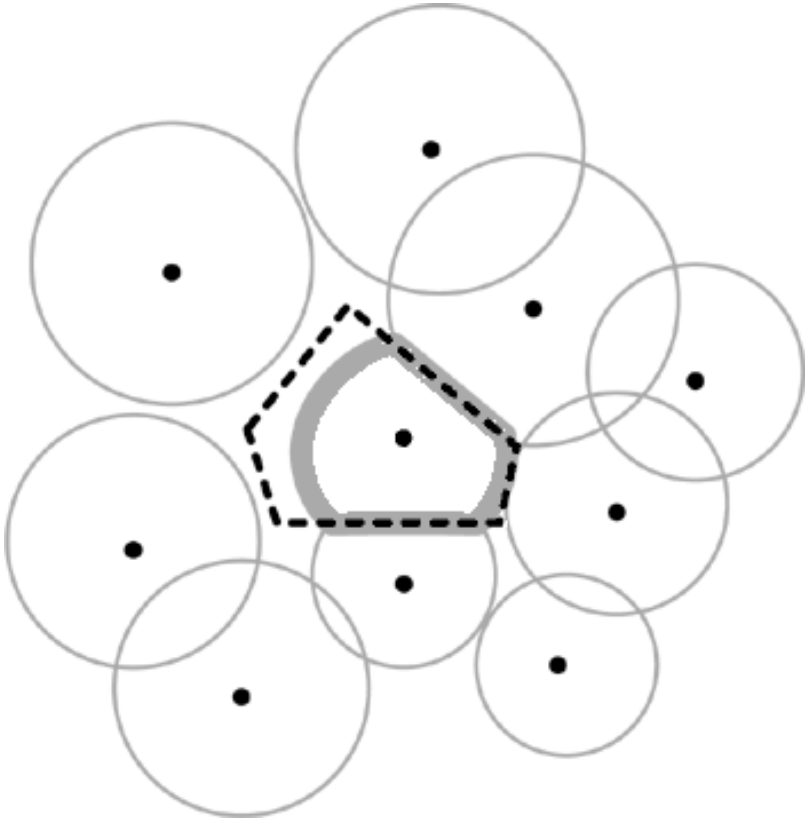
Space-it-occupies

= V(VDW) / V(Voronoi)

- Absolute v relative eff.

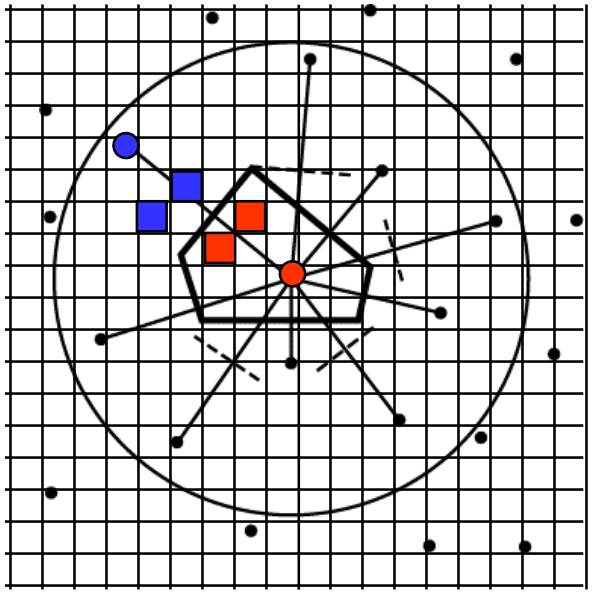
  V1 / V2

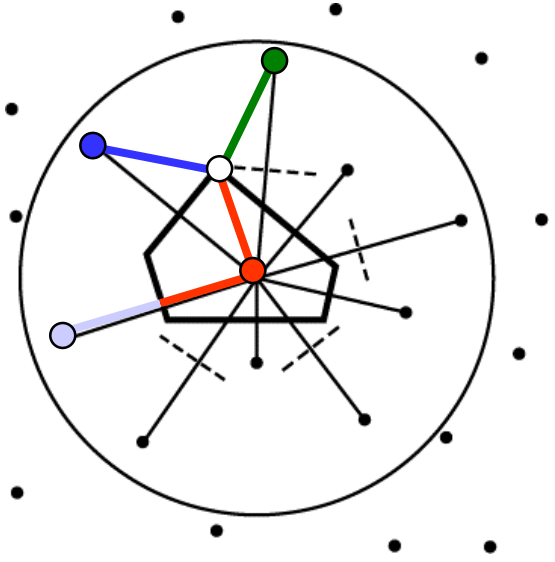
- Other methods
  - ◊ Measure Cavity Volume (grids, constructions, &c)



Calculating Voronoi Volumes

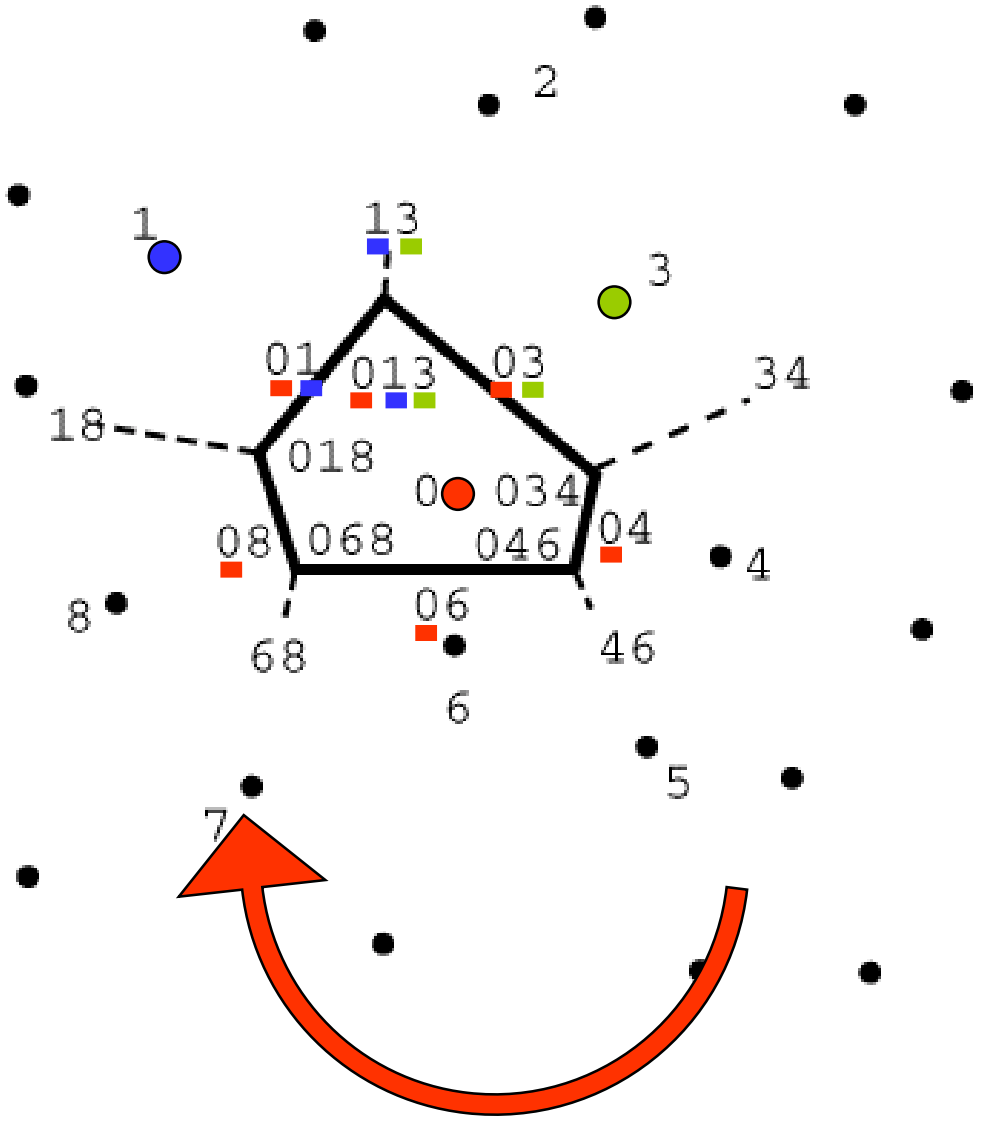
- Integrating on a Grid
  - ◊ The simplest method for calculating volumes with Voronoi polyhedra is to put all atoms in the system on a fine grid. Then go to each grid-point (i.e., voxel) and add its infinitesimal volume to the atom center closest to it. This is prohibitively slow for a real protein structure, but it can be made somewhat faster by randomly sampling grid-points. It is, furthermore, a useful approach for high-dimensional integration.
- Solving for the Vertices
  - ◊ In the basic Voronoi construction, each atom is surrounded by a unique limiting polyhedron such that all points within an atom's polyhedron are closer to this atom than all other atoms. Points equidistant from 2 atoms lie on a dividing plane; those equidistant from 3 atoms are on a line, and those equidistant from 4 centers form a vertex.
  - ◊ It is straightforward to solve for possible vertex coordinates using the equation of a sphere. (That is, one uses four sets of coordinates (x,y,z) and the equation (x-a) 2 + (y-b) 2 + (z-c) 2 = r2 to solve for the center (a,b,c) and radius (r) of the sphere.) One then checks whether this putative vertex is closer to these four atoms than any other atom; if so, it is a real vertex.



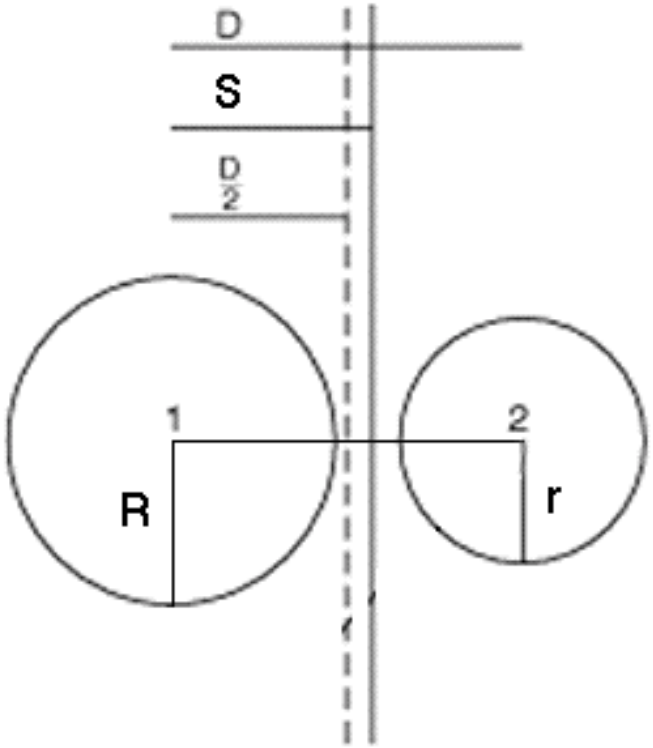


Collecting Vertices & Calculating Volumes

- To systematically collect the vertices associated with an atom, label each one by the indices of the four atoms with which it is associated. To traverse the vertices on one face of a polyhedron, find all vertices that share two indices and thus have two atoms in common — e.g., a central atom (atom 0) and another atom (atom 1). Arbitrarily pick a vertex to start and walk around the perimeter of the face. One can tell which vertices are connected by edges because they will have a third atom in common (in addition to atom 0 and atom 1). This sequential walking procedure also provides a way to draw polyhedra on a graphics device. More importantly, with reference to the starting vertex, the face can be divided into triangles, for which it is trivial to calculate areas and volumes.



Atoms have different sizes



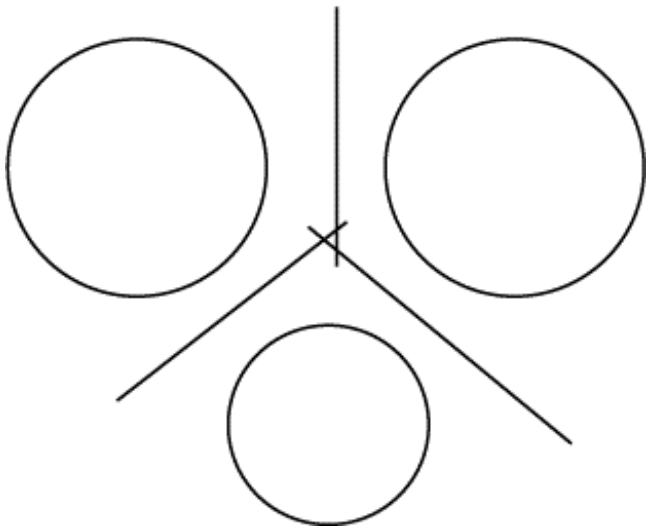
- Difficulty with Voronoi Meth.
  Not all atoms created equal
- Solutions
  - ◊ Bisection -- plane midway between atoms
  - ◊ Method B (Richards)
    Positions the dividing plane according to ratio
  - ◊ Radical Plane
- VDW Radii Set

ProtOr

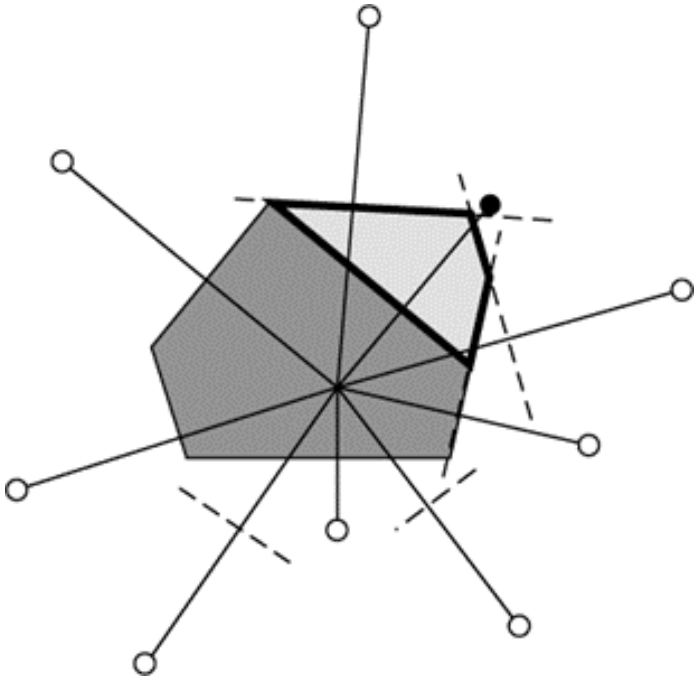
type	radii[1]
C3H0	1.61
C3H1	1.76
C4H1	1.88
C4H2	1.88
C4H3	1.88
N3H0	1.64
N3H1	1.64
N3H2	1.64
N4H3	1.64
O1H0	1.42
O2H1	1.46
S2H0	1.77
S2H1	1.77

Complexity from different atom sizes requires new ways to calculate polyhedra

Vertex Error

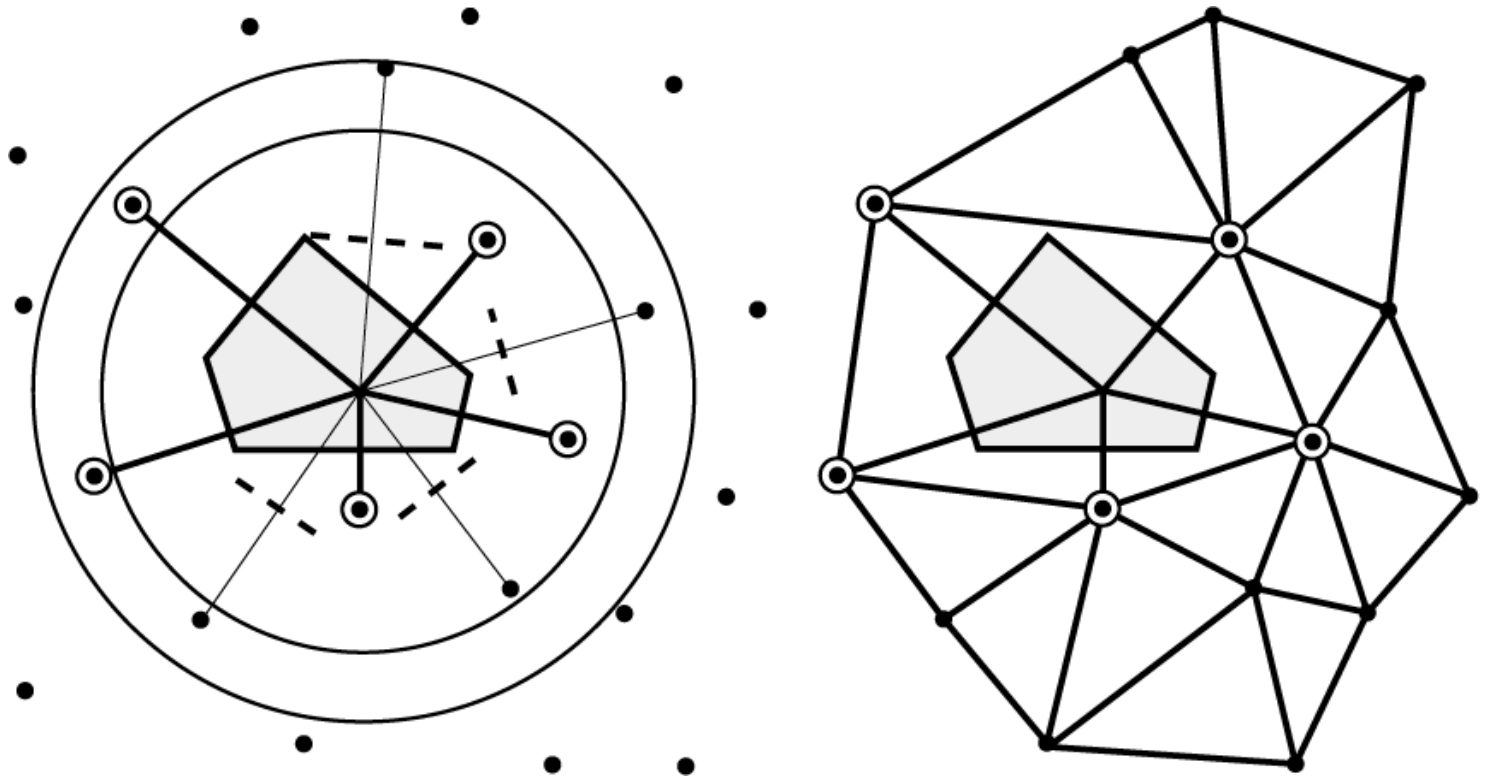


Chopping Down Method of Calculating Polyhedra



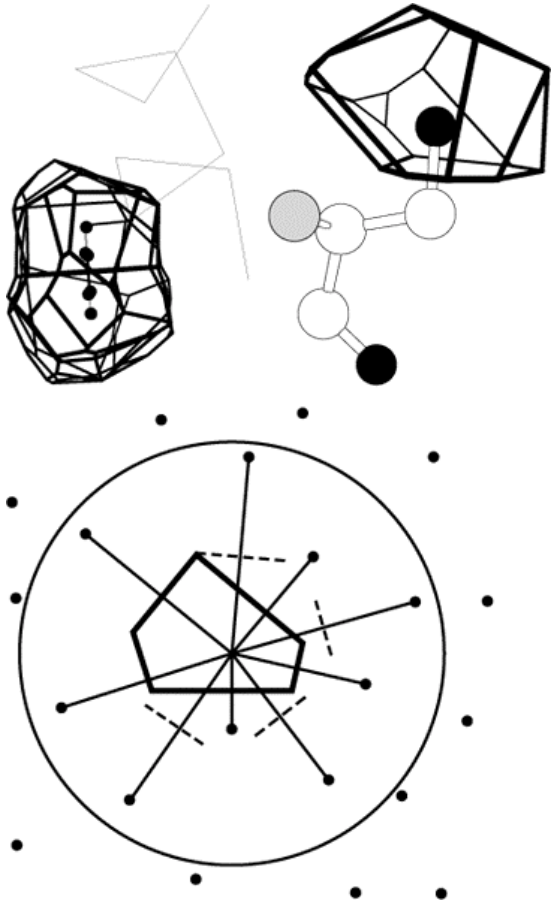
Delauney Triangulation, the Natural Way to Define Packing Neighbors

- Related to Voronoi polyhedra (dual)
- What “coordination number” does an atom have? Doesn’t depend on distance
- alpha shape (Edelsbrunner)



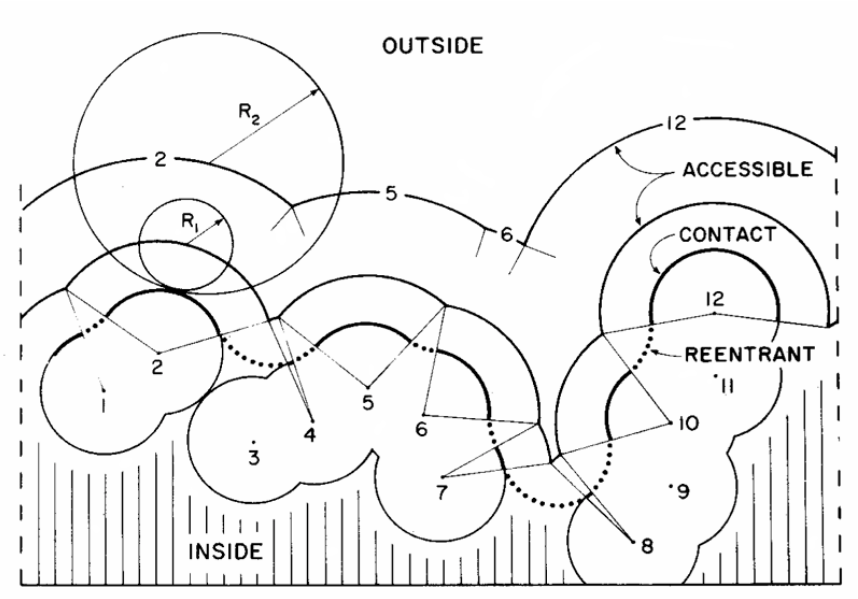
Properties & uses of Voronoi Polyhedra

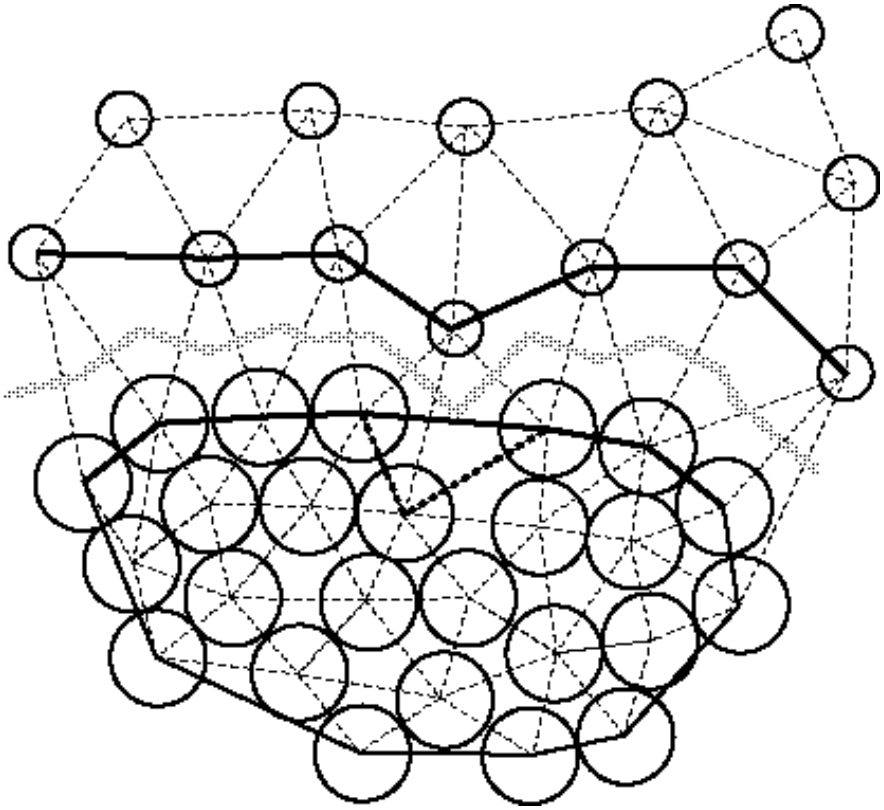
- All Space Filled
  - ◊ If Voronoi polyhedra are constructed around atoms in a periodic system, such as in a crystal, all the volume in the unit cell will be apportioned to the atoms. There will be no gaps or cavities as there would be if one, for instance, simply drew spheres around the atoms.
- V = A d
  - ◊ Voronoi volume of an atom is a weighted average of distances to all its neighbors, where the weighting factor is the contact area with the neighbor.
- DT
  - ◊ Border of D.T. is Convex Hull
  - ◊ D.T. produces "fatest" possible triangles which makes it convenient for things such as finite element analysis.
- Applications
  - ◊ Nearest neighbor problems. The nearest neighbor of a query point in center of the Voronoi diagram in which it resides
  - ◊ Largest empty circle in a collection of points has center at a Voronoi vertex
- Weighting



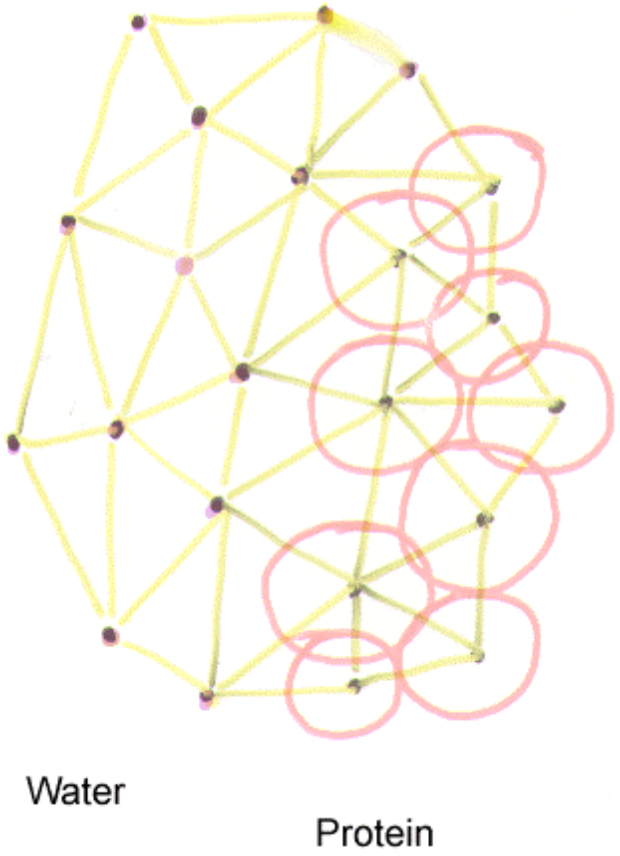
Packing defines the Protein Surface: Surface & Volume Definitions Linked

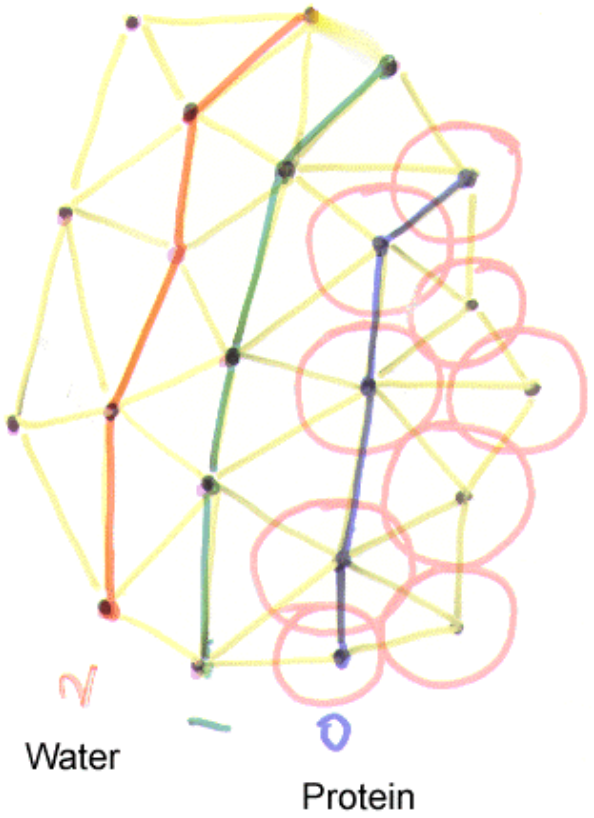
- Voronoi polyhedra are the *Natural* way to study surfaces!
- The relationship between
  - ◊ accessible surface
  - ◊ molecular surface
  - ◊ Delauney Triangulation (Convex Hull)
  - ◊ polyhedra faces
  - ◊ hydration surface



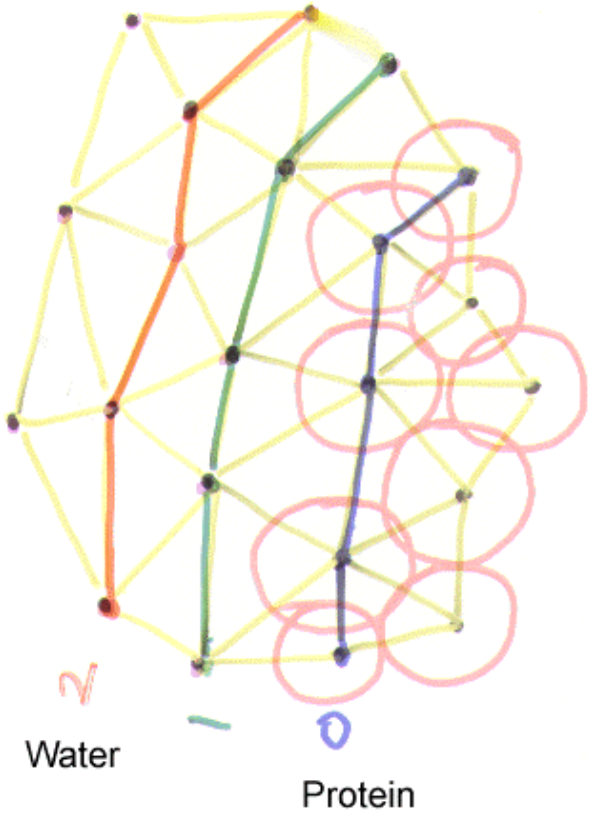


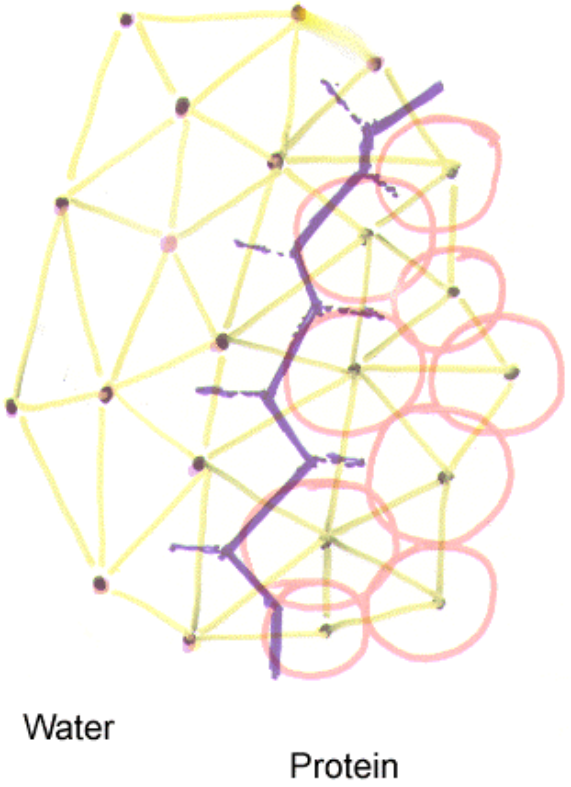
Defining Surfaces from Packing: Convex Hull and Layers of Waters



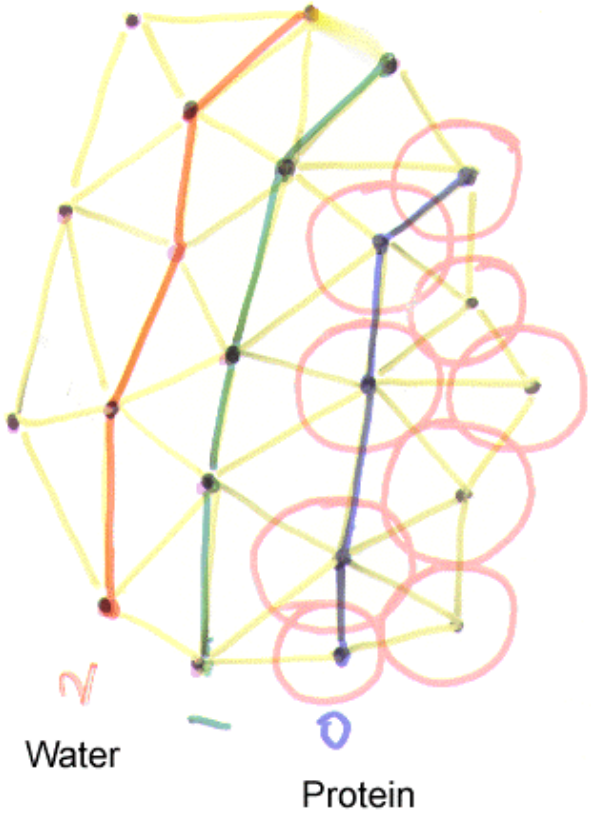


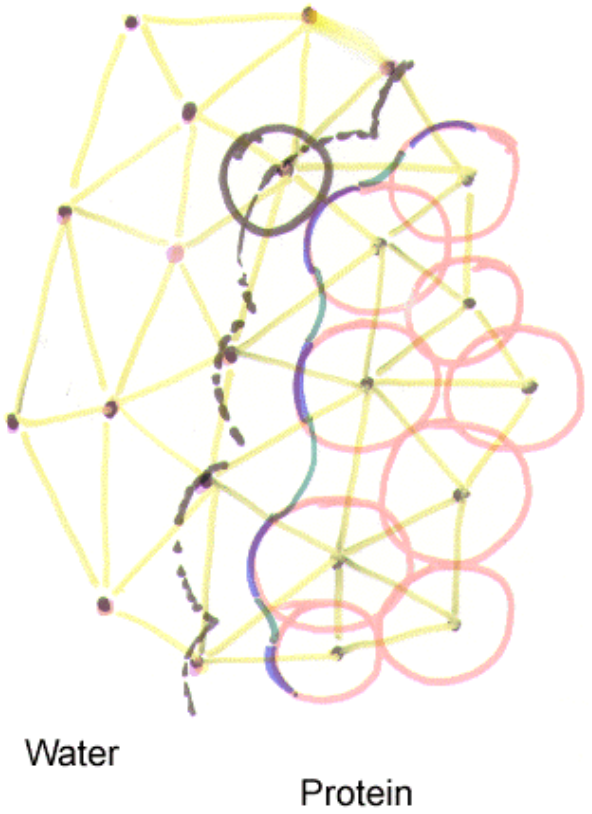
Defining a Surface from the Faces of Voronoi Polyhedra





Accessible Surface as a Time-averaged Water Layer





Analysis of Macromolecular Geometry & Motions

W Krebs, C Chothia,
J Tsai, Y Harpaz, R Taylor

bioinfo.mbb.yale.edu/MolMovDB

1 Database of Macromolecular Motions

Interface packing, shear v. hinge, web DB

2 Morph Movie Server

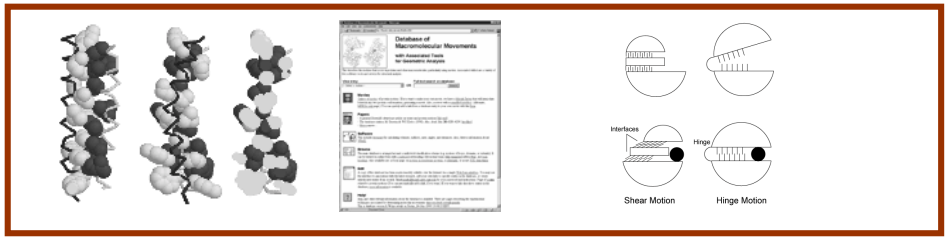
Restrained Interpolation between submitted endpoint conformations, standardized stats, hinge finder

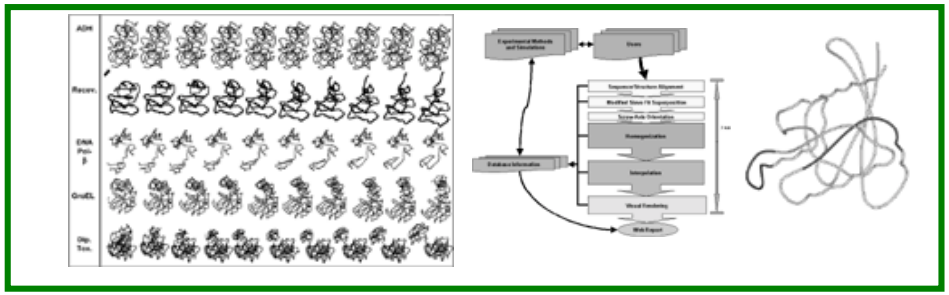
3 Geometry Software (Volumes)

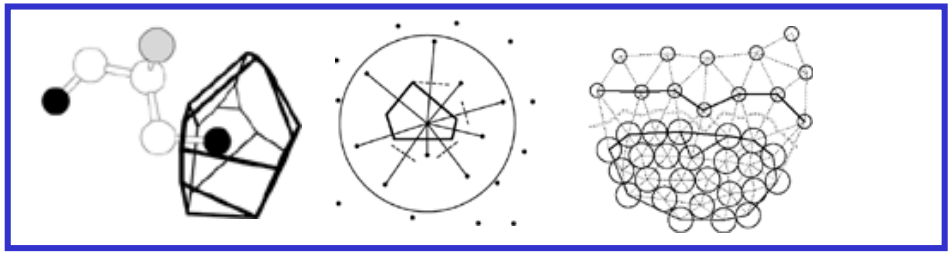
Voronoi volumes, relation to surface, radii problem, packing measurement

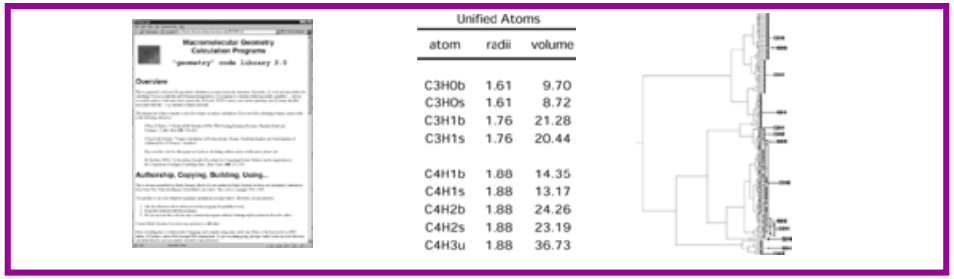
4 Standard Radii & Volumes

Needed for volume calculation, 18 basic types, Sensitivity Analysis, radii set, atom selection criteria, structure set









Unified Atoms		
atom	radii	volume
C3H0b	1.61	9.70
C3H0s	1.61	8.72
C3H1b	1.76	21.28
C3H1s	1.76	20.44
C4H1b	1.88	14.35
C4H1s	1.88	13.17
C4H2b	1.88	24.26
C4H2s	1.88	23.19
C4H3u	1.88	36.73
N3H0u	1.64	8.65
N3H1b	1.64	15.72
N3H1s	1.64	13.62
N3H2u	1.64	22.69
N4H3u	1.64	21.41
O1H0u	1.42	15.91
O2H1u	1.46	17.98
S2H0u	1.77	29.17
S2H1u	1.77	36.75

Residues	
aa	volume
Gly	63.8
Ala	89.3
Val	138.2
Leu	163.1
Ile	163.0
Met	165.8
Pro	121.6
His	157.5
Phe	190.8
Tyr	194.6
Trp	226.4
Cyh	112.8
Cys	102.5
Ser	94.2
Thr	119.6
Asn	112.4
Gln	146.9
Asp	114.4
Glu	138.8
Lys	165.1
Arg	190.3

Parameters used in Protor Volume Derivation	
Typing Scheme	Hybrid chemical and numerical typing with 18 basic types
Radii Set	ProtOr Radii, Tsai et al. (1999)
Plane-Positioning Method	Ratio
Atom Selection Criteria	BL+
Structure Set	SCOP (87 structures)

ProtOr Parameter Set: Standard Radii & Volumes

- Consistent Radii, Typing, and Volumes for Packing Calculations
- For comparison

Sample Results of Volume Calculations: Significant but Small Changes in Packing

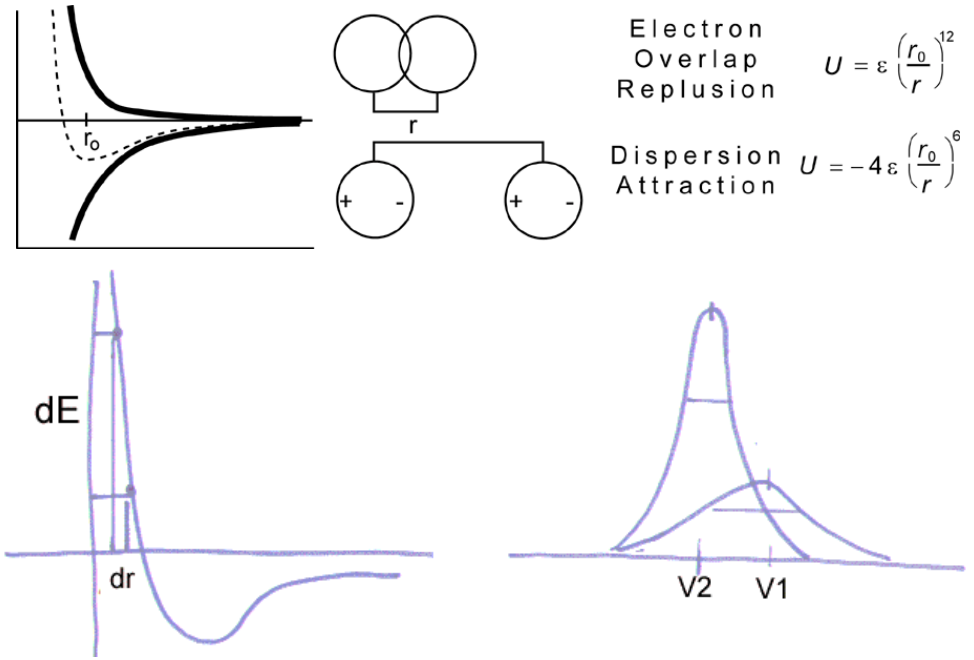
```
-CH2- volume in cubic A

23.7    standard volume
        in protein core

23.6    mobile helix-helix
        interface (CS, TrpR)

24.8    grooves on protein surface
        (in high-res. struc.
        via SurFractal)
```

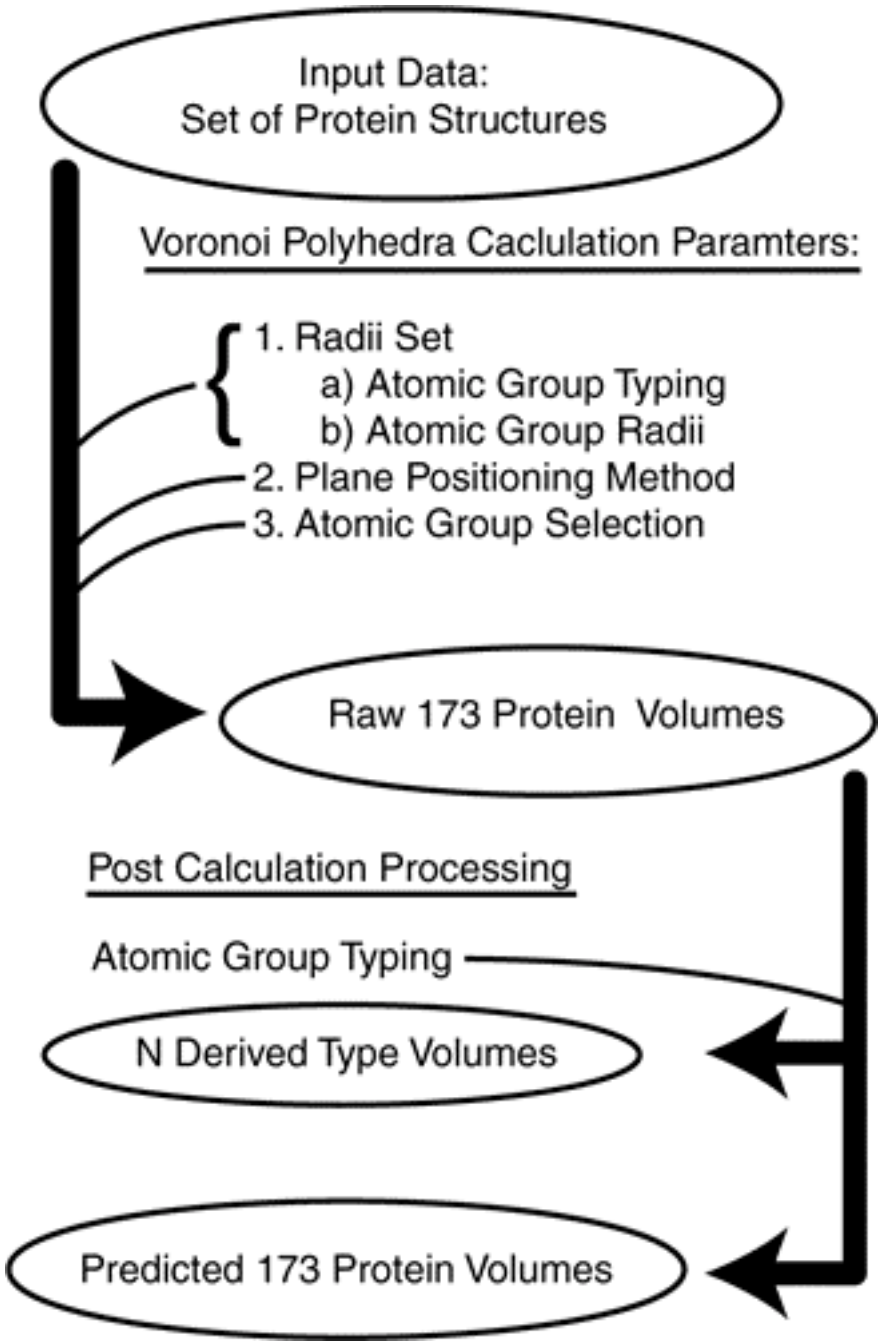
- VDW ~ Packing
  - ◊ Exponential Repulsion
- Many observations (>10K) in standard volumes gives small error about the mean (SD/sqrt(N))



atom	Count	Mean volume	SD	SD(%)	Minimum	Maximum
GLY						
O	452	16.154	1.313	8.127	13.087	21.208
C	685	9.652	0.715	7.404	7.627	12.060
CA	272	23.470	1.924	8.200	19.386	29.000
N	577	14.480	1.127	7.782	10.917	19.363
Total	1986	63.756	2.685	4.211		
ALA						
O	610	16.026	1.230	7.675	11.753	21.098
C	1070	8.858	0.519	5.857	7.828	11.063
CA	678	13.959	1.120	8.025	11.595	18.123
N	786	13.872	1.038	7.483	11.564	18.014
CB	397	36.551	2.793	7.642	29.350	45.481
Total	3541	89.266	3.452	3.867		
VAL						
O	622	15.998	1.176	7.351	13.257	20.987
C	1027	8.528	0.540	6.327	7.148	10.390
CA	781	13.078	0.934	7.145	10.791	17.340
N	864	13.553	0.809	5.968	11.340	17.500
CB	776	14.514	1.510	10.406	2.247	18.973
CG1	481	36.320	2.937	8.086	16.906	46.985
CG2	512	36.173	2.957	8.175	20.029	45.018
Total	5063	138.164	4.780	3.460		
LEU						
O	715	15.957	1.233	7.728	12.453	19.813
C	1040	8.781	0.568	6.474	7.153	11.923
...						

The Raw Output -- 173 Volumes

- Trying to find **Mean V**
- Want High **Counts**
- Minimum **SD**

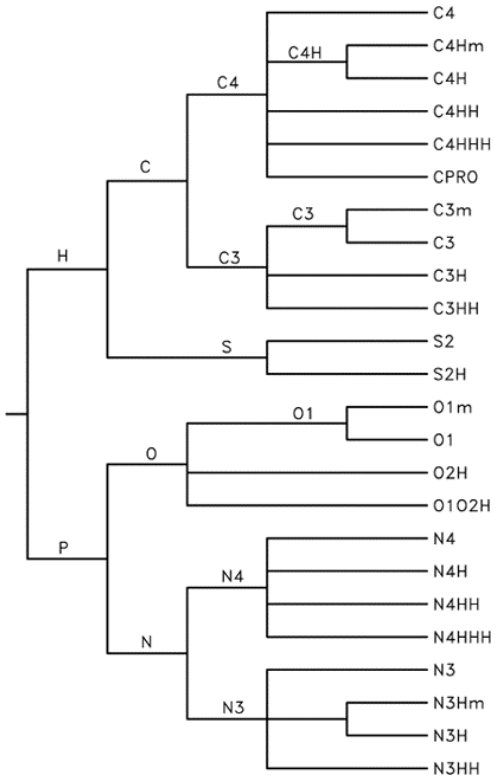


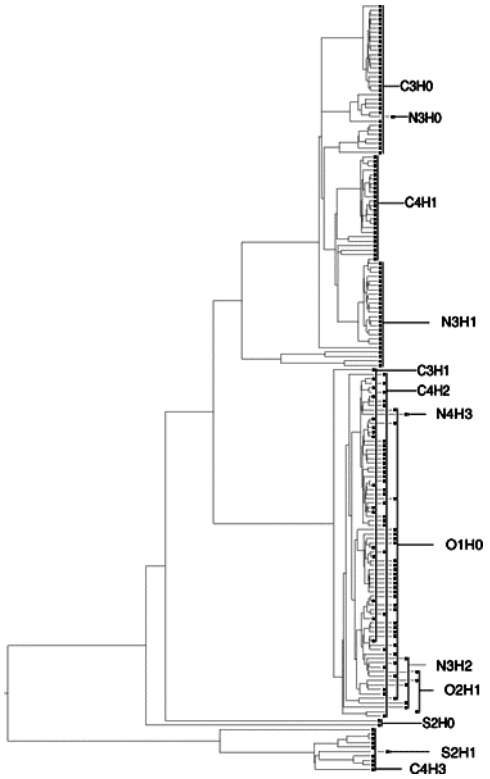
Sensitivity Analysis on Volume Calculation: Number of Types, Collapsing down from 173

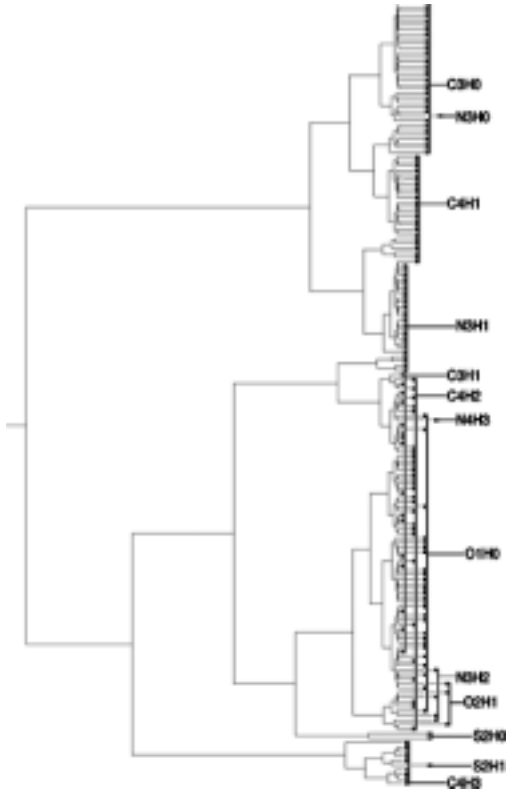
Clustering into a set of Atom Types I

- Which atoms are equivalent? How many types valid?
- 18 types, [CNOS][34]H[123][bsu]

Chemical **Single-link** **Multi-link**

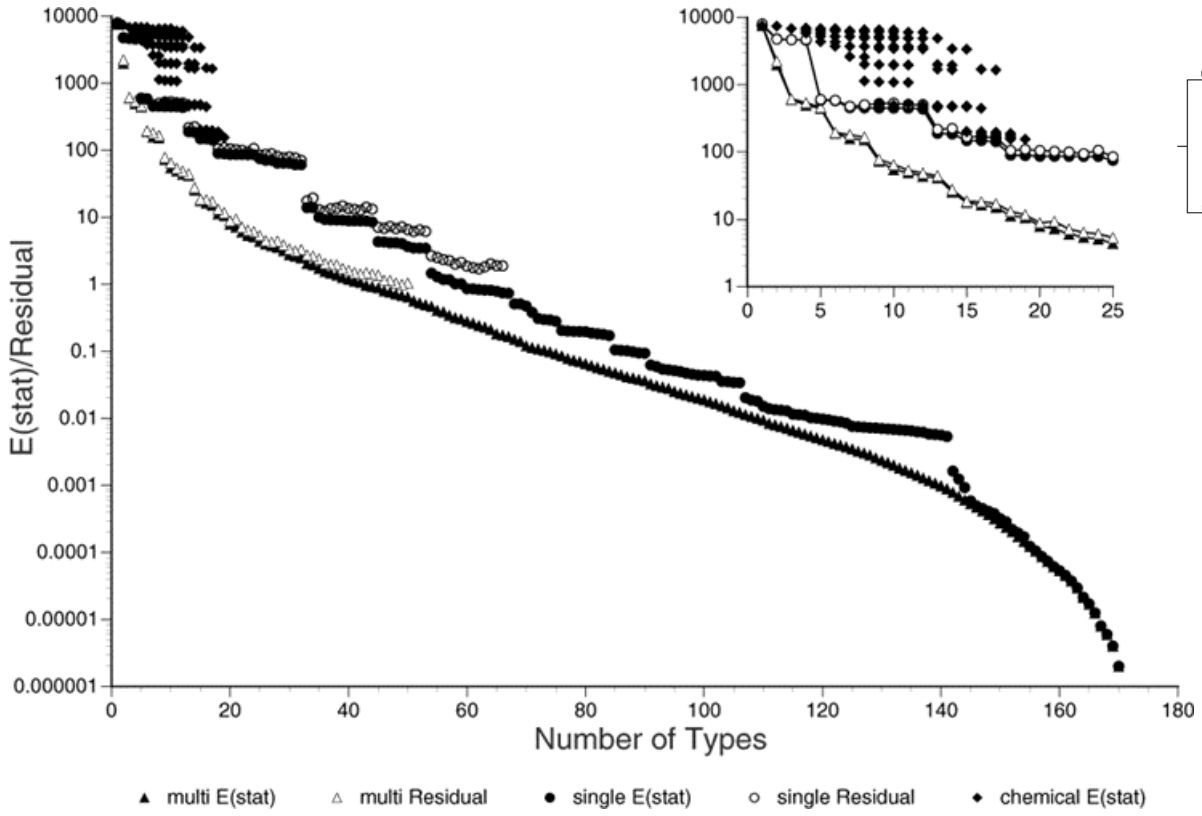


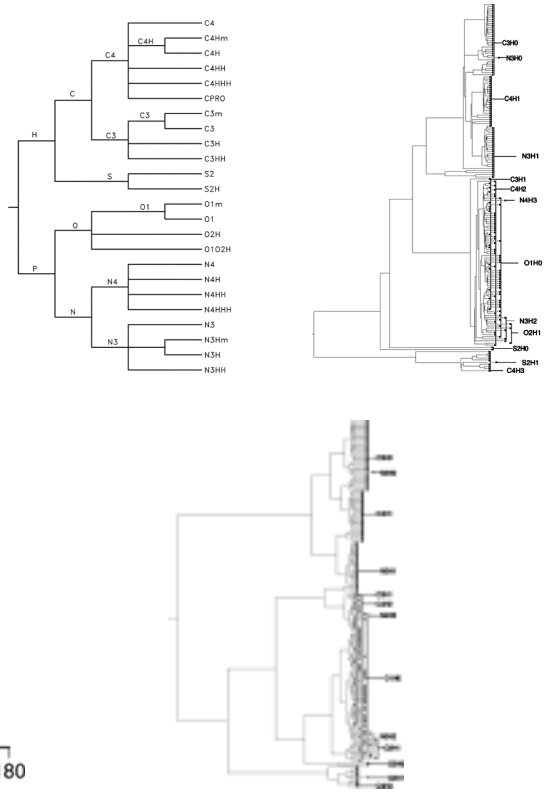




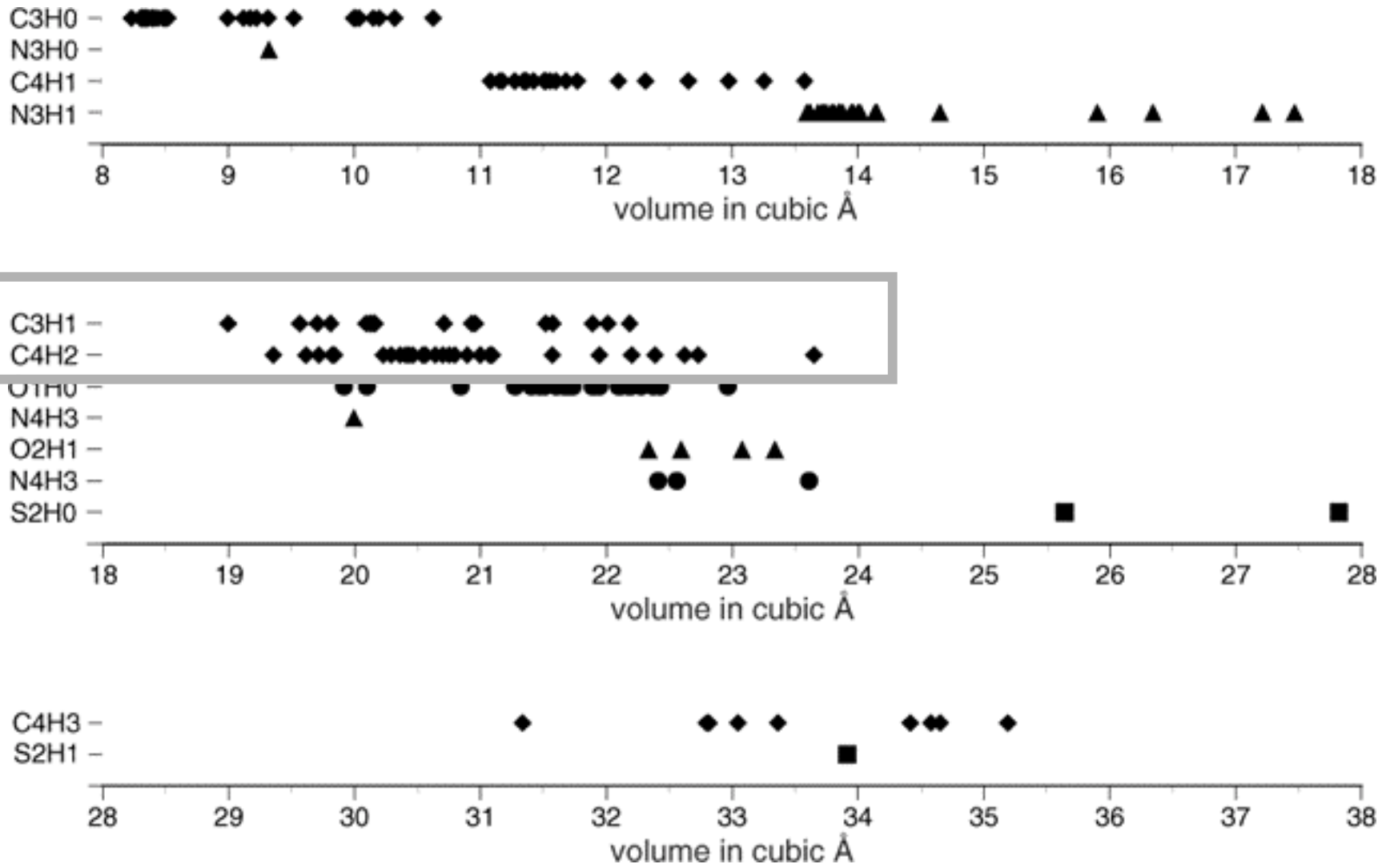
Clustering into a set of Atom Types II

- Which atoms are equivalent? How many types valid?
- 18 types, [CNOS][34]H[123][bsu]
- Residual to tell apart





Overlap of Volumes of Aromatic C3H1 and Aliphatic C4H2



Different Sets of Radii

C3H0	1.61
C3H1	1.76
C4H1	1.88
C4H2	1.88
C4H3	1.88
N3H0	1.64
N3H1	1.64
N3H2	1.64
N4H3	1.64
O1H0	1.42
O2H1	1.46
S2H0	1.77
S2H1	1.77

Atom Type & Symbol		Bondi 1968	Lee & Richards 1971	Shrake & Rupley 1973	Richards 1974	Chothia 1975	Rich-mond & Richards 1978	Gelin & Karplus 1979	Dunfield et al. 1979	ENCAD derived 1995	CHARMM derived 1995	Tsai et al. 1998
$-CH_3$	Aliphatic, methyl	2.00	1.80	2.00	2.00	1.87	1.90	1.95	2.13	1.82	1.88	1.88
$-CH_2-$	Aliphatic, methyl	2.00	1.80	2.00	2.00	1.87	1.90	1.90	2.23	1.82	1.88	1.88
>CH-	Aliphatic, CH	-	1.70	2.00	2.00	1.87	1.90	1.85	2.38	1.82	1.88	1.88
=CH	Aromatic, CH	-	1.80	1.85	*	1.76	1.70	1.90	2.10	1.74	1.80	1.76
>C=	Trigonal, aromatic	1.74	1.80	*	1.70	1.76	1.70	1.80	1.85	1.74	1.80	1.61
$-NH_3+$	Amino, protonated	-	1.80	1.50	2.00	1.50	0.70	1.75		1.68	1.40	1.64
$-NH_2$	Amino or amide	1.75	1.80	1.50	-	1.65	1.70	1.70		1.68	1.40	1.64
>NH	Peptide, NH or N	1.65	1.52	1.40	1.70	1.65	1.70	1.65	1.75	1.68	1.40	1.64
=O	Carbonyl Oxygen	1.50	1.80	1.40	1.40	1.40	1.40	1.60	1.56	1.34	1.38	1.42
-OH	Alcoholic hydroxyl	-	1.80	1.40	1.60	1.40	1.40	1.70		1.54	1.53	1.46
-OM	Carboxyl Oxygen	-	1.80	1.89	1.50	1.40	1.40	1.60	1.62	1.34	1.41	1.42
-SH	Sulfhydryl	-	1.80	1.85	-	1.85	1.80	1.90		1.82	1.56	1.77
-S-	Thioether or –S-S-	1.80	-	-	1.80	1.85	1.80	1.90	2.08	1.82	1.56	1.77

7386 *J. Phys. Chem., Vol. 100, No. 18, 1996*

$$f = \sum_{i} \sum_{j \le i} w_{ij} (d_{ij} - [r_i + r_j])^2$$

R Taylor at CCDC optimized **r's** to fit **d's**

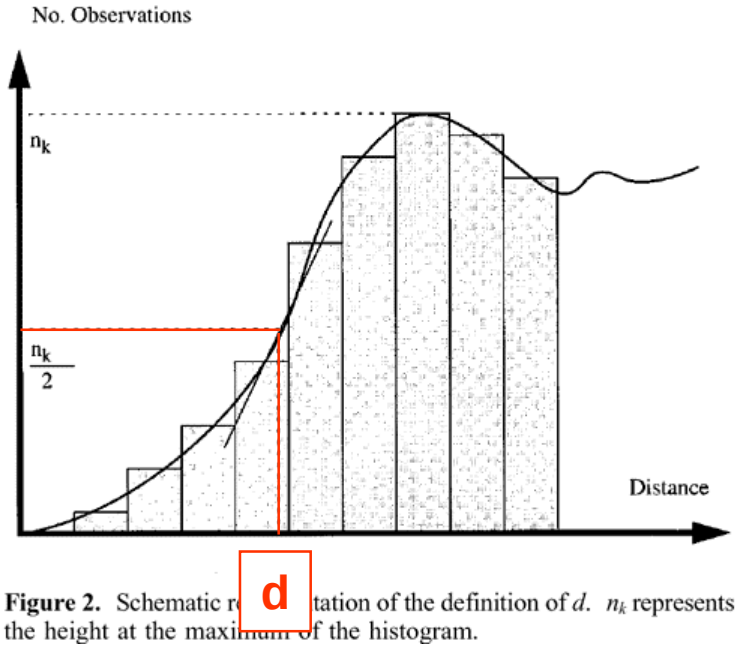


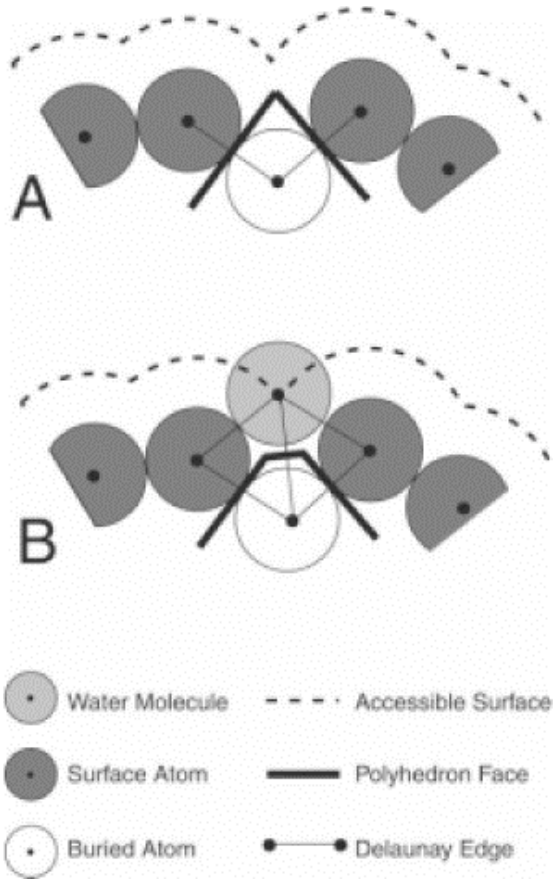
Figure 2. Schematic r tation of the definition of d. n_k represents the height at the maxi of the histogram.

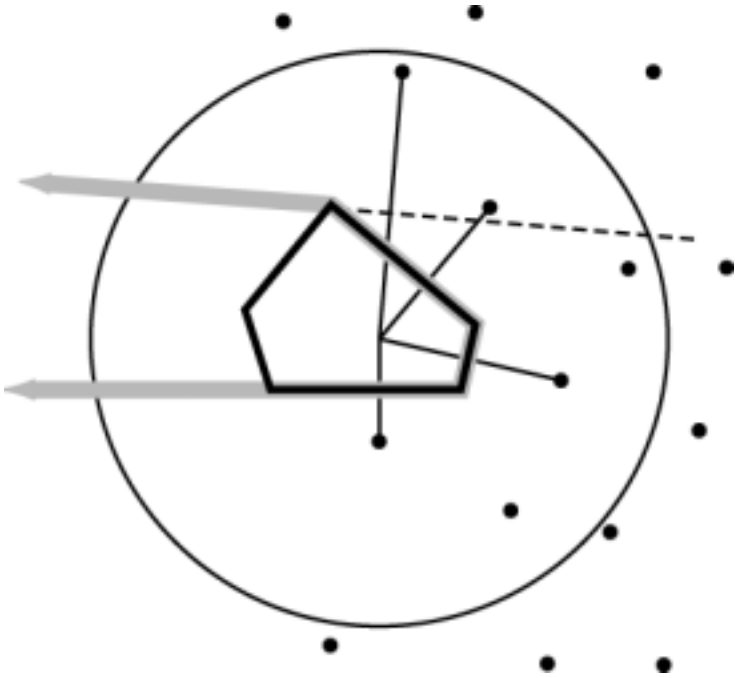
ProtOr atom type	PDB Sets[a]													
	SCOP		Standard		High		Low		NMR		New		Obsolete	
	Vol.[b]	SD	Vol.[b]	SD	Vol.[b]	SD	Vol.[b]	SD	Vol.[b]	SD	Vol.[b]	SD	Vol.[b]	SD
C3HOb	9.64	0.72	9.67	0.68	9.65	0.68	9.68	0.69	9.53	1.05	9.78	0.79	9.83	0.86
C3HOs	8.66	0.58	8.68	0.59	8.65	0.57	8.70	0.60	8.65	0.80	8.77	0.69	8.84	0.76
C3H1b	21.33	1.87	21.38	1.89	21.36	1.85	21.39	1.91	19.40	2.73	21.26	2.11	20.96	2.30
C3H1s	20.45	1.76	20.41	1.77	20.27	1.72	20.50	1.80	18.48	2.78	20.42	2.02	20.43	2.21
C4H3u	14.35	1.35	14.41	1.22	14.38	1.20	14.43	1.23	13.89	1.55	14.40	1.48	14.42	1.59
C4H1b	13.14	0.97	13.17	0.96	13.20	0.94	13.15	0.97	13.20	1.27	13.11	1.11	13.18	1.20
C4H1s	24.14	2.07	24.25	2.13	24.11	1.95	24.33	2.21	20.48	5.89	24.26	2.43	24.07	2.76
C4H2b	23.17	2.35	23.29	1.94	23.28	1.96	23.29	1.93	19.13	6.40	23.14	2.23	22.92	2.46
C4H2s	36.84	3.24	36.94	2.99	36.93	3.00	36.94	2.98	30.38	8.26	36.43	3.75	35.76	3.95
N3HOu	8.62	0.59	8.57	0.65	8.60	0.70	8.56	0.6						
N3H1b	15.65	1.55	15.73	1.70	15.55	1.48	15.80	1.7						
N3H1s	13.54	0.99	13.53	1.00	13.52	0.97	13.53	1.0						
N3H2u	22.61	2.36	22.07	2.13	22.12	2.22	22.04	2.0						
N4H3u	21.56	1.28	21.03	1.29	20.30	0.55	21.76	1.4						
O1HOu	15.91	1.29	15.92	1.28	15.87	1.23	15.94	1.3						
O2H1u	18.11	1.78	18.09	1.86	18.10	1.97	18.09	1.7						
S2HOu	29.29	2.68	28.79	2.67	28.66	2.68	28.90	2.6						
S2H1u	36.82	3.48	35.93	2.44	37.15	2.46	35.71	2.3						

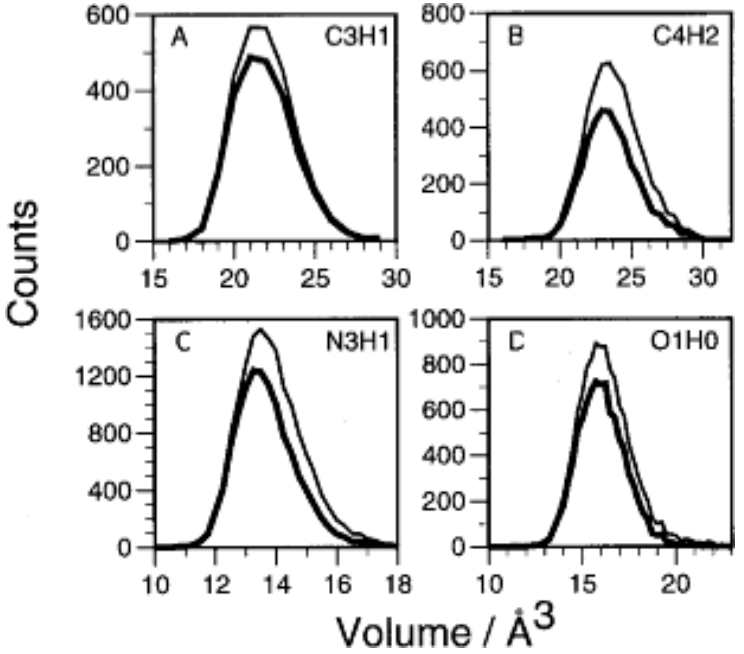
Compare Different Structure Sets

Set	Number	Identifier
Standard	130	135l, 1aaj, 1aap, 1ake, 1arb, 1bbh, 1bp2, 1ccr, 1cdp, 1cmb, 1cpc, 1crn, 1cse, 1ctf, 1cus, 1dfn, 1dr1, 1eco, 1ezm, 1fkf, 1fus, 1fxd, 1gct, 1gd1, 1gpr, 1hbg, 1hel, 1hne, 1ifc, 1igd, 1lmb, 1lz1, 1lz3, 1mba, 1mbd, 1ofv, 1omd, 1paz, 1pgx, 1pk4, 1plc, 1ppn, 1ppt, 1ptx, 1rcf, 1rdg, 1rms, 1rop, 1rpg, 1rpo, 1rro, 1sar, 1sgt, 1snc, 1st3, 1thm, 1ubq, 1ycc, 256b, 2act, 2alp, 2apr, 2aza, 2cba, 2ccy, 2cdv, 2cpp, 2ctc, 2cyp, 2er7, 2fb4, 2fcr, 2fx2, 2gbp, 2hhb, 2ihl, 2ltn, 2mcm, 2mhr, 2msb, 2ovo, 2por, 2prk, 2rhe, 2rn2, 2sga, 2sn3, 2trx, 2utg, 2wrp, 2zta, 3app, 3b5c, 3bcl, 3c2c, 3cla, 3dfr, 3ebx, 3est, 3fxn, 3grs, 3lzm, 3rp2, 3sgb, 451c, 4dfr, 4enl, 4icb, 4ins, 4ptp, 5cpa, 5cyt, 5p21, 5pal, 5pti, 5rub, 5rxn, 5tim, 6ebx, 6rlx, 6rxn, 6xia, 7aat, 7rsa, 8dfr, 8fab, 8rxn, 9pti, 9rnt, 9wga
SCOP	87	1cbn, 1lkk, 2erl, 8rxn, 1bpi, 1ctj, 1igd, 1rge, 1amm, 1arb, 1cse, 1jbc, 2sn3, 1cus, 7rsa, 1rro, 1aac, 193l, 1utg, 5p21, 1hms, 1xyz, 256b, 2olb, 2phy, 3ebx, 3sdh, 2end, 1xso, 1cka, 1cyo, 1edm, 1ezm, 1isu, 1mla, 1poa, 1rie, 1whi, 2ctb, 2eng, 2ovo, 2cba, 3grs, 1lit, 1ra9, 1tca, 1csh, 1epn, 1mrj, 1phc, 1ptf, 1smd, 1vcc, 2dri, 2ilk, 2sil, 3pte, 4fgf, 2cpl, 1kap, 1lcp, 1php, 1snc, 1sri, 2wrp, 1krn, 2trx, 1ctf, 1fnb, 1gai, 1gof, 1knb, 1llp, 1mol, 1pdo, 1rop, 1tad, 1tfe, 1vhh, 1vsd, 2act, 1fkd, 1chd, 1kpt, 1thw, 2bbk, 3cla
NMR	125	1aab, 1aaf, 1aca, 1acp, 1afp, 1ahd, 1ale, 1alf, 1bbo, 1bus, 1bw3, 1bw4, 1cdb, 1cdn, 1cis, 1clb, 1crp, 1crq, 1crr, 1csy, 1csz, 1ctl, 1dhm, 1erg, 1erh, 1fht, 1fkr, 1fks, 1fkt, 1ftz, 1gb1, 1gbr, 1gfc, 1gfd, 1hcc, 1hdn, 1hme, 1hmf, 1hom, 1hrq, 1hrr, 1hsm, 1hsn, 1hue, 1hum, 1hun, 1il8, 1iml, 1irp, 1kb7, 1kb8, 1ldl, 1ldr, 1lip, 1lpt, 1mbe, 1mbf, 1mbg, 1mbj, 1mbk, 1mef, 1ncp, 1neh, 1neq, 1ner, 1nhm, 1nhn, 1nil, 1nim, 1nmf, 1nmg, 1noe, 1odp, 1odq, 1odr, 1oef, 1oeg, 1pan, 1pao, 1pcp, 1pdc, 1pih, 1pij, 1pmc, 1pog, 1pra, 1prr, 1prs, 1pse, 1psf, 1qwe, 1qwf, 1rht, 1rip, 1rod, 1rpv, 1san, 1sap, 1srl, 1srm, 1stu, 1sxl, 1tam, 1tiv, 1tvs, 1tvt, 1ums, 1umt, 1utr, 1vnd, 1zer, 2abd, 2bus, 2gb1, 2gva, 2gvb, 2hid, 2hmx, 2hoa, 2igg, 2igh, 2il8, 2ptl, 2znf, 3ci2
Current	69	116l, 1act, 1alp, 1alr, 1anh, 251c, 156b, 1apd, 2bcl, 1abk, 1abp, 1abx, 1afg, 1ace, 1afn, 1ak3, 1asi, 1aza, 1baa, 1bjl, 2grs, 1cab, 1cae, 1cd4, 1ci2, 1cpk, 1cln, 1dhb, 1dri, 1eip, 1end, 7atc, 1fnr, 1gap, 1gbp, 1gcr, 1gmf, 1gn5, 2hvt, 1gsr, 1gyi, 1hft, 1hid, 1hmg, 1hmx, 1lrd, 3fab, 1mev, 1omf, 1ora, 1pab, 1pel, 1pgk, 1phy, 1ptc, 1r04, 1r1e, 1rsl, 1sod, 1srt, 1tbs, 1tct, 1trt, 1yhx, 2adk, 1vaa, 1ts1, 1ada
Obsolete	69	1abe, 1cdh, 1eri, 1fnb, 1lmb, 216l, 256b, 2abk, 2abx, 2ace, 2act, 2ada, 2afg, 2afn, 2ak3, 2alp, 2alr, 2anh, 2apd, 2asi, 2aza, 2baa, 2cab, 2cae, 2ci2, 2cpk, 2cyh, 2dhb, 2dri, 2eip, 2end, 2gmf, 2gn5, 2gsr, 2gyi, 2hft, 2hid, 2hmg, 2hmx, 2mev, 2omf, 2ora, 2pab, 2pel, 2phy, 2ptc, 2r04, 2rsl, 2sod, 2srt, 2tbs,2tct, 2trt, 2ts1, 2vaa, 2yhx, 351c, 3adk, 3bcl, 3bjl, 3cln, 3gap, 3gbp, 3grs, 3hvt, 3pgk, 4gcr, 5at1, 7fab

Sensitivity of Voronoi Construction to Surface Structure







Sensitivity Analysis for Different Atom Selections

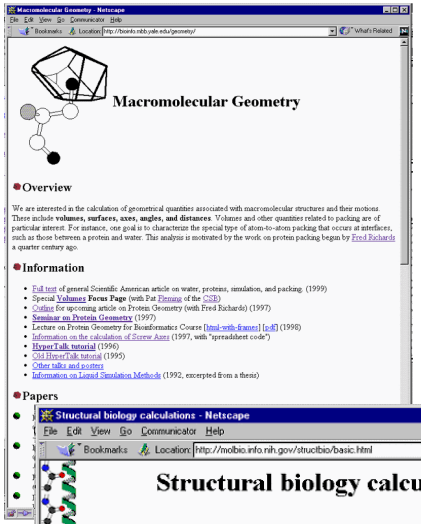
B: This set contains protein atoms that are buried by other protein atoms and by ligands and/or cofactors. In selecting this set, the crystallographically determined water structure is ignored: *i.e.* the protein atoms used are those that have zero accessible surface area (Lee & Richards, 1971, Connolly, 1983) as calculated using just the atoms in the proteins, ligands, and cofactors.

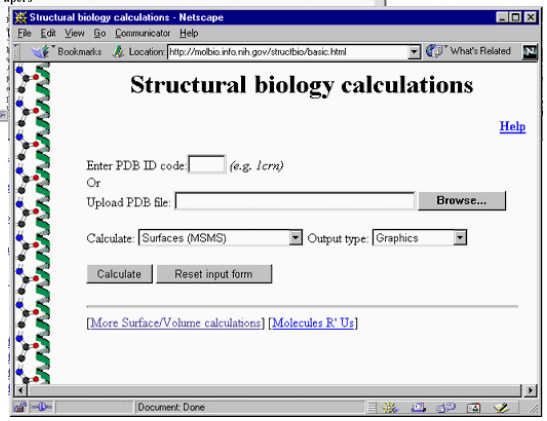
BL: This set contains atoms that are buried as defined by the B set *less those whose volumes are affected by ligands and cofactors*. The set was selected by removing from set B those atoms whose volumes are different when they are calculated in the presence and absence of ligands and cofactors. The L in the name of this set indicates this extra filtering of atoms.

BD: The atoms excluded from this set are (i) all those that have surface accessible to the solvent (as in set B) and (ii) all those in contact to these surface atoms. Thus both surface atoms and those that form the first layer below the surface are removed from the calculation to leave only those that are deeply buried. Therefore, the volumes produced by this set of atoms are named BD, where the D indicates that the resulting set of atoms are buried deep in the protein.

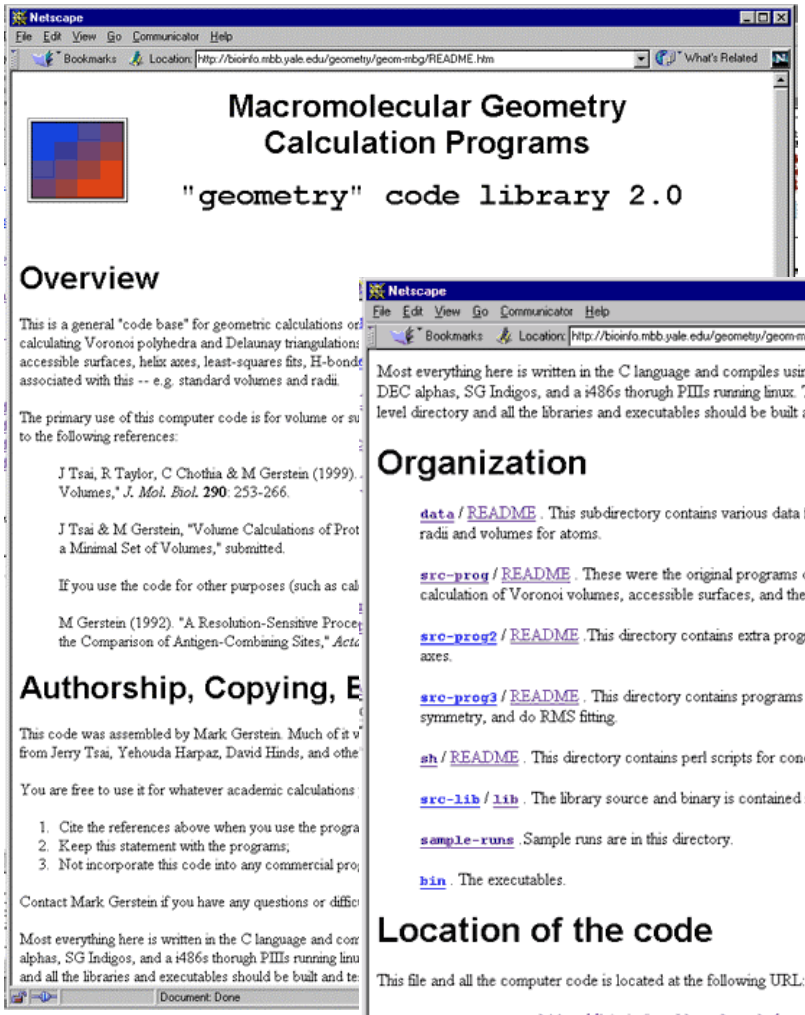
	Atom Selection																				
	base			B+			B-			BL+			BL-			BT			BD		
	All atomic groups for which polyhedra can be constructed			Atomic groups with zero accessible surface area, waters included			Atomic groups with zero accessible surface area, waters excluded			Same as B+ except those atomic groups touching non-protein atomic groups			Same as B- except those atomic groups touching non-protein atomic groups			Same as B+ less those touching the exposed atomic groups			Same as BL+ less those atomic groups touching non-selected atomic groups		
ProtOr type	Count	Vol.[b]	SD	Count	Vol.[b]	SD	Count	Vol.[b]	SD	Count	Vol.[b]	SD	Count	Vol.[b]	SD	Count	Vol.[b]	SD	Count	Vol.[b]	SD
C3H0b	12148	11.11	2.55	6539	9.79	0.73	7802	9.78	0.72	6090	9.79	0.74	4255	9.67	0.68	4618	9.68	0.69	1434	9.64	0.63
C3H0s	24588	9.05	1.08	18341	8.79	0.64	20429	8.79	0.62	17911	8.78	0.64	13269	8.68	0.59	14011	8.69	0.59	3647	8.73	0.58
C3H1b	5129	23.87	6.65	3024	21.39	1.88	3668	21.40	1.88	2773	21.40	1.87	2363	21.38	1.89	2376	21.38	1.89	1046	21.33	1.84
C3H1s	5902	23.95	7.19	2941	20.56	1.82	3834	20.57	1.79	2627	20.56	1.81	1938	20.41	1.71	1973	20.43	1.78	767	20.52	1.71
C4H1b	10248	16.05	4.12	5934	14.47	1.23	7440	14.47	1.23	5719	14.46	1.23	4579	14.41	1.22	4634	14.42	1.22	1947	14.37	1.17
C4H1s	20389	14.66	3.25	10882	13.27	1.01	14128	13.31	0.99	10455	13.26	1.00	8280	13.17	0.96	8358	13.17	0.97	3274	13.17	0.90
C4H2b	4391	30.62	11.82	1712	24.41	2.19	2411	24.43	2.20	1564	24.46	2.20	1152	24.25	2.13	1162	24.28	2.14	559	24.13	2.06
C4H2s	21531	29.18	10.06	6421	23.52	2.01	10234	23.48	1.97	5927	23.54	2.00	3948	23.29	1.94	4010	23.30	1.94	1989	23.33	1.88
C4H3u	12891	41.84	11.13	5711	36.96	3.06	7635	36.87	3.05	5241	36.98	3.02	4160	36.94	2.99	4179	36.94	2.99	1890	36.74	2.90
N3H0u	1119	9.14	1.82	848	8.80	0.74	948	8.70	0.73	830	8.79	0.72	495	8.57	0.65	501	8.58	0.66	107	8.48	0.56
N3H1b	2047	20.18	8.98	769	15.76	2.42	1212	15.74	2.02	648	16.07	1.98	432	15.73	1.30	449	15.75	1.33	228	15.31	2.09
N3H1s	25551	14.87	3.91	16057	13.77	1.19	20712	13.72	1.06	15354	13.76	1.18	11180	13.53	1.00	11224	13.53	1.00	3282	13.74	0.99
N3H2u	2276	33.13	15.23	486	23.05	2.84	878	22.82	2.33	345	23.03	2.89	178	22.07	2.13	179	22.08	2.13	142	22.54	2.10
N4H3u	401	41.77	22.36	26	20.20	6.51	88	21.32	3.77	19	23.07	3.40	6	21.03	1.29	6	21.03	1.29	9	21.56	2.61
O1H0u	28430	20.84	10.21	12545	16.14	1.55	18670	16.07	1.45	11781	16.15	1.51	8847	15.92	1.28	8865	15.92	1.28	2759	16.07	1.30
O2H1u	3289	25.29	12.66	910	18.63	2.38	1842	18.08	1.93	780	18.67	2.34	477	18.09	1.86	479	18.10	1.87	369	17.82	1.49
S2H0u	643	32.05	9.04	507	28.68	2.87	610	28.64	2.84	461	28.80	2.61	352	28.79	2.67	354	28.80	2.67	115	28.16	3.29
S2H1u	167	35.75	8.92	94	33.32	4.85	119	33.39	5.38	55	36.23	2.62	47	35.93	2.44	47	35.93	2.44	34	33.92	3.08

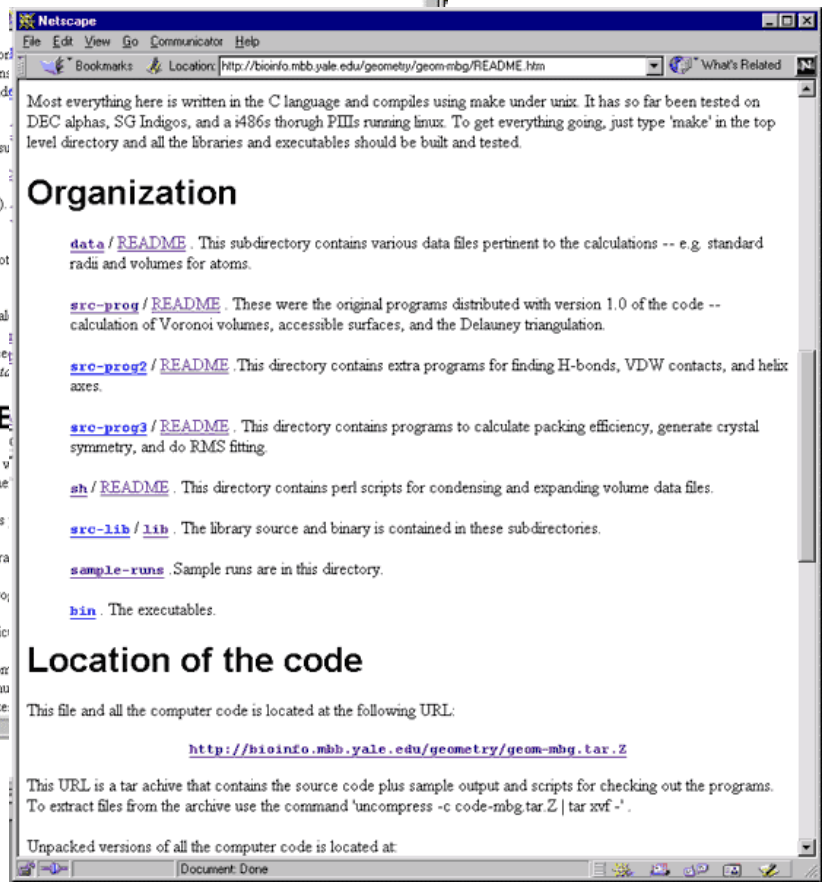
C Source for Geometry Code





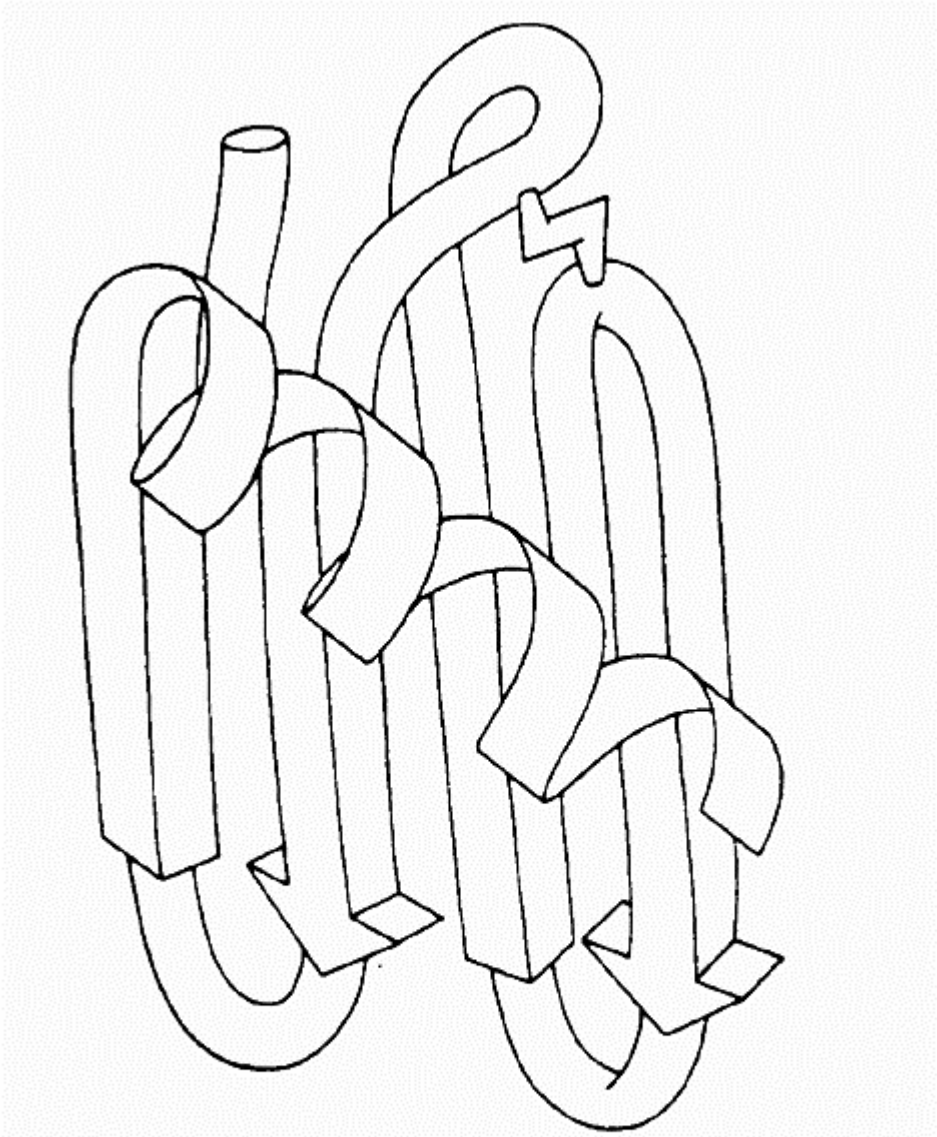
Server (Chacko)





make, sample runs (reg. tests), **parms**

Volumes, Polyhedra, Surfaces, Packing Eff., LSQ Fits, Axes, Angles, Distances, Contacts/Hbonds, Crystal Sym.



Incredulase

J.S. Richardson and D.C. Richardson, "*Some design principles: Betabellin*", in D.L. Oxender and C.F. Fox (Eds.), "*Protein Engineering*", Alan R. Liss, 1987, p. 149-163

Incredulase

Analysis of Macromolecular Geometry & Motions

W Krebs, C Chothia,
J Tsai, Y Harpaz, R Taylor

bioinfo.mbb.yale.edu/MolMovDB

1 Database of Macromolecular Motions

Interface packing, shear v. hinge, web DB

2 Morph Movie Server

Restrained Interpolation between submitted endpoint conformations, standardized stats, hinge finder

3 Geometry Software (Volumes)

Voronoi volumes, relation to surface, radii problem, packing measurement

4 Standard Radii & Volumes

Needed for volume calculation, 18 basic types, Sensitivity Analysis, radii set, atom selection criteria, structure set