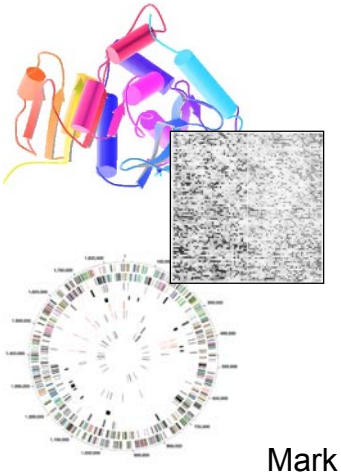
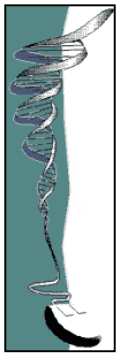
BIOINFORMATICS
Structures





Mark Gerstein, Yale University
bioinfo.mbb.yale.edu/mbb452a

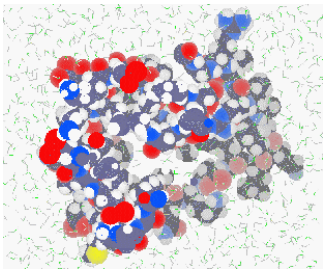
Contents: Structures

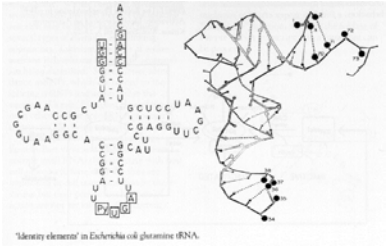
- What Structures Look Like?
- Structural Alignment by Iterated Dynamic Programming
  - ◊ RMS Superposition
  - ◊ Rotating and Translating Structures
- Scoring Structural Similarity
- Other Aspects of Structural Alignment
  - ◊ Distance Matrix based methods
- Elaborating structures
  - ◊ Surfaces and volumes

Molecular Biology Information: Macromolecular Structure

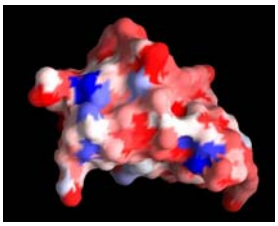
- DNA/RNA/Protein
  - ◊ Almost all protein

(RNA Adapted From D Soll Web Page,
Right Hand Top Protein from M Levitt web page)





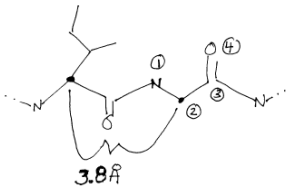
'Identity elements' in Escherichia coli glutamine tRNA.



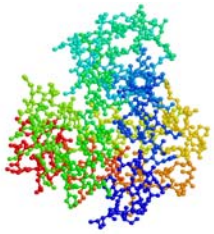
Molecular Biology Information: Protein Structure Details

- Statistics on Number of XYZ triplets
  - ◊ 200 residues/domain –> 200 CA atoms, separated by 3.8 A
  - ◊ Avg. Residue is Leu: 4 backbone atoms + 4 sidechain atoms, 150 cubic A
    - o => ~1500 xyz triplets (=8x200) per protein domain
  - ◊ 10 K known domain, ~300 folds

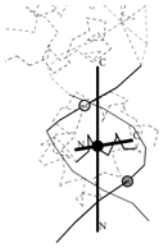
```
ATOM      1  C   ACE     0       9.401  30.166  60.595  1.00 49.88      1GKY  67
ATOM      2  O   ACE     0      10.432  30.832  60.722  1.00 50.35      1GKY  68
ATOM      3  CH3 ACE     0       8.876  29.767  59.226  1.00 50.04      1GKY  69
ATOM      4  N   SER     1       8.753  29.755  61.685  1.00 49.13      1GKY  70
ATOM      5  CA  SER     1       9.242  30.200  62.974  1.00 46.62      1GKY  71
ATOM      6  C   SER     1      10.453  29.500  63.579  1.00 41.99      1GKY  72
ATOM      7  O   SER     1      10.593  29.607  64.814  1.00 43.24      1GKY  73
ATOM      8  CB  SER     1       8.052  30.189  63.974  1.00 53.00      1GKY  74
ATOM      9  OG  SER     1       7.294  31.409  63.930  1.00 57.79      1GKY  75
ATOM     10  N   ARG     2      11.360  28.819  62.827  1.00 36.48      1GKY  76
ATOM     11  CA  ARG     2      12.548  28.316  63.532  1.00 30.20      1GKY  77
ATOM     12  C   ARG     2      13.502  29.501  63.500  1.00 25.54      1GKY  78
...
ATOM   1444  CB  LYS   186      13.836  22.263  57.567  1.00 55.06      1GKY1510
ATOM   1445  CG  LYS   186      12.422  22.452  58.180  1.00 53.45      1GKY1511
ATOM   1446  CD  LYS   186      11.531  21.198  58.185  1.00 49.88      1GKY1512
ATOM   1447  CE  LYS   186      11.452  20.402  56.860  1.00 48.15      1GKY1513
ATOM   1448  NZ  LYS   186      10.735  21.104  55.811  1.00 48.41      1GKY1514
ATOM   1449  OXT LYS   186      16.887  23.841  56.647  1.00 62.94      1GKY1515
TER    1450      LYS   186                                              1GKY1516
```



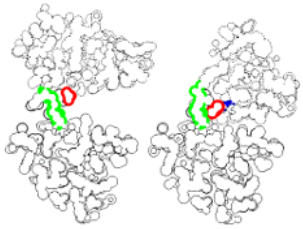
Other Aspects of Structure, Besides just Comparing Atom Positions



Atom
Position,
XYZ triplets



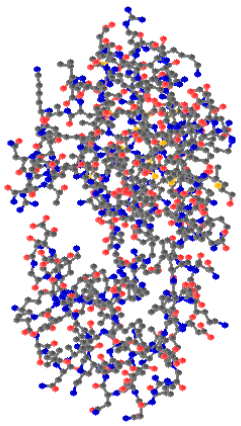
Lines, Axes,
Angles



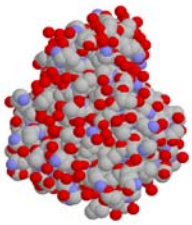
Surfaces, Volumes

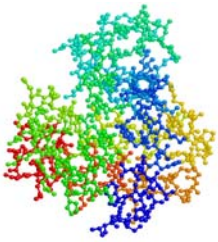
What is Protein Geometry?

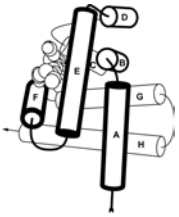
- Coordinates (X, Y, Z's)
- Derivative Concepts
  - ◊ Distance, Surface Area, Volume, Cavity, Groove, Axes, Angle, &c
- Relation to
  - ◊ Function, Energies (E(x)), Dynamics (dx/dt)



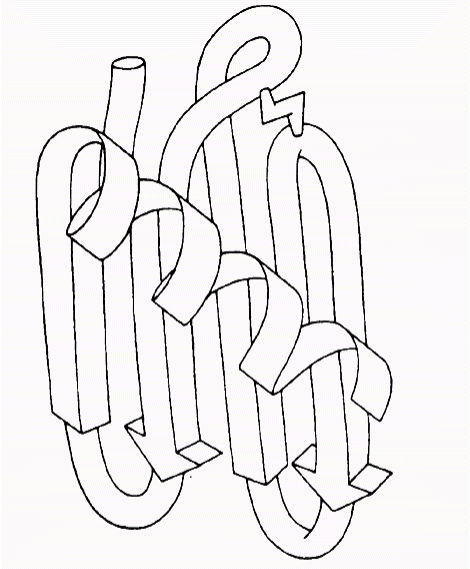
Depicting Protein Structure: Sperm Whale Myoglobin







Incredulase



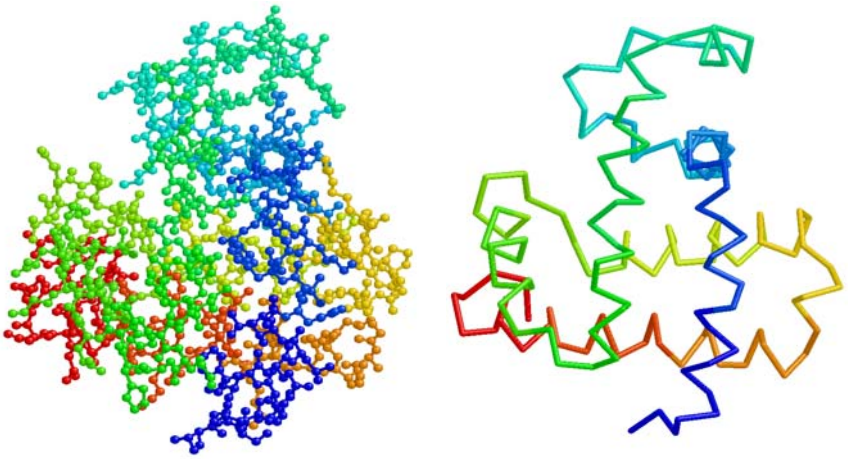
Incredulase

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

Structure alignment - Method

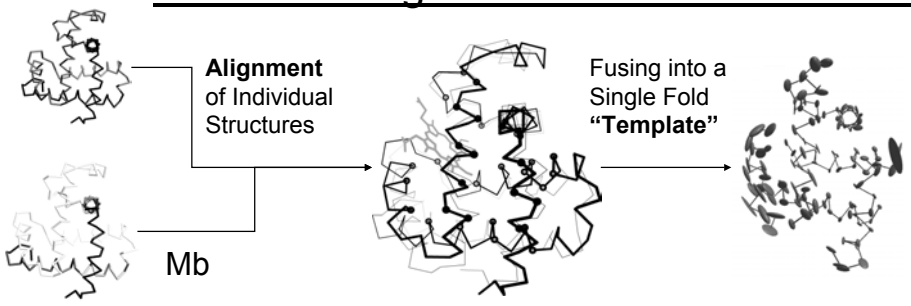
- What Structures Look Like?
- Structural Alignment by Iterated Dynamic Programming
  - ◊ RMS Superposition
- Scoring Structural Similarity
- Other Aspects of Structural Alignment
  - ◊ Distance Matrix based methods
  - ◊ Fold Library
- Relation of Sequence Similarity to Structural and Functional Similarity
- Protein Geometry
- Surface I (Calculation)
- Calculation of Volume
- Voronoi Volumes & Packing
- Standard Volumes & Radii
- Surfaces II (Relationship to Volumes)
- Other Applications of Volumes -- Motions, Docking

Sperm Whale Myoglobin



Structural Alignment of Two Globins

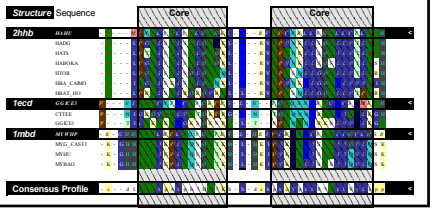
Automatic Alignment to Build Fold Library




```
Hb VLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHF-DLS-----HGSAQVKGHGKKVADALTNAV
     ||| ..   |   |.|| |  .  | .  | | | |   |  |  |       .|  .| ||  |  ||   .
Mb VLSEGEWQLVLHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDRFKHLKTEAEMKASEDLKKHGVTVLTALGAIL

Hb AHVD-DMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKYR------
             |  | . || | .. .     .|  ..  |    |..|     . . | |      .  ||.
Mb KK-KGHHEAELKPLAQSHATKHKIPIKYLEFISEAIIHVLHSRHPGDFGADAQGAMNKALELFRKDIAAKYKELGYQG
```

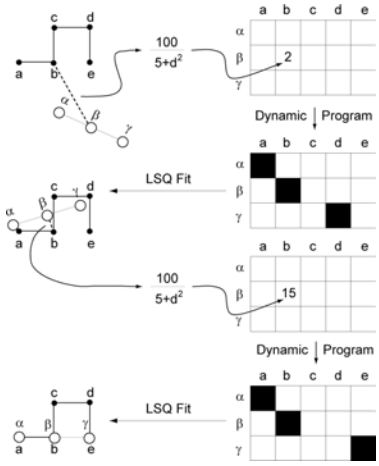
Elements: Domain definitions; Aligned structures, collecting together Non-homologous Sequences; Core annotation



Previous work: Remington, Matthews '80; **Taylor, Orengo '89**, '94; Artymiuk, Rice, Willett '89; Sali, Blundell, '90; Vriend, Sander '91; Russell, Barton '92; **Holm, Sander '93**; Godzik, Skolnick '94; Gibrat, Madej, Bryant '96; Falicov, F Cohen, '96; Feng, Sippl '96; G Cohen '97; Singh & Brutlag, '98

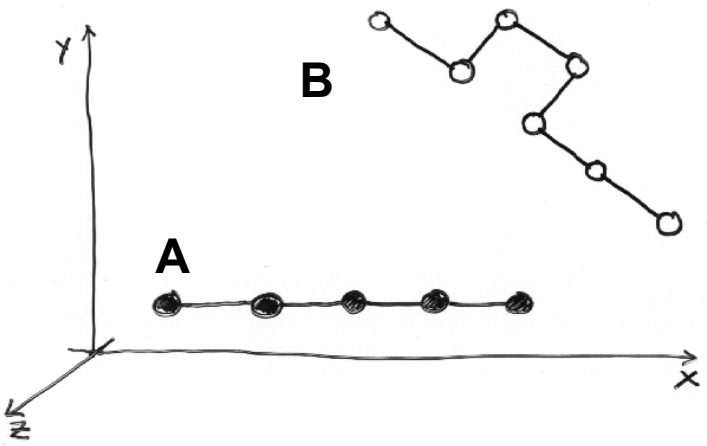
Automatically Comparing Protein Structures

- Given
  2 Structures (A & B),
  2 Basic
  Comparison Operations
  1 Given an alignment optimally **SUPERIMPOSE** A onto B
    Find Best R & T to move A onto B
  2 **Find an Alignment** between A and B based on their 3D coordinates



Core

RMS Superposition (1)

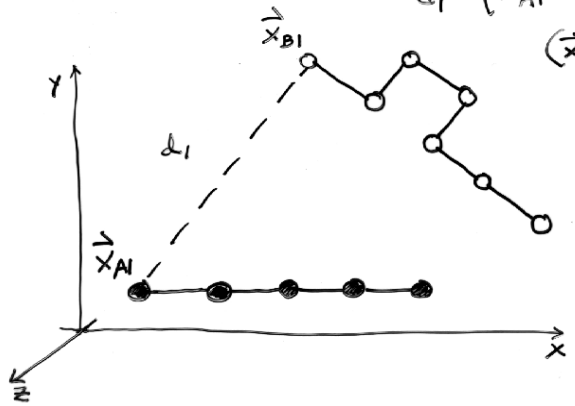


Core

RMS Superposition (2): Distance Between an Atom in 2 Structures

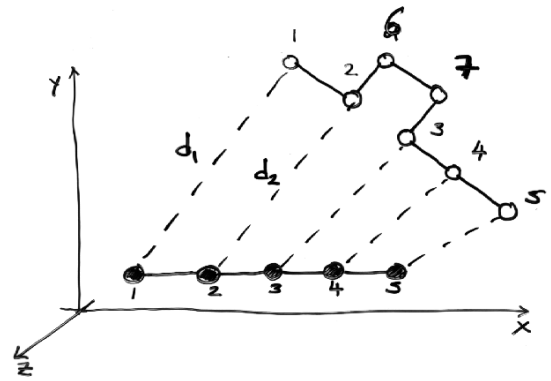
Core

$$d_1^2 = (\vec{x}_{A1} - \vec{x}_{B1}) \cdot (\vec{x}_{A1} - \vec{x}_{B1})$$



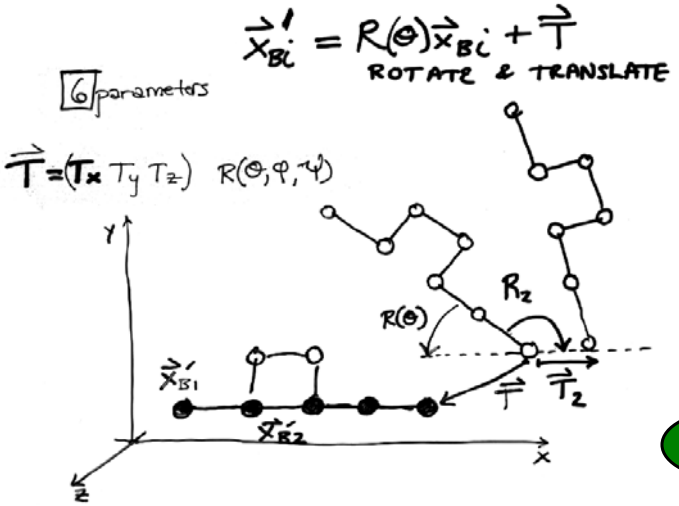
RMS Superposition (3): RMS Distance Between Aligned Atoms in 2 Structures

$$RMS = \sqrt{\frac{\sum_{i=1}^{5} (\vec{x}_{Ai} - \vec{x}_{Bi})^2}{5}} \approx \frac{d_1 + d_2 + d_3 + d_4 + d_5}{5}$$



Core

RMS Superposition (4): Rigid-Body Rotation and Translation of One Structure (B)



Core

RMS Superposition (5): Optimal Movement of One Structure to Minimize the RMS

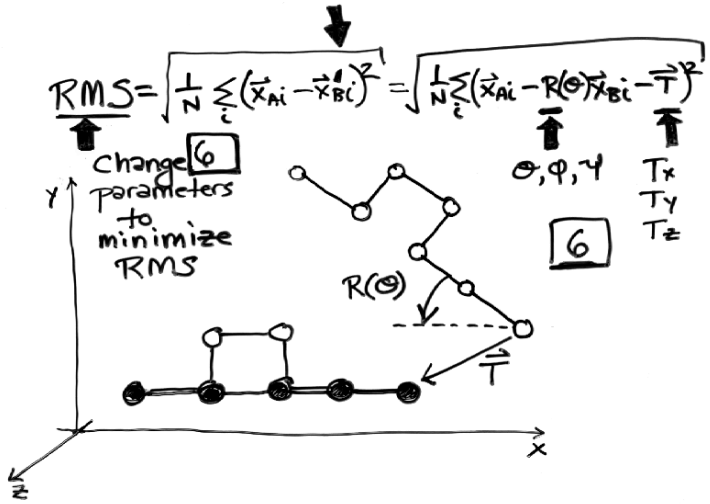
Core

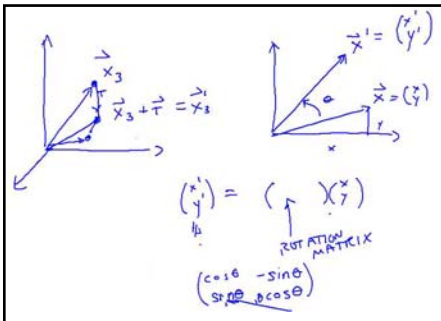
Methods of Solution:

springs (F ~ kx)

SVD

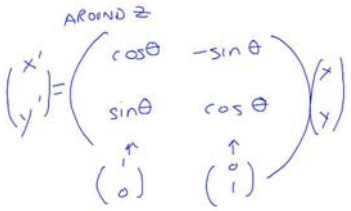
Kabsch



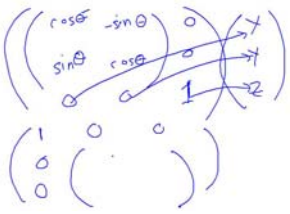


Rotation Matrices

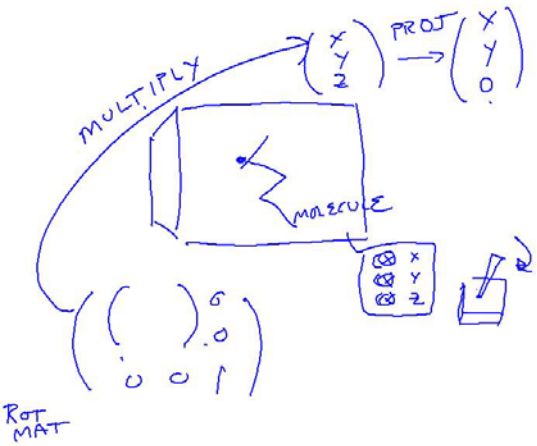
2D

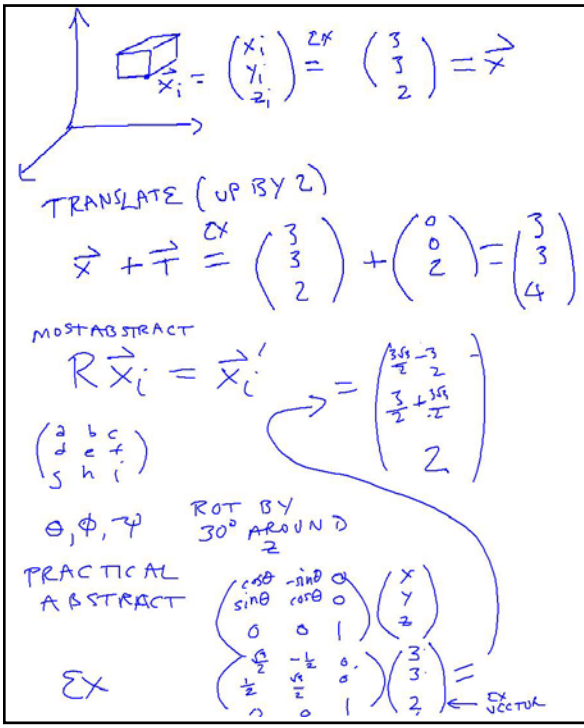


3D



Computer Graphics Systems: Rotation, Translation, and XY Projection





Worked Example of Rigid Body Movement

End of class 2002,11.11

(Bioinfo-9)

[starting in sequences handout]

Alignment (1)
Make a Similarity Matrix
(Like Dot Plot)

	A	B	C	N	Y	R	Q	C	L	C	R	P	M
A	1												
Y					1								
C			1					1		1			
Y					1								
N				1									
R						1					1		
C			1					1		1			
K													
C			1					1		1			
R						1					1		
B		1											
P												1	

Structural Alignment (1b)
Make a Similarity Matrix
(Generalized Similarity Matrix)

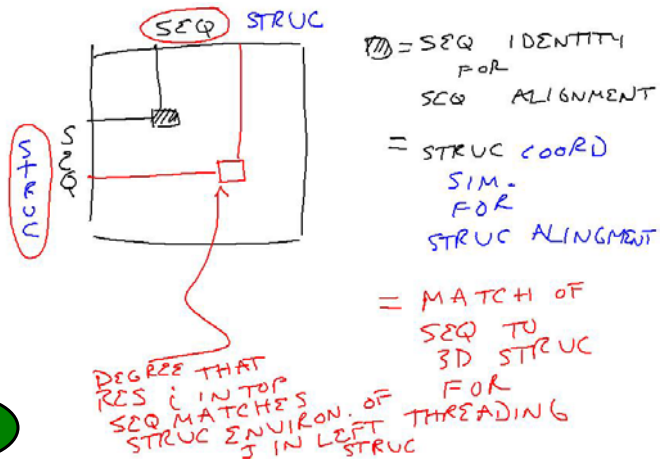
- PAM(A,V) = 0.5
  - ◊ Applies at every position
- S(aa @ i, aa @ J)
  - ◊ Specific Matrix for each pair of residues
    i in protein 1 and
    J in protein 2
  - ◊ Example is Y near N-term. matches any C-term. residue (Y at J=2)
- S(i,J)
  - ◊ Doesn't need to depend on a.a. identities at all!
  - ◊ Just need to make up a score for matching residue i in protein 1 with residue J in protein 2

i →

		1	2	3	4	5	6	7	8	9	10	11	12	13
		A	B	C	N	Y	R	Q	C	L	C	R	P	M
1	A	1												
2	**Y**					1			5	5	5	5	5	5
3	C			1					1		1			
4	Y					1								
5	N				1									
6	R						1					1		
7	C			1					1		1			
8	K													
9	C			1					1		1			
10	R						1					1		
11	B		1											
12	P												1	

J ↓

Seq. Alignment, Struc. Alignment, Threading



Core

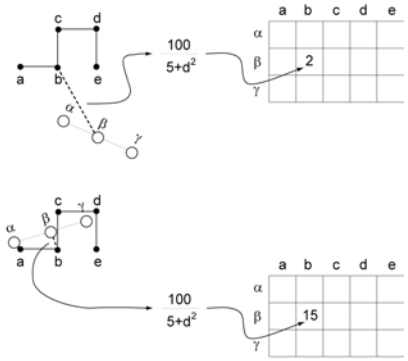
Structural Alignment (1c*) Similarity Matrix for Structural Alignment

- Structural Alignment
  - ◊ Similarity Matrix S(i,J) depends on the 3D coordinates of residues i and J
  - ◊ Distance between CA of i and J

$$d = \sqrt{(x_i - x_J)^2 + (y_i - y_J)^2 + (z_i - z_J)^2}$$

  - ◊ M(i,j) = 100 / (5 + d^2)

- Threading
  - ◊ S(i,J) depends on the how well the amino acid at position i in protein 1 fits into the 3D structural environment at position J of protein 2



Alignment (2): Dynamic Programming, Start Computing the Sum Matrix

```
new_value_cell(R,C) <=
  cell(R,C)                       { Old value, either 1 or 0    }
  + Max[
        cell (R+1, C+1),          { Diagonally Down, no gaps    }
        cells(R+1, C+2 to C_max),{ Down a row, making col. gap }
        cells(R+2 to R_max, C+2) { Down a col., making row gap }
       ]
```

	A	B	C	N	Y	R	Q	C	L	C	R	P	M
A	**1**												
Y					**1**								
C			**1**					**1**		**1**			
Y					**1**								
N				**1**									
R						**1**					**1**		
C			**1**					**1**		**1**			
K													
C			**1**					**1**		**1**			
R						**1**					**1**		
B		**1**											
P												**1**	

	A	B	C	N	Y	R	Q	C	L	C	R	P	M
A	1												
Y					1								
C			1					1		1			
Y					1								
N				1									
R						1					1		
C			1					1		1			
K													
C			1					1		1			
R						1					*2*	**0**	**0**
B	**1**	**2**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**0**	**0**
P	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**0**

Alignment (3):Dynamic Programming, Keep Going

	A	B	C	N	Y	R	Q	C	L	C	R	P	M
A	1												
Y					1								
C			1					1		1			
Y					1								
N				1									
R						1					**1**		
C			1					1		1			
K													
C			1					1		1			
R						1					*2*	**0**	**0**
B	**1**	**2**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**0**	**0**
P	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**0**

	A	B	C	N	Y	R	Q	C	L	C	R	P	M
A	1												
Y					1								
C			1					1		1			
Y					1								
N				1									
R						*5*	**4**	**3**	**3**	**2**	**2**	**0**	**0**
C	**3**	**3**	**4**	**3**	**3**	**3**	**3**	**4**	**3**	**3**	**1**	**0**	**0**
K	**3**	**3**	**3**	**3**	**3**	**3**	**3**	**3**	**3**	**2**	**1**	**0**	**0**
C	**2**	**2**	**3**	**2**	**2**	**2**	**2**	**3**	**2**	**3**	**1**	**0**	**0**
R	**2**	**1**	**1**	**1**	**1**	**2**	**1**	**1**	**1**	**1**	**2**	**0**	**0**
B	**1**	**2**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**0**	**0**
P	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**0**

Alignment (4): Dynamic Programming, Sum Matrix All Done

	A	B	C	N	Y	R	Q	C	L	C	R	P	M
A	1												
Y					1								
C			1					1		1			
Y					1								
N				1									
R						5	4	3	3	2	2	0	0
C	3	3	4	3	3	3	3	4	3	3	1	0	0
K	3	3	3	3	3	3	3	3	3	2	1	0	0
C	2	2	3	2	2	2	2	3	2	3	1	0	0
R	2	1	1	1	1	2	1	1	1	1	2	0	0
B	1	2	1	1	1	1	1	1	1	1	1	0	0
P	0	0	0	0	0	0	0	0	0	0	0	1	0

	A	B	C	N	Y	R	Q	C	L	C	R	P	M
A	8	7	6	6	5	4	4	3	3	2	1	0	0
Y	7	7	6	6	6	4	4	3	3	2	1	0	0
C	6	6	7	6	5	4	4	4	3	3	1	0	0
Y	6	6	6	5	6	4	4	3	3	2	1	0	0
N	5	5	5	6	5	4	4	3	3	2	1	0	0
R	4	4	4	4	4	5	4	3	3	2	2	0	0
C	3	3	4	3	3	3	3	4	3	3	1	0	0
K	3	3	3	3	3	3	3	3	3	2	1	0	0
C	2	2	3	2	2	2	2	3	2	3	1	0	0
R	2	1	1	1	1	2	1	1	1	1	2	0	0
B	1	2	1	1	1	1	1	1	1	1	1	0	0
P	0	0	0	0	0	0	0	0	0	0	0	1	0

Alignment (5): Traceback

Find Best Score (8) and Trace Back

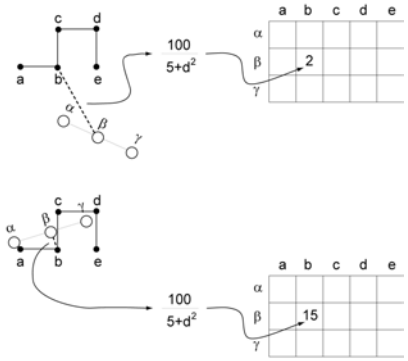
A B C N Y - R Q C L C R - P M

A Y C - Y N R - C K C R B P

	A	B	C	N	Y	R	Q	C	L	C	R	P	M
A	8	7	6	6	5	4	4	3	3	2	1	0	0
Y	7	7	6	6	6	4	4	3	3	2	1	0	0
C	6	6	7	6	5	4	4	4	3	3	1	0	0
Y	6	6	6	5	6	4	4	3	3	2	1	0	0
N	5	5	5	6	5	4	4	3	3	2	1	0	0
R	4	4	4	4	4	5	4	3	3	2	2	0	0
C	3	3	4	3	3	3	3	4	3	3	1	0	0
K	3	3	3	3	3	3	3	3	3	2	1	0	0
C	2	2	3	2	2	2	2	3	2	3	1	0	0
R	2	1	1	1	1	2	1	1	1	1	2	0	0
B	1	2	1	1	1	1	1	1	1	1	1	0	0
P	0	0	0	0	0	0	0	0	0	0	0	1	0

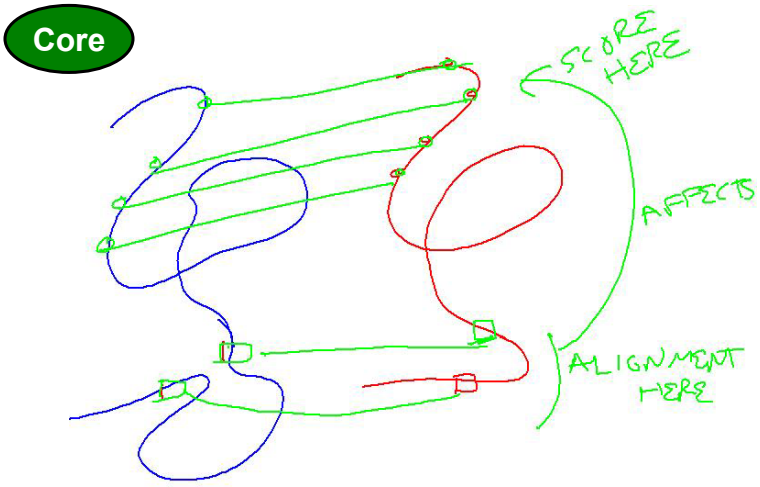
In Structural Alignment, Not Yet Done (Step 6*)

- Use Alignment to LSQ Fit Structure B onto Structure A
  - ◊ However, movement of B will now change the Similarity Matrix
- This Violates Fundamental Premise of Dynamic Programming
  - ◊ Way Residue at i is aligned can now affect previously optimal alignment of residues (from 1 to i-1)



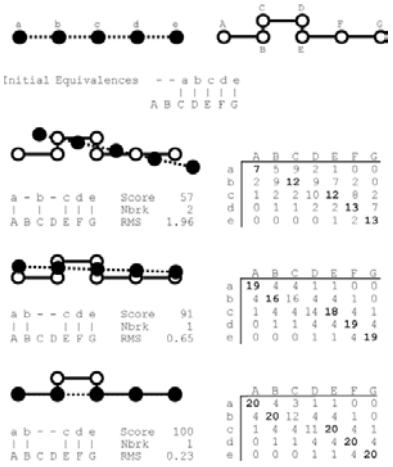

```
ACSQRP--LRV-SH  -R  SENCV
A-SNKPQLVKLMTH  VK  DFCV-
```


How central idea of dynamic programming is violated in structural alignment



Structural Alignment (7*), Iterate Until Convergence

1 Compute Sim. Matrix
2 Align via Dyn. Prog.
3 RMS Fit Based on Alignment
4 Move Structure B
5 Re-compute Sim. Matrix
6 If changed from #1, GOTO #2



Structure alignment - Scoring

- What Structures Look Like?
- Structural Alignment by Iterated Dynamic Programming
  - ◊ RMS Superposition
- Scoring Structural Similarity
- Other Aspects of Structural Alignment
  - ◊ Distance Matrix based methods
  - ◊ Fold Library
- Relation of Sequence Similarity to Structural and Functional Similarity
- Protein Geometry
- Surface I (Calculation)
- Calculation of Volume
- Voronoi Volumes & Packing
- Standard Volumes & Radii
- Surfaces II (Relationship to Volumes)
- Other Applications of Volumes -- Motions, Docking

Score S at End Just Like SW Score, but also have final RMS

S = Total Score

S(i,j) = similarity matrix score for aligning i and j

Sum is carried out over all aligned i and j

n = number of gaps (assuming no gap ext. penalty)

G = gap penalty

$$S = \sum_{i,j} S(i,j) - nG$$

Some Similarities are Readily Apparent others are more Subtle

Easy: Globins

125 res., ~1.5 Å

Tricky: Ig C & V

85 res., ~3 Å

Very Subtle: G3P-dehydrogenase, C-term. Domain

>5 Å

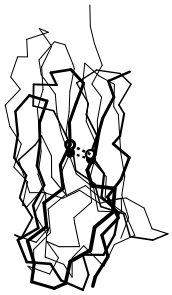
Some Similarities are Readily Apparent others are more Subtle

Easy: Globins

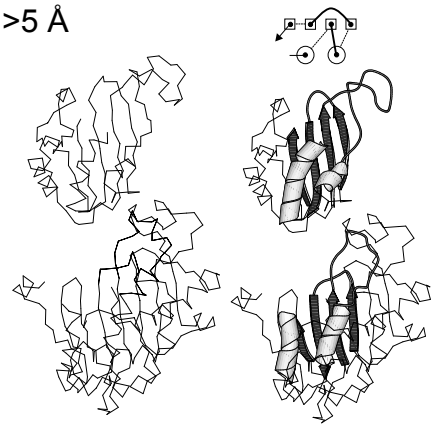
125 res., ~1.5 Å

Tricky: Ig C & V

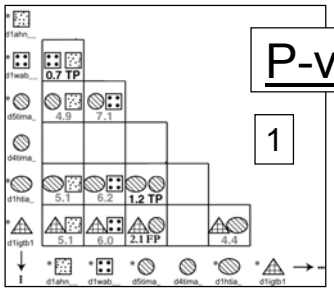
85 res., ~3 Å



Very Subtle: G3P-dehydro-genase, C-term. Domain >5 Å



P-values

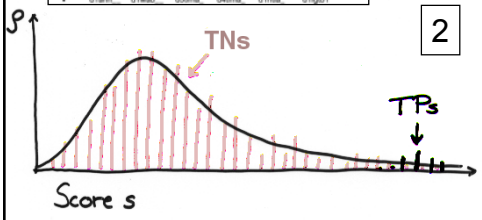


1

Extra

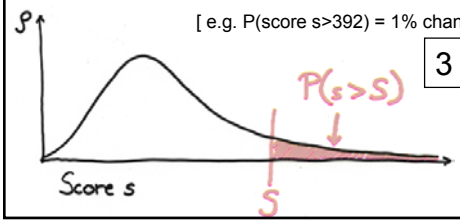
- Significance Statistics
  - ◊ For sequences, originally used in Blast (Karlin-Altschul). Then in FASTA, &c.
  - ◊ Extrapolated Percentile Rank: How does a Score Rank Relative to all Other Scores?
- Our Strategy: Fit to Observed Distribution
  1) All-vs-All comparison
  2) Graph Distribution of Scores in 2D (N dependence); 1K x 1K families -> ~1M scores; ~2K included TPs
  3) Fit a function ρ(S) to TN distribution (TNs from scop); Integrating ρ gives P(s>S), the CDF, chance of getting a score better than threshold S randomly
  4) Use same formalism for sequence & structure

2



[e.g. P(score s>392) = 1% chance]

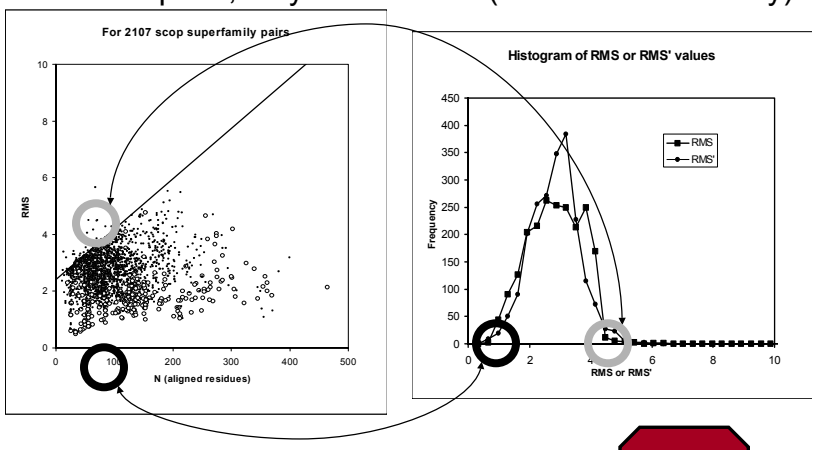
3



Statistics on Range of Similarities

For 2107 pairs, only 2% Outliers (with subtle similarity)

RMS

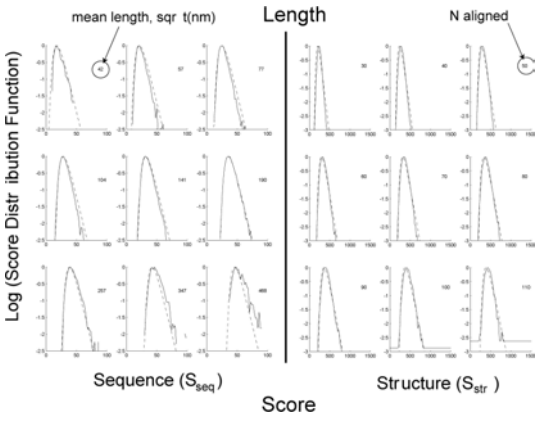


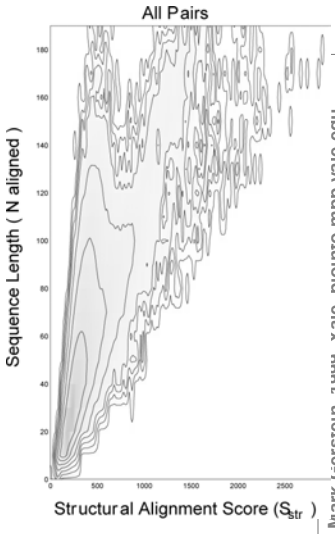
Num. Aligned

Extra

Scores from Structural Alignment Distributed Just Like Ones from Sequence Alignment (E.V.D.)

Extra





Same Results for Sequence & Structure

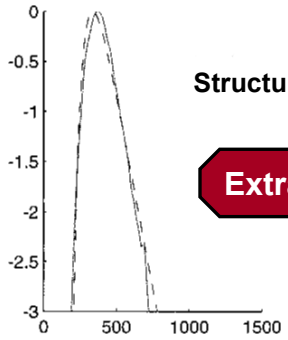
3 Free Parm. fit to EVD involving: ***a, b,σ***. These are the only difference betw. sequence and structure.

$$Z = \frac{S - (a \ln N + b)}{\sigma}$$

$$S = \sum_{i,j} M(i,j) - G$$

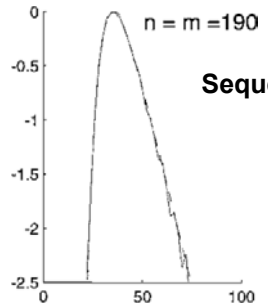
$$\rho(z) = \exp\left(-z - e^{-z}\right)$$

N, G, M also defined differently for sequence and structure.
N = number of residues matched.
G = total gap penalty.
M(i,j) = similarity matrix
(Blossum for seq. or M_{str}(i,j) struc.)



Structure

Extra



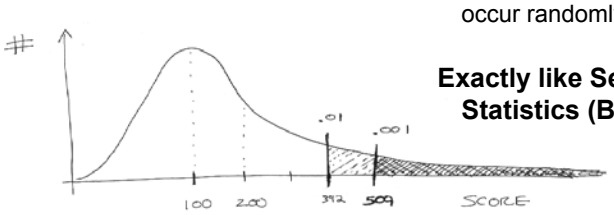
Sequence

Score Significance (P-value) derived from Extreme Value Distribution (just like BLAST, FASTA)

F(s) = E.V.D of scores
F(s) = exp(-Z(s) - exp(-Z(s)))

Z(s) = As + ln(N) + B
s = Score from random alignment
N length of sequence matched
A & B are fit parameters

P(s>S) = CDF = integral[F(s)]
P(s>S) = 1 - exp(-exp(-Z(s)))
Given Score S (1%), P (s > S) is the chance that a given random score **s** is greater than the threshold
i.e. P-value gives chance score would occur randomly



Exactly like Sequence Matching Statistics (BLAST and FASTA)

Extra

RMS is a similarity Score

Extra

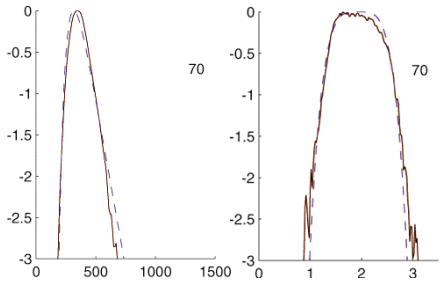
- Also, RMS doesn't work instead of structural alignment (no EVD fit)
  - ◊ RMS penalizes worst fitting atoms, easily skewed

S_{str}

$$\sum \frac{100}{5+\mathbf{d}_i^2} vs$$

RMS

$$\sqrt{\sum \mathbf{d}_i^2}$$



Structure alignment - Other methods

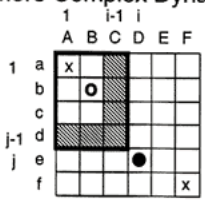
- What Structures Look Like?
- Structural Alignment by Iterated Dynamic Programming
  - ◊ RMS Superposition
- Scoring Structural Similarity
- Other Aspects of Structural Alignment
  - ◊ Distance Matrix based methods
  - ◊ Fold Library
- Relation of Sequence Similarity to Structural and Functional Similarity
- Protein Geometry
- Surface I (Calculation)
- Calculation of Volume
- Voronoi Volumes & Packing
- Standard Volumes & Radii
- Surfaces II (Relationship to Volumes)
- Other Applications of Volumes -- Motions, Docking

Refine Method

Extra

- Multiple Aligment by aligning to central structure

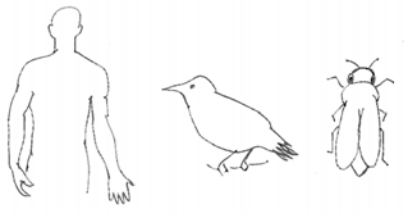
- More Complex Dynamic Programming



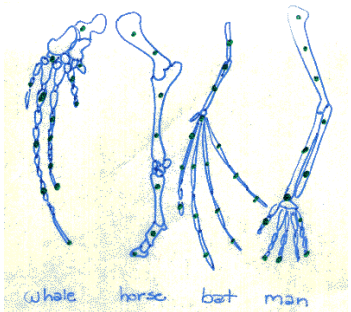
- Find "best" aligned regions
  - "Core-finding" to remove outliers
  - "Noisy" suboptimal paths

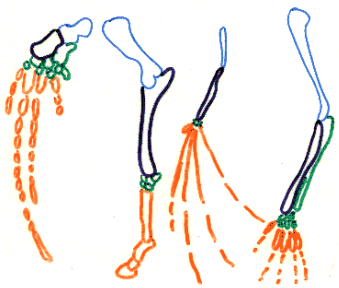
	A	y	B	C	D	E
A	2	0	0	0	0	0
B	0	2	3	1	1	1
C	0	1	2	5	2	2
x	0	1	1	2	5	4
D	0	1	1	2	6	5
E	0	1	1	2	4	8

Significance Ignoring Crucial Features in Structural Similarity



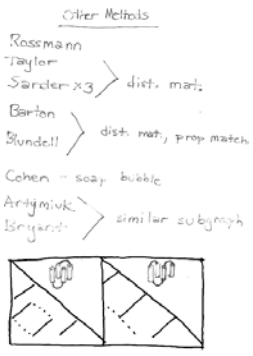
Extra

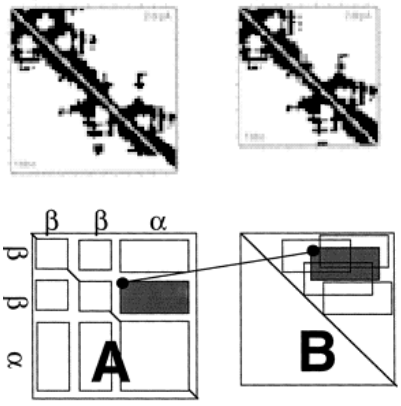




Other Methods of Structural Alignment

- RMS fitting used universally, but other alignment methods
- Comparison of Distance Matrices
  - ◊ Holm & Sander, DALI
  - ◊ Taylor & Orengo

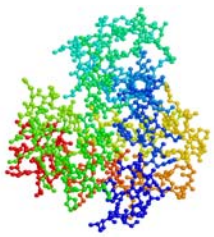




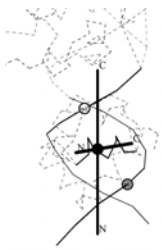
Structure Hashing
- Bryant, VAST
- Rice, Artymiuk

Others
- Cohen (Soap)
- Sippl
- Godzik (Lattice)

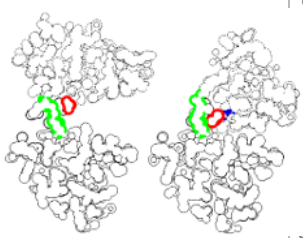
Other Aspects of Structure, Besides just Comparing Atom Positions



Atom Position, XYZ triplets



Lines, Axes, Angles



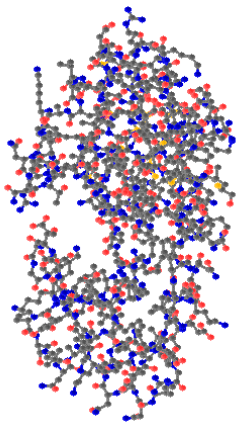
Surfaces, Volumes

Extra

From here to end of Structures all is "extra" unless otherwise marked.

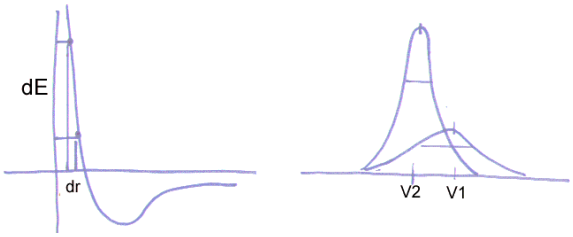
What is Protein Geometry?

- Coordinates (X, Y, Z's)
- Derivative Concepts
  - ◊ Distance, Surface Area, Volume, Cavity, Groove, Axes, Angle, &c
- Relation to
  - ◊ Function, Energies (E(x)), Dynamics (dx/dt)



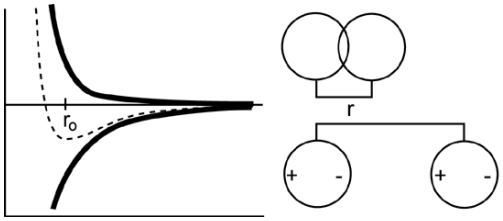
Small Packing Changes Significant

- Exponential dependence
- Bounded within a range of 0.5 (.8 and .3)
- Many observations in standard volumes gives small error about the mean (SD/sqrt(N))



Packing ~ VDW force

- Longer-range isotropic attractive tail provides general cohesion
- Shorter-ranged repulsion determines detailed geometry of interaction
- Billiard Ball model, WCA Theory

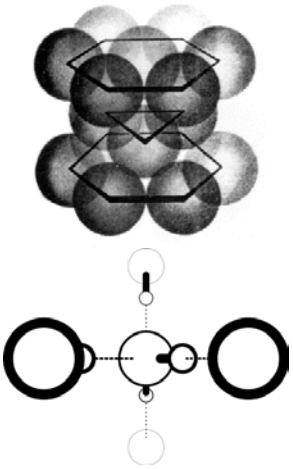


Electron Overlap Replusion $U = \varepsilon \left(\frac{r_0}{r}\right)^{12}$

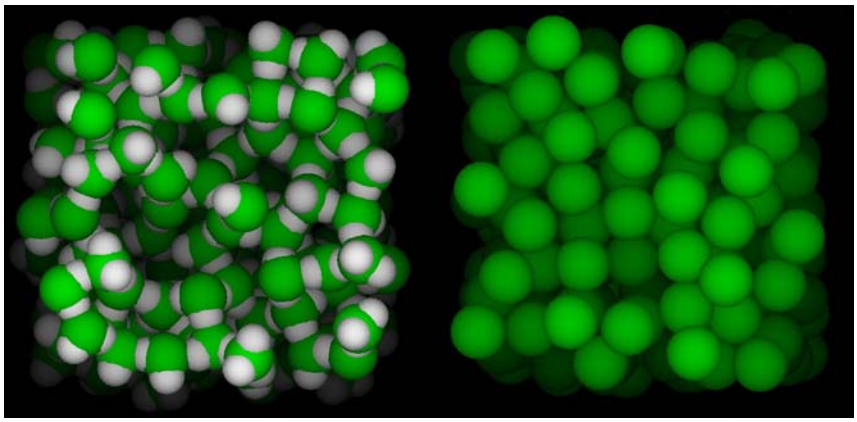
Dispersion Attraction $U = -4\varepsilon \left(\frac{r_0}{r}\right)^{6}$

Close-packing is Default

- No tight packing when highly directional interactions (such as H-bonds) need to be satisfied
- Packing spheres (.74), hexagonal
- Water (~.35), “Open” tetrahedral, H-bonds



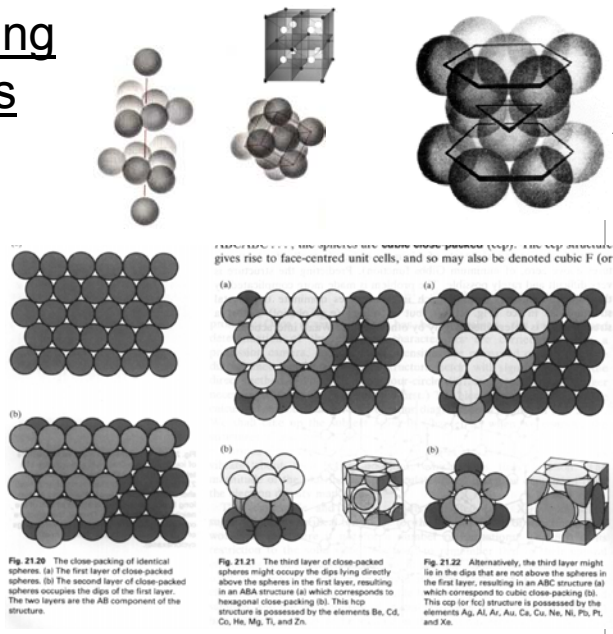
Water v. Argon



More Complex Systems -- what to do?

Close-Packing of Spheres

- Efficiency
  - ◊ Volume Spheres / Volume of space
- Close packed spheres
  - ◊ 74% volume filled
  - ◊ Coordination of 12
  - ◊ Two Ways of laying out
- Fcc
  - ◊ cubic close packing
  - ◊ ABC layers
- hcp
  - ◊ Hexagonally close packed
  - ◊ ABABAB



gives rise to face-centred unit cells, and so may also be denoted cubic F (or

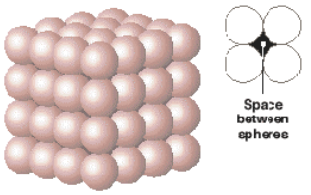
Fig. 21.20 The close-packing of identical spheres. (a) The first layer of close-packed spheres. (b) The second layer of close-packed spheres occupies the dips of the first layer. The two layers are the AB component of the structure.

Fig. 21.21 The third layer of close-packed spheres might occupy the dips lying directly above the spheres in the first layer, resulting in an ABA structure (a) which corresponds to hexagonal close-packing (b). This hcp structure is possessed by the elements Be, Cd, Co, He, Mg, Ti, and Zn.

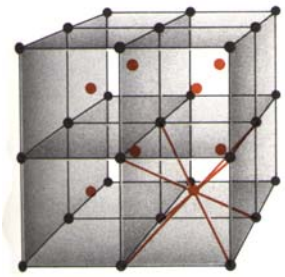
Fig. 21.22 Alternatively, the third layer might lie in the dips that are not above the spheres in the first layer, resulting in an ABC structure (a) which correspond to cubic close-packing (b). This ccp (or fcc) structure is possessed by the elements Ag, Al, Ar, Au, Ca, Cu, Ne, Ni, Pb, Pt, and Xe.

Illustration Credits: Atkins, Pchem, 634

Other Well Known Sphere Arrangements



- Simple cubic packing
  - ◊ 8 nbrs
  - ◊ 52% efficiency
- bcc cubic packing
  - ◊ one sphere sits in middle of 8 others (body-centered)
  - ◊ 8 nbrs
  - ◊ 68% efficiency
- fcc -> bcc -> simple
  - ◊ apx 3/4, 2/3, 1/2



Optimal Packing Finally Proved

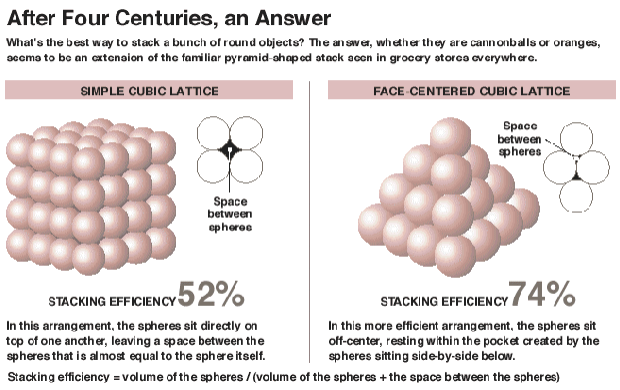
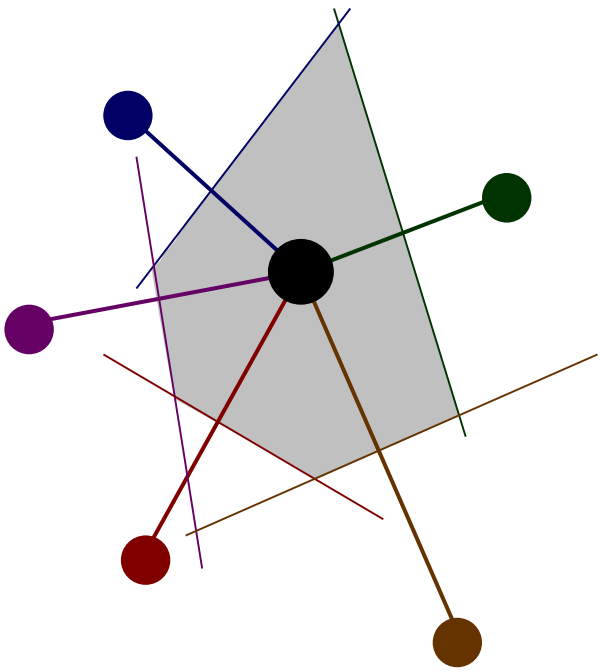
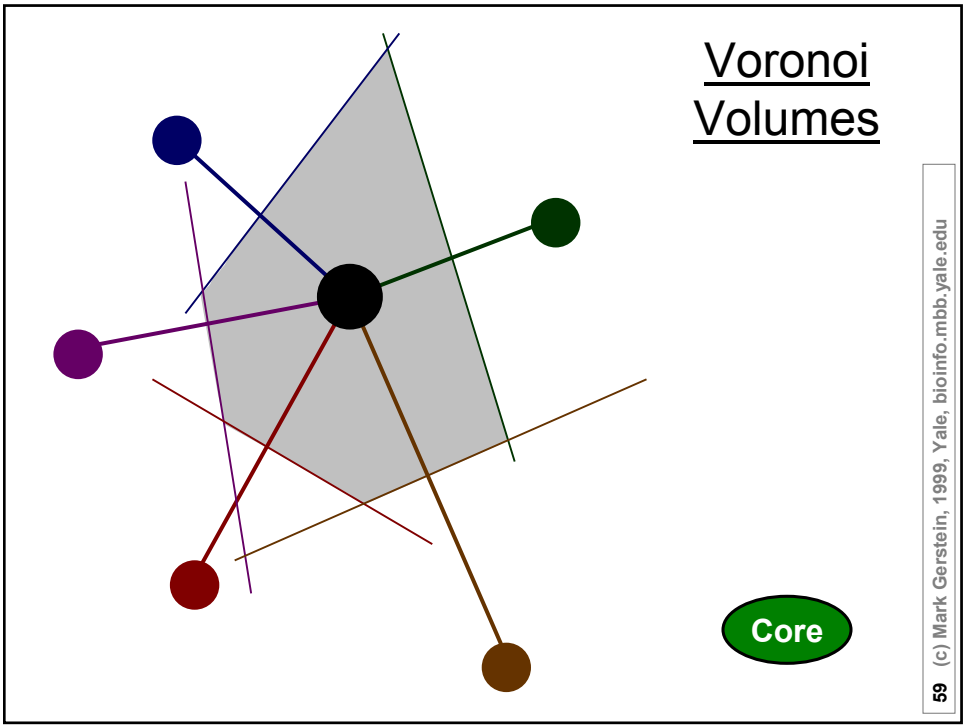


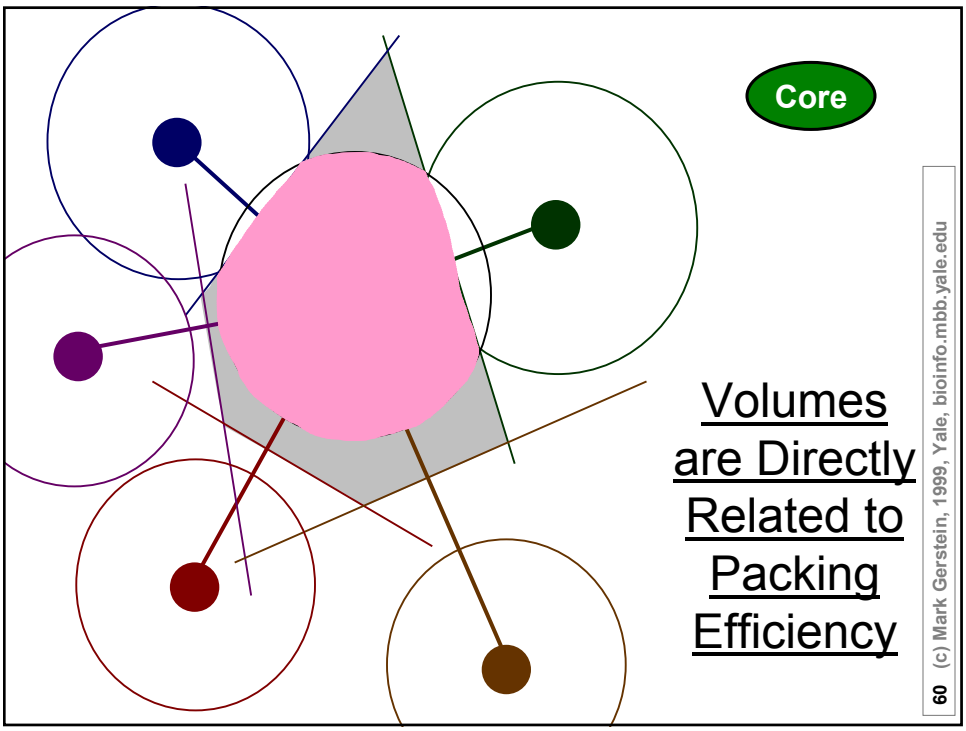
Illustration Credits: Singh, New York Times

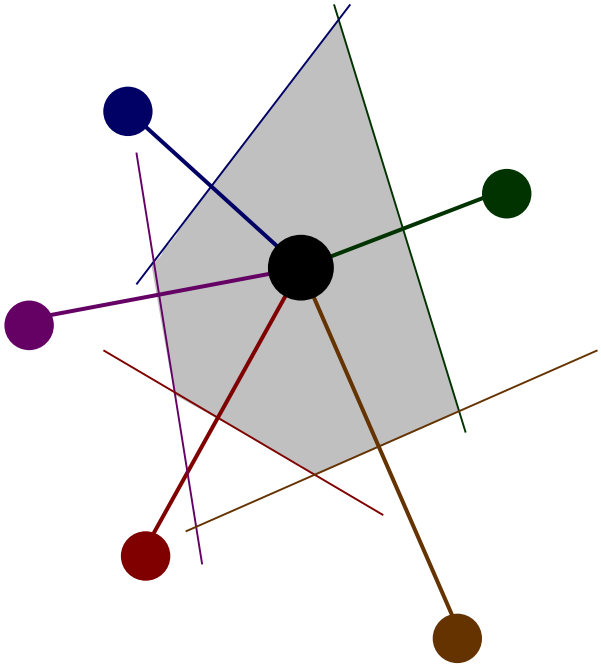
Voronoi Volumes



Core

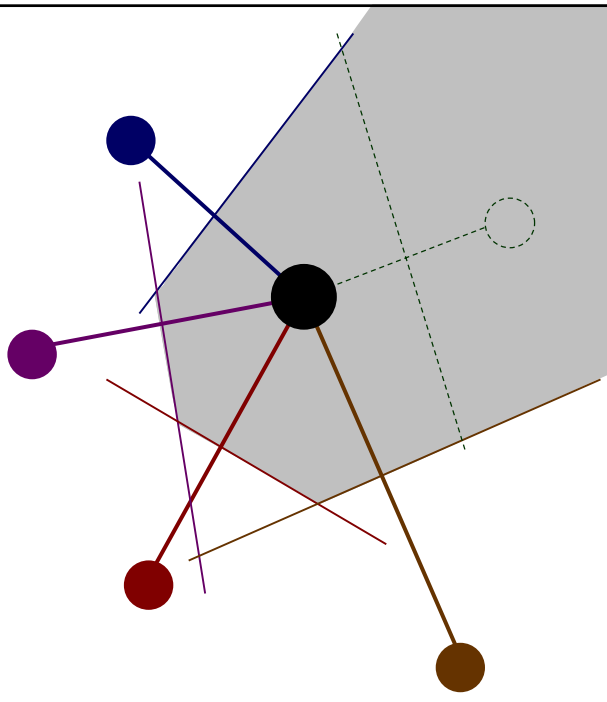

Voronoi
Volumes
Core
59 (c) Mark Gerstein, 1999, Yale, bioinfo.mbb.yale.edu


Core
Volumes
are Directly
Related to
Packing
Efficiency
60 (c) Mark Gerstein, 1999, Yale, bioinfo.mbb.yale.edu



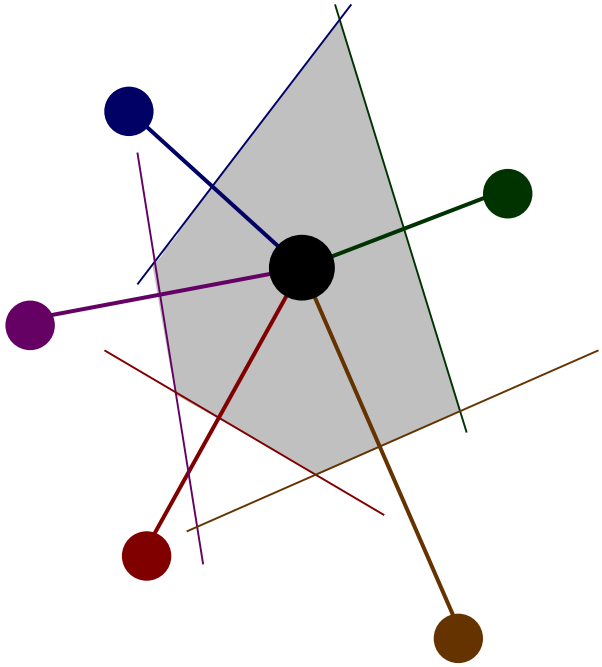
Voronoi Volumes

Core



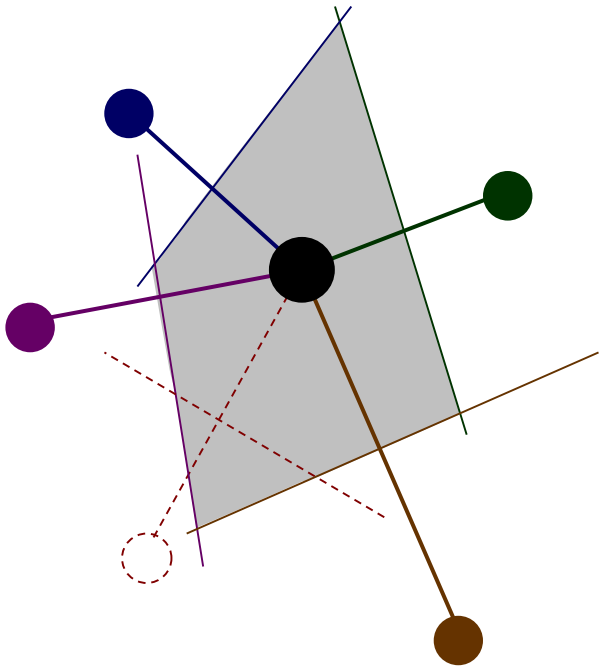
Problem of Protein Surface

Extra



Voronoi Volumes

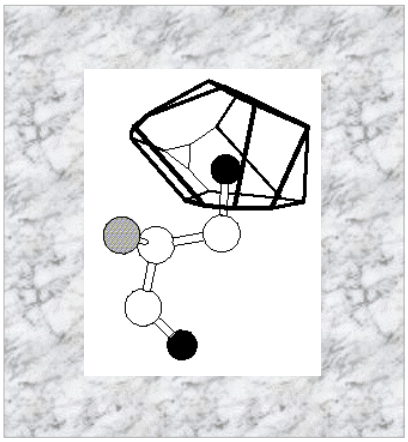
Extra



Missing Atoms Give Looser Packing

Extra

Finding Voronoi Volumes



- Draw Lines in Between Atoms
- Draw Equidistant Planes Between Atoms
- Intersect Planes for a Volume

Classic Papers

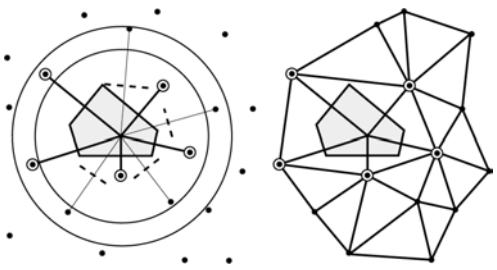
- Lee, B. & Richards, F. M. (1971). "The Interpretation of Protein Structures: Estimation of Static Accessibility," *J. Mol. Biol*. **55**, 379-400.
- Richards, F. M. (1974). "The Interpretation of Protein Structures: Total Volume, Group Volume Distributions and Packing Density," *J. Mol. Biol.* **82**, 1-14.
- Richards, F. M. (1977). "Areas, Volumes, Packing, and Protein Structure," Ann. Rev. Biophys. Bioeng. 6, 151-76.

Properties of Voronoi Polyhedra

- 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.
- 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.

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
- threading

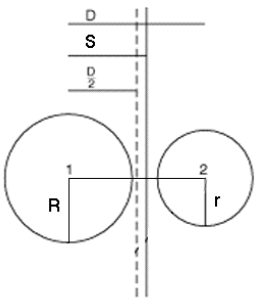


Voronoi diagrams are generally useful, beyond proteins

- Border of D.T. is Convex Hull
- D.T. produces "fatest" possible triangles which makes it convenient for things such as finite element analysis.
- 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
- Voronoi volume of "something" often is a useful weighting factor. This fact can be used, for instance, to weight sequences in alignment to correct for over or under-representation

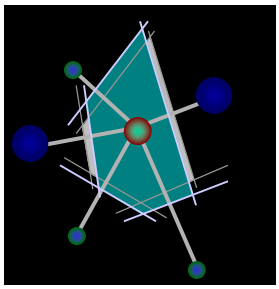
Atoms have different sizes

- 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



Why Type the Atoms?

- Calculate Average Volumes
- Compare to Protein Atoms of Similar Type
- Allows for Modified Voronoi Volumes: Instead of Equidistant Planes, Use the Ratio of Their Radius



Courtesy of N Voss

Set of VDW Radii

- Great differences in a sensitive parameter (Radii for carbon 1.87 vs 2.00)
- Complex calculation: minimizing SD, iterative procedure, from protein structures
- Look for common distances in CCD
- Preliminary Solution

Atom	Bondi	New
C4__	1.87	1.88
C3H1	1.76	1.76
C3H0	1.76	1.61
O1HO	1.40	1.42
O2H1	1.40	1.46
N___	1.65	1.64
S___	1.85	1.77

Different Sets of Radii

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

Voronoi Volumes, the Natural Way to Measure Packing

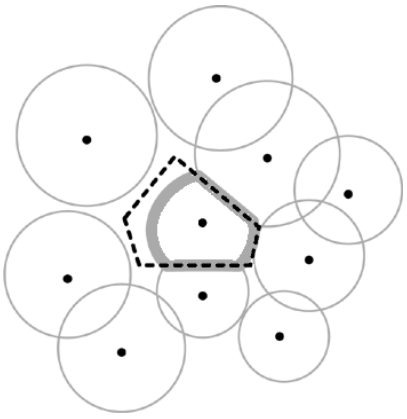
Packing Efficiency

= Volume-of-Object

Space-it-occupies

= V(VDW) / V(Voronoi)

- Absolute v relative eff.
  V1 / V2
- Other methods
  - ◊ Measure Cavity Volume (grids, constructions, &c)



Standard Residue Volumes

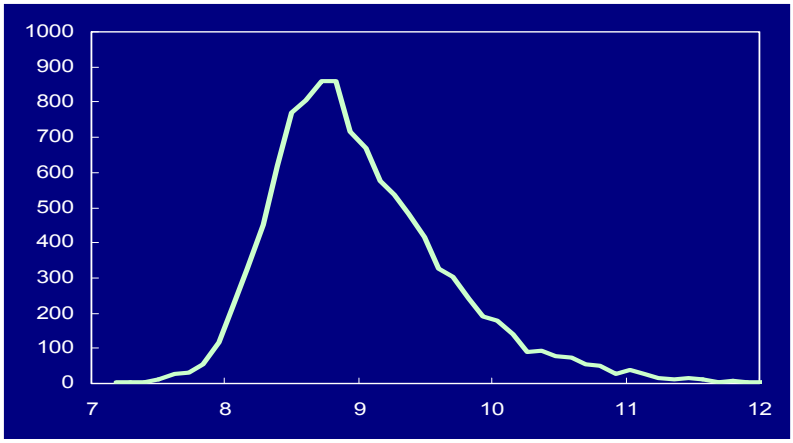
- Database of many hi-res structures (~100, 2 Å)
- Volumes statistics for buried residues (various selections, resample, &c)
- Standard atomic volumes harder… parameter set development...

G	64	c	105	T	120	V	139	H	159	M	168	R	194
A	90	C	113	P	124	E	140	L	165	K	170	Y	198
S	94	D	117	N	128	N	150	I	165	F	193	W	233

Standard Core Volumes (Prelim.)

Atom Types		Num.	Volume (Å^3)	Error (%)
Mainchain Atoms				
carbonyl carbon (except G)	C	8361	9.2	.08
alpha carbon (except G)	CA	7686	13.4	.09
nitrogen (except P)	N	9042	13.9	.09
carbonyl oxygen	O	7831	15.8	.10
Gly C		811	10.2	.27
Gly CA		522	23.5	.39
Pro N		334	8.6	.39
Sidechain atoms				
trigonal or aromatic carbon	>C=	3026	10.3	.13
aromatic CH (H,F,W,Y)	-CH=	4333	21.1	.14
aliphatic CH	>CH-	3411	14.6	.14
methylene group	-CH2-	5427	23.7	.12
methyl group (A,V,L,I)	-CH3	5273	36.7	.11
hydroxyl oxygen (S,T)	-OH	851	17.2	.36
carbonyl oxygen (N,Q)	=O	272	16.8	.76
carboxyl oxygen (D,E)	-O	517	16.0	.53
2° amine (R,H,W)	-NH-	530	15.6	.53
1° amine or amide (R,N,Q)	-NH2	355	23.4	.52
tetrahedral nitrogen (K)	-NH3	31	20.0	1.40
thioether or disulfide (C,M)	-S-	1242	19.3	1.22
sulfhydryl (C)	-SH	67	37.8	1.33

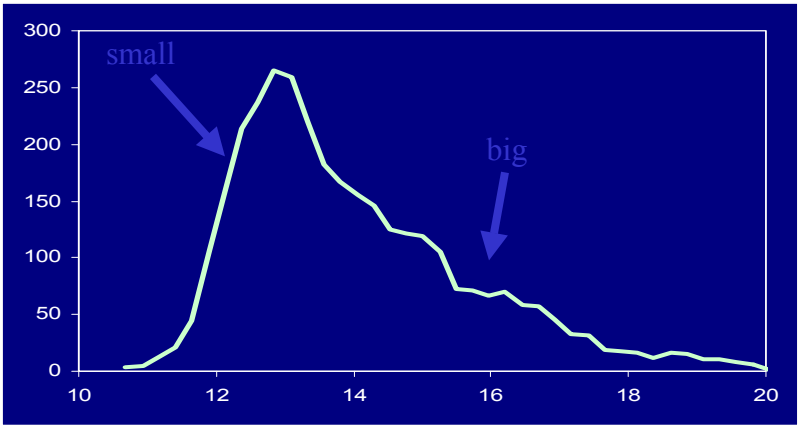
Histogram Analysis of the C3H0s



Volume

Courtesy of N Voss

Histogram Analysis of the N3H1

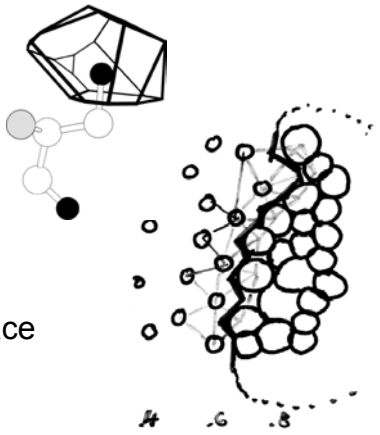


Volume

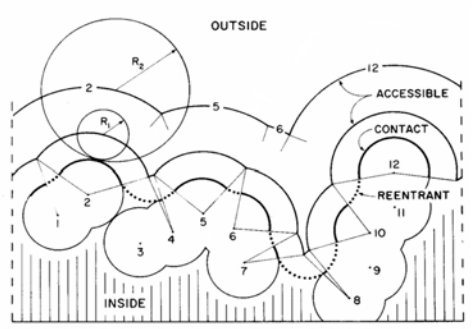
Courtesy of N Voss

Packing at Interfaces

- Voronoi volumes (and D. triangulation) to measure packing
- Tight core packing v. Loose surface packing
- Grooves & ridges: close-packing v. H-bonding
- How packing defines a surface (hydration surface)
- Implications for Motions



Richards' Molecular and Accessible Surfaces



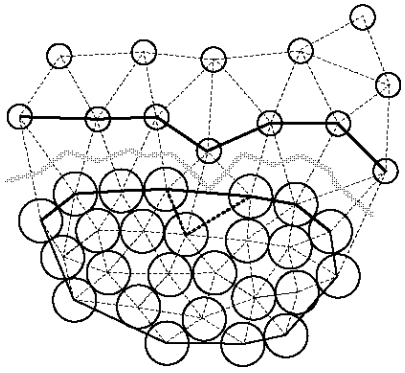
Probe Radius	Part of Probe Sphere	Type of Surface
0	Center (or Tangent)	Van der Waals Surface (vdWS)
1.4 Å	Center	Solvent Accessible Surface (SAS)
""	Tangent (1 atom)	Contact Surface (CS, from parts of atoms)
""	Tangent (2 or 3 atoms)	Reentrant Surface (RS, from parts of Probe)
""	Tangent (1,2, or 3 atoms)	Molecular Surface (MS = CS + RS)
10 Å	Center	A Ligand or Reagent Accessible Surface
∞	Tangent	Minimum limit of MS (related to convex hull)
""	Center	Undefined

Core

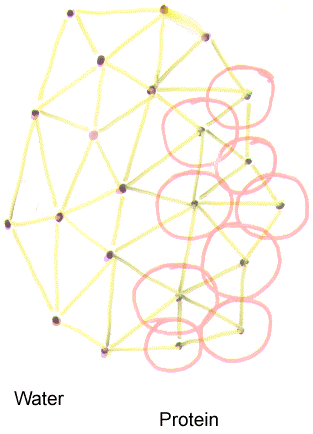
Packing defines the "Correct Definition" of the Protein Surface

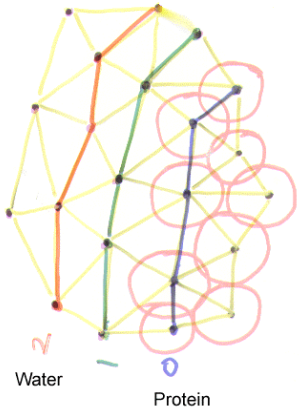
- Voronoi polyhedra are the *Natural* way to study packing!
- How reasonable is a geometric definition of the surface in light of what we know about packing
- The relationship between
  - ◊ accessible surface
  - ◊ molecular surface
  - ◊ Delauney Triangulation (Convex Hull)
  - ◊ polyhedra faces
  - ◊ hydration surface

Surface and Volume Definitions Linked

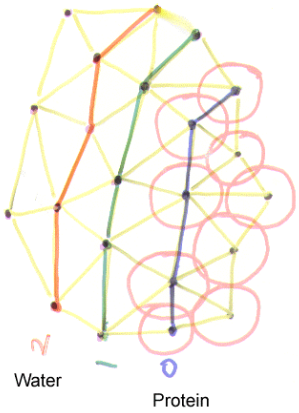


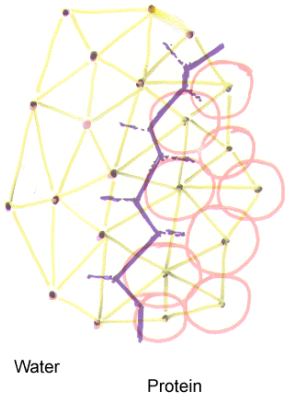
Defining Surfaces from Packing: Convex Hull and Layers of Waters



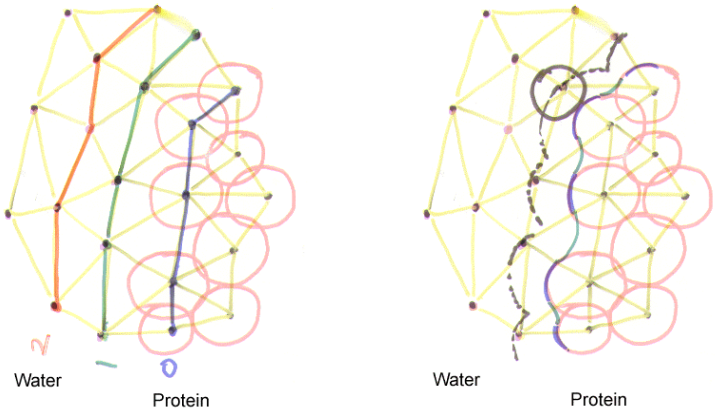


Defining a Surface from the Faces of Voronoi Polyhedra

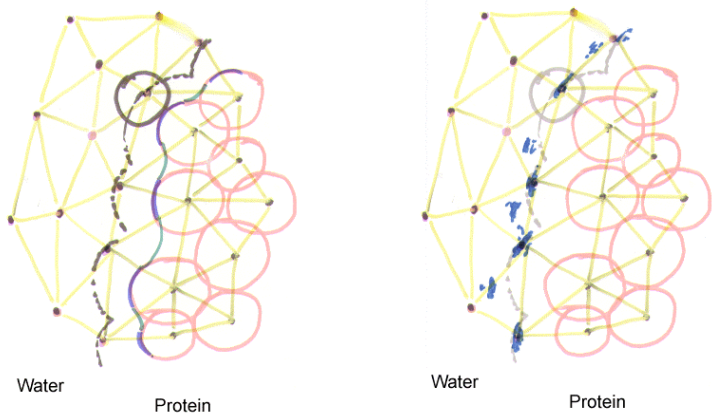




Accessible Surface as a Time-averaged Water Layer



The Hydration Surface: Trying to Model Real Water



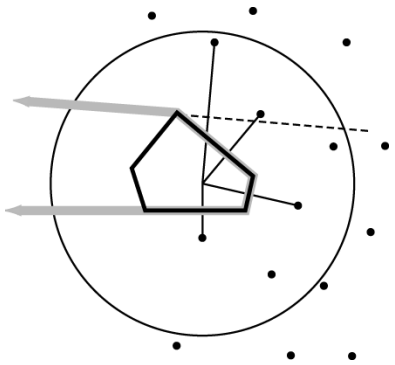
Surfaces
(slides 1-10, 20-40 from website)

These are detailed slides on how to do Voronoi construction.

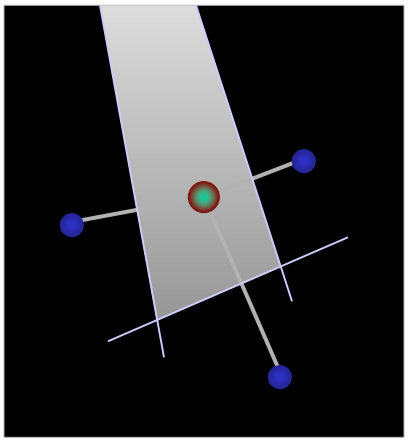
Go to http://bioinfo.mbb.yale.edu/geometry

and follow links to “HyperTalk” tutorial on surfaces and volumes

Problem of Protein Surface for Voronoi Construction



Problems with Voronoi Volumes



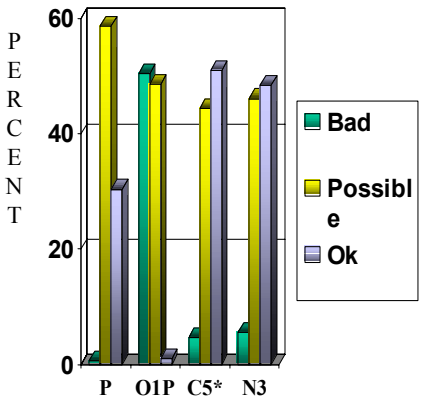
Courtesy of N Voss

- When the Atoms are located on the Surface the Voronoi Volume fails

Ways to Correct for Unclosed Volumes

- Obtain Asymmetric Crystal Packing
- Adjust for Atom Size

Quality of Measurement

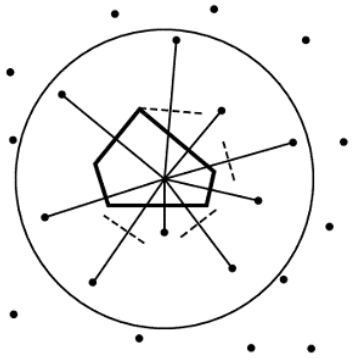


- Determine Quality By Surface Area
- If Surface Falls Below Certain Value Classified as “Ok”

Courtesy of N Voss

Voronoi Volumes

- 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



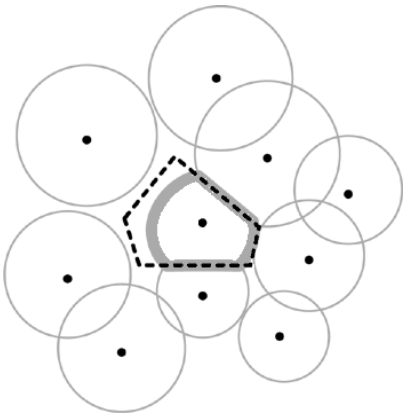
Voronoi Volumes, the Natural Way to Measure Packing

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)

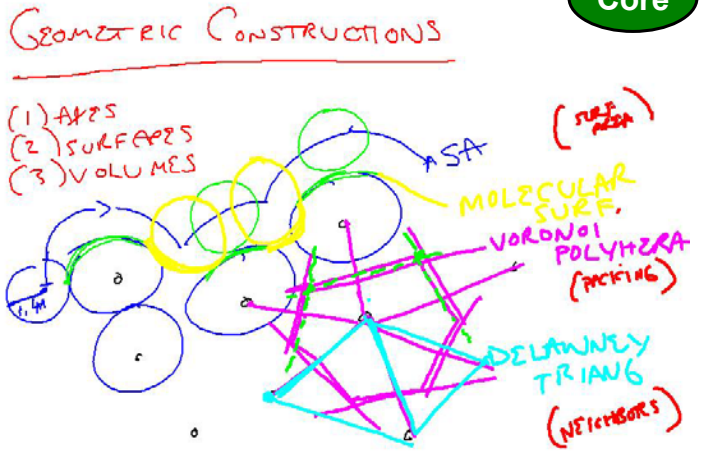


Calclating Volumes with Voronoi polyhedra

- 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.

Summary of Geometric Constructions

Core



End of class 2002,11.13

(Bioinfo-10)

[Next is discussion of expression]