Applications of Voronoi Polyhedra:

Analysis of Molecular Dynamics Simulations

Jerry Tsai, Mark Gerstein, & Michael Levitt

Stanford University

Department of Structural Biology

http://hyper.stanford.edu/~mbg/SurfVolTalk

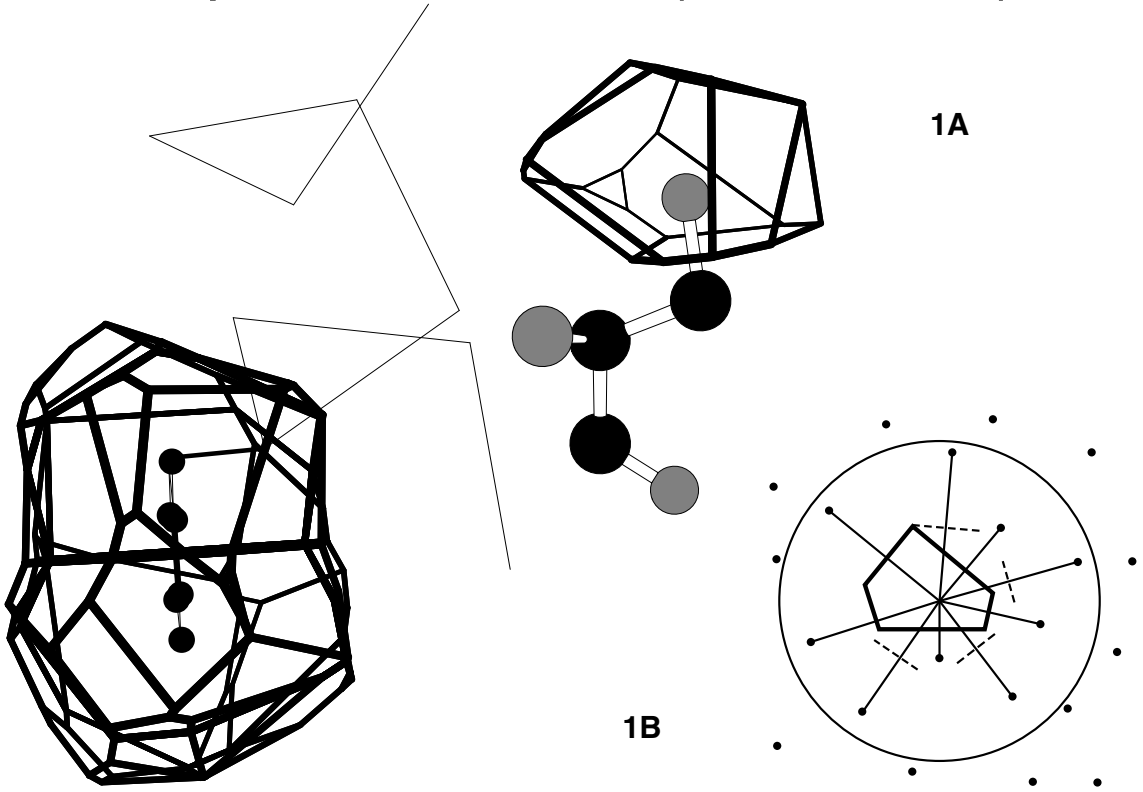
Molecular Dynamics Simulation

- Create a detailed view of molecular interactions between individual atoms and/or molecules, i.e. protein and surrounding water solution
- Program: ENCAD,
  ENergy Calculation And Dynamics
  - 1 ns requires 5.5 CPU days for a system of 5pti (58 aa) in 2557 waters
  - Reproduces structural and dynamic qualities of pure water solutions.
  - Stable simulation of proteins and helices.
- Analysis and Systems Studied
  - Surface packing in a protein simulation trypsin inhibitor in solution (see above)
  - Aggregation in small solute simulations
    a) 32 ureas and 183 waters
    b) 32 isobutylenes and 175 waters
    c) 1-48 methanes and 214-147 waters, respectively

5pti in Solution

Voronoi Polyhedra

- Each atom is surrounded by a single convex polyhedron and allocated the space within it
  - complete allocation (no cavities)



- History
  - Developed by Voronoi in 1908
  - Bernal & Finney in 1967 used w/ liquid sim.
  - Richards in 1974 applied to proteins

Uses of Polyhedra in Simulation

1 Explicit determination of protein volumes and packing densities

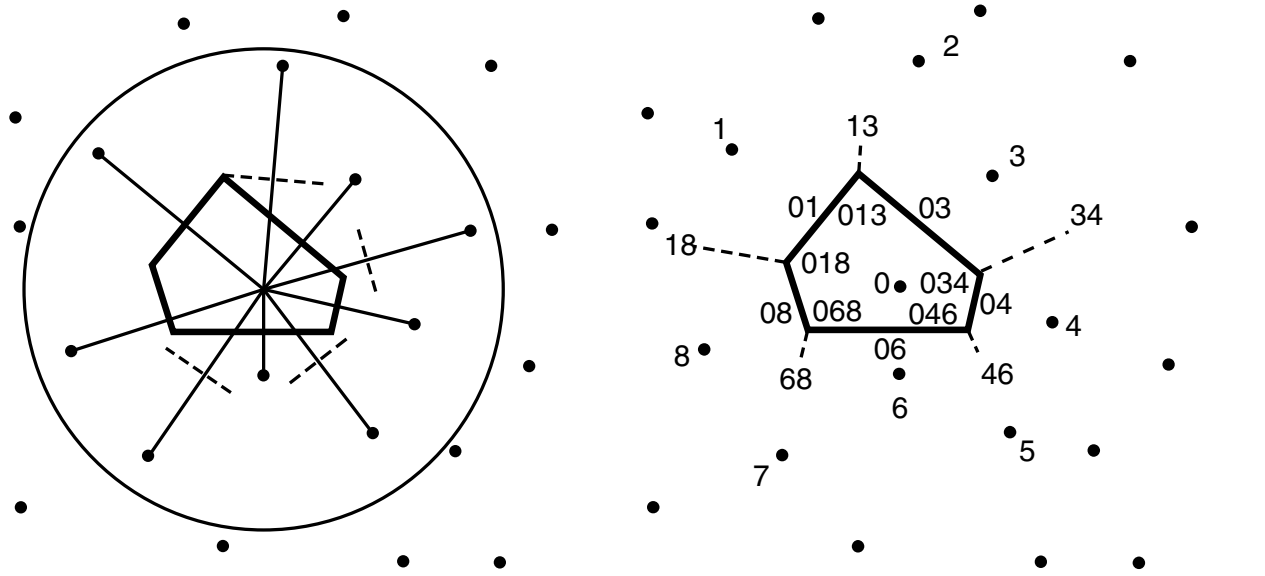
- Packing efficiency = $V_{VDW} / V_{Voronoi}$
- Compressibility

2 Precise calculation of surface area and contacts

- Area of a face “weights” extent of contact
- Volume as weighted average of contact distances

- Also common in CS
  - “Nearest-neighbor” problems (find the neighbor of a query point)
  - Largest empty circle in a collection of points
  - Dual of a Voronoi diagram is Delaunay triangulation (Convex Hull)

Basic Voronoi



1 Find Vertices

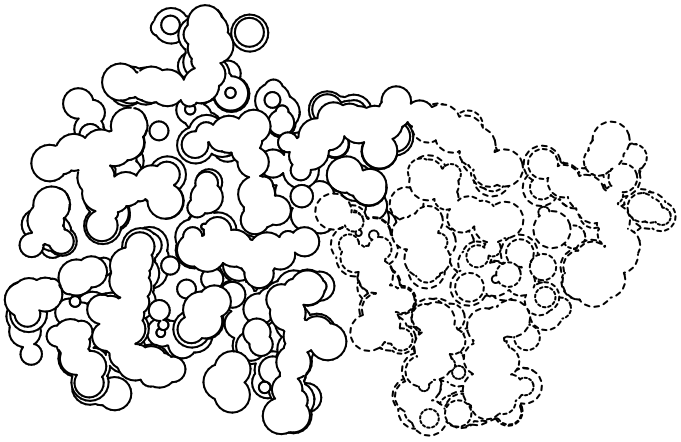
- Point in space assigned to closest atom
- In 3D, points equidistant from 2 atoms form a face; 4 atoms, a vertex
- Solve for vertex using 4 atoms and equation of sphere. (Make sure vertex is closer to these 4 than any others)

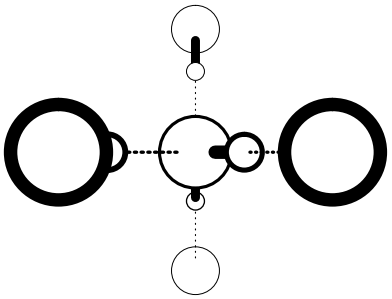
2 Traverse Vertices $\rightarrow$ Calculate Volume

- Label Vertices: those with 2 atoms in common are a face; those with 3, connected by edge
- Systematically traverse: face areas $\rightarrow$ pyramid volumes $\rightarrow$ polyhedra volumes

Why Packing at the Protein Surface is Interesting

- Surface is where function happens
- Helps to understand why ligands bind
  - Elements of molecular recognition
  - Docking & rational drug design
- Surface is where two dissimilar substances meet





  - Protein: close-packing apolar atoms in the core
  - Water: “Open,” tetrahedral structure dictated by H-bonds
  - Similar contrast to water v. simple liquid such as argon

Water vs. Argon

20 Standard Core Volumes

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

- These values imply an exceptionally tight packing in the protein core
  - Tighter than in organic crystals

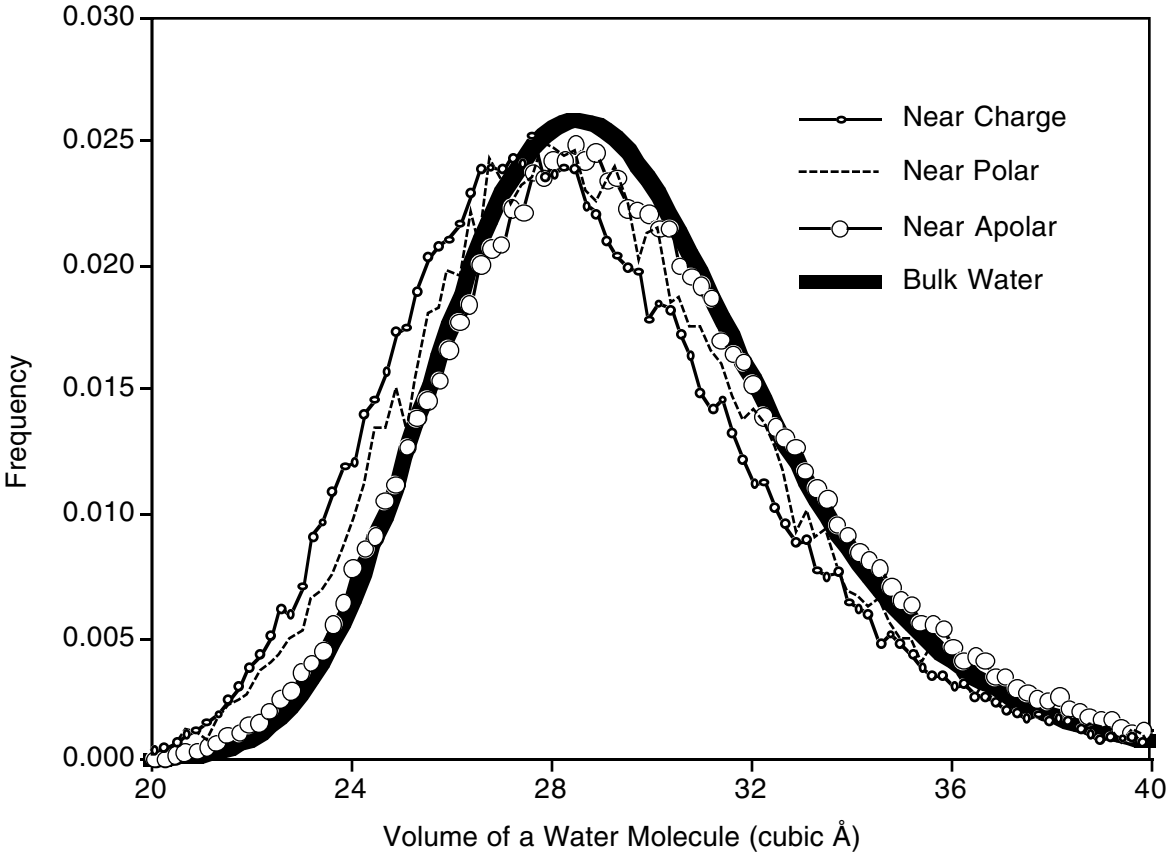
Trypsin Inhibitor Results: Protein

- Most protein atoms are larger on the surface than in the core, by ~6% on average
  - Except charged atoms, which decrease in size, presumably due to electroconstriction

Atom Type		Charge	Buried (a) Num.	Volume	Surface (a) Num.	Volume	Difference Volume (b)
C			33	9.7	19	9.9	+ 2.8%
CA			20	14.7	32	15.1	+ 2.7%
O	(c)		18	16.8	39	18.1	+ 8.0%
N			24	14.2	24	15.3	+ 7.3%
>C=			11	10.7	15	10.8	+ 0.6%
—CH=			16	22.8	20	24.0	+ 5.6%
>CH-			3	14.9	5	15.8	+ 6.5%
—CH2—			11	25.5	67	27.4	+ 7.5%
—CH3			0	—	20	42.7	
=O	(c)		2	17.5	2	18.1	+ 3.5%
—O	(c)	-.9	1	13.7	9	13.2	-3.6%
—OH			2	17.3	7	23.5	+ 35.8%
—NH2		+.4	2	29.7	14	28.0	-5.7%
—NH3		+1.	0	—	4	20.6	
—S—			4	32.2	2	34.5	+ 7.1%

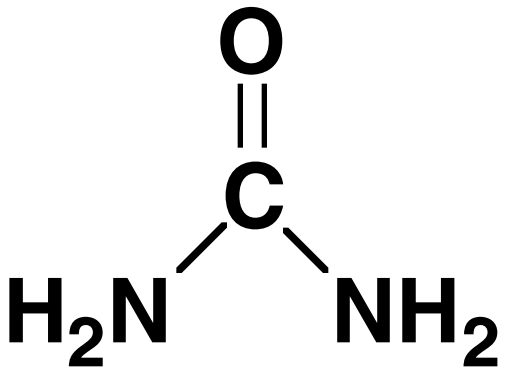
Trypsin Inhibitor Results: Water

- In aggregate, water molecules in contact with protein are smaller in size, counter-balancing the increase in protein size
- Water molecules near the protein surface are marginally larger near apolar atoms than those near polar or charged atoms



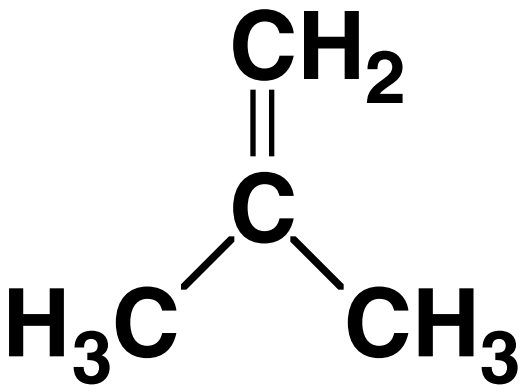
Simulations of Binary Solutions

- Characterization of denaturant for use in further simulation studies with proteins
  - A well characterized denaturant:



Urea

  - For comparison, a molecule with the same planar,Y-shape structure as urea but non polar:



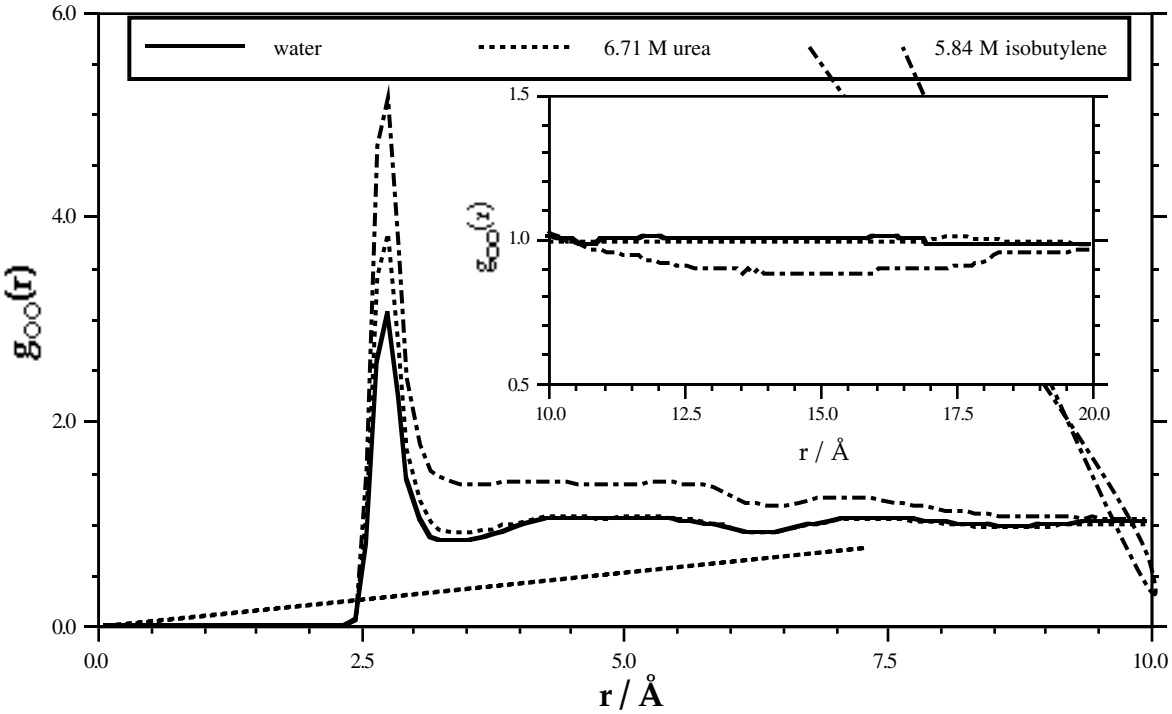
Isobutylene

- Also, simulations of methane in solution to study hydrophobic aggregation

Urea in Solution

Isobutylene in Solution

Water O..O RDF

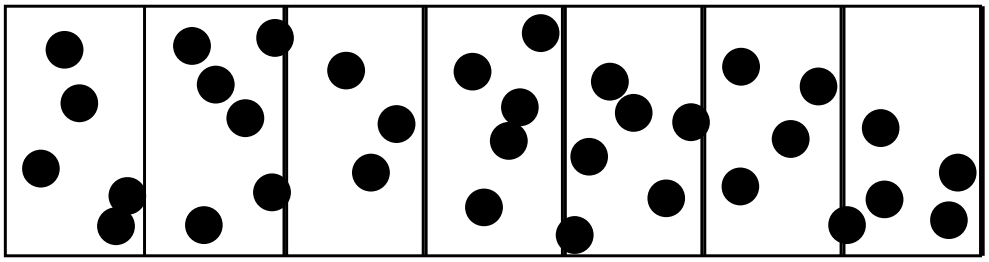


- Results suggested from RDF
  - isobutylene disrupts water structure more than urea
  - decrease in water density past 10 Å for the isobutylene simulation indicates solute aggregation

- However, these conclusions are less intuitive and do not provide a descriptive illustration of what is occuring in solution

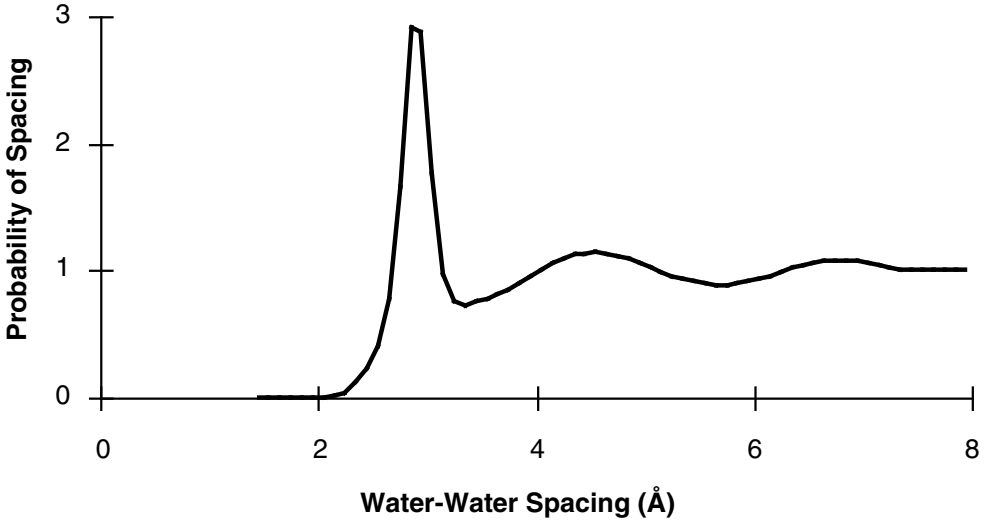
Number Density

= Number of waters per unit volume averaged over simulation



5 6 3 5 6 4 5

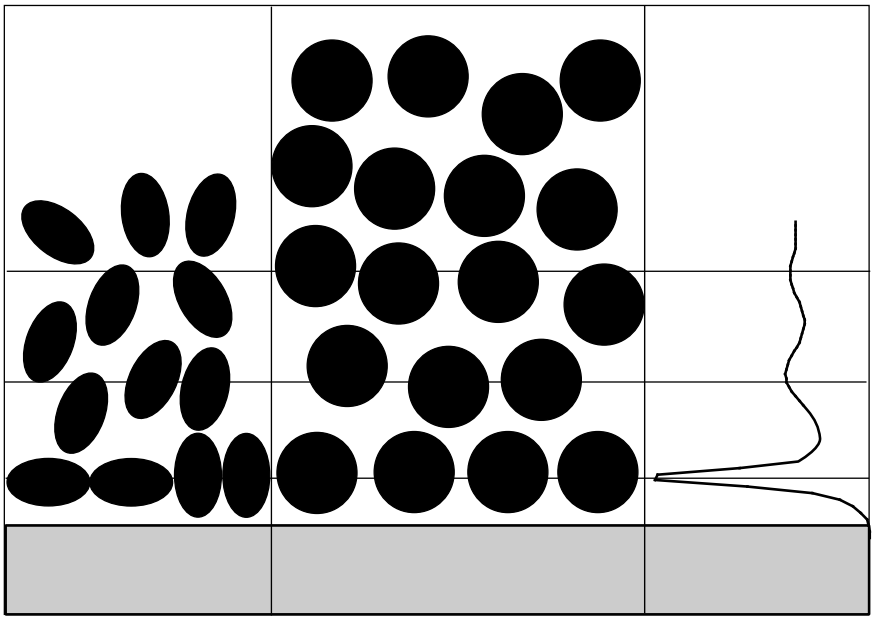
- Advantages: Intuitive, Relates to scattering expts
- Spatially average in different ways
  - Over all directions gives 1D RDF



  - Project along axis of helix gives 2D distribution

Problems with Number Density

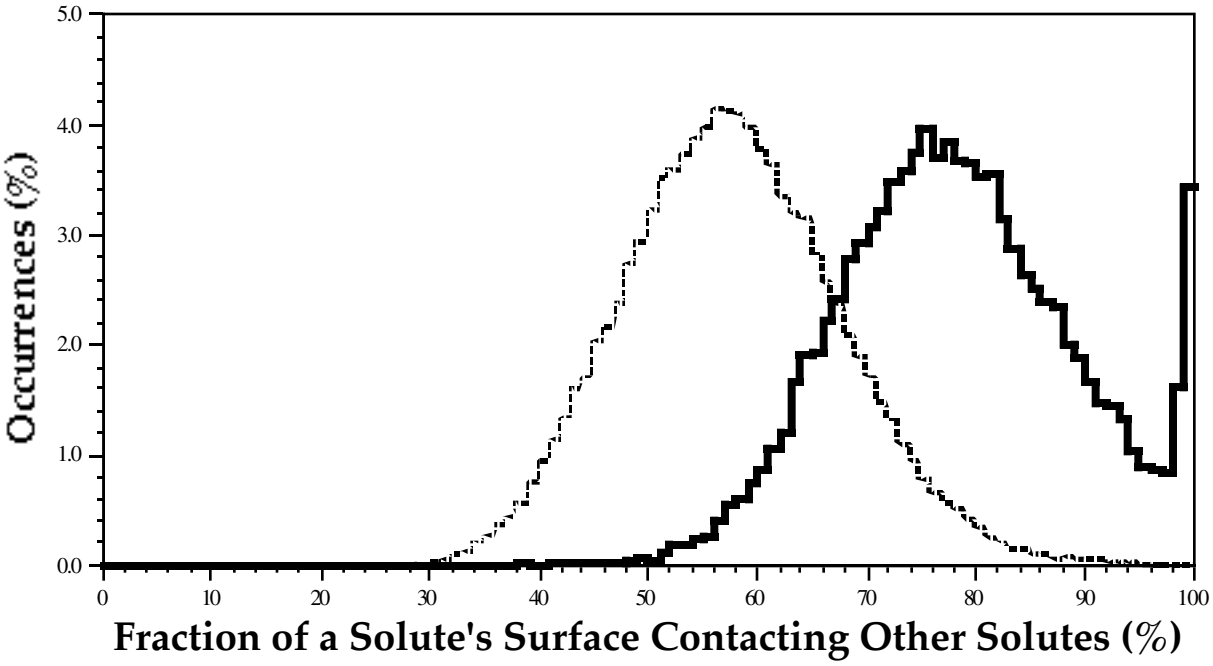
- Number densities measure spatial correlations, not packing
  - 1 D average of 3 D quality
  - Complicated by asymmetric molecules
  - How solutes aggregate is property of instantaneous structure — not average



- Alternative: Measure Aggregation from Surface Area Calculations using Voronoi Polyhedra

Surface Area Calculations

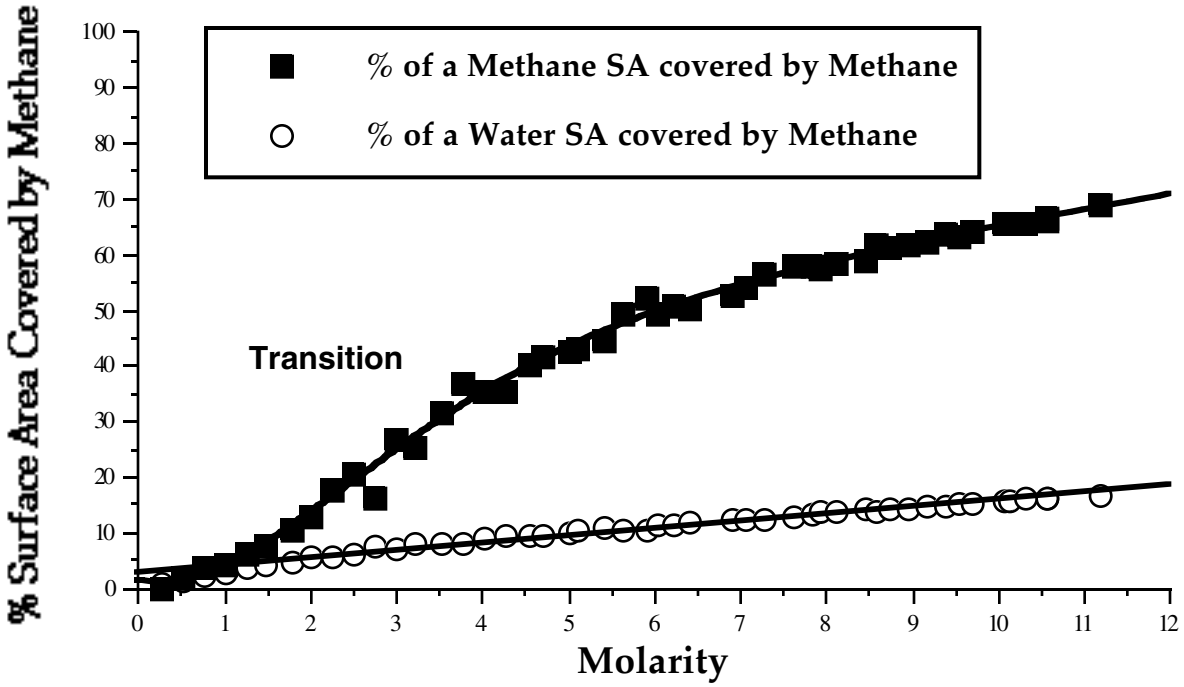
- Find the percent surface area (%SA) of one solute molecule that touches other solute molecules
  - high values indicate a lot of solute-solute interaction: aggregation
  - 100% coverage shows phase separation
- Distribution of %SA in solute simulations



  - urea (broken line) always contacting solution
  - isobutylene (solid line) displays two species: completely sequestered & at water interface

Phase Transition in Methane Simulations

- Simulations of up to 1 to 48 methanes with 216 to 147 waters, respectively



- Average %SA of methane (filled squares) or water (open circles) covered by methane molecules
  - Linear increase of water's %SA
  - Sigmoidal curve of methane %SA suggests a phase transition due to burial of methane away from water