Protein Models

Geometric/Kinematic Models

- A geometric model just take into account how atoms are statically displaced within a molecule to form a geometric shape. Different models (shapes) are aimed at showing different features.
- Kinematics studies possible movements in a system, regardless of the causing forces.

Atomistic Model (ordinary…)

- In the atomistic model, the molecule is described by the coordinates of all the composing atoms.
- With p atoms,
  the model is described by 3p parameters
- Vibrational degrees of freedom: 3p-6
  (3 coord. for the center of mass, 3 for possible rotations around it)
  3p-5 for linear molecules
- No information about bonds is explicitly present: it doesn't account for the *kinematic structure*

Linkage Model

- Kinematics is defined through *internal coordinates*:
  - Bond lengths
  - Bond angles
  - Torsion angles around bonds
- Internal coordinates are also commonly known as "curvilinear coordinates" or "valence coordinates"

Rotatable Bonds Model

- Each protein conformation on n residues will be specified as a sequence of dihedral angles

Cα Trace

- Aka "Virtual Bonds Model"
- The whole protein is modeled by the sequence of alpha carbons
- The Cα-Cα distance is fixed at 3.8 Å
- The model is described by a sequence (N-3) of angle pairs (θ,τ)
- θ in [0,π] is the angle between three consecutive Cα
- τ in [- π, π] is the dihedral between four consecutive Cα

Further Abstraction: Coarse Grain & Toy Models

- (revise!)
- United-atom model: non-polar H atoms are incorporated into the heavy atoms to which they are bonded
- Lollipop model: the side-chains are approximated as single spheres with varying radii
- Bead model: Each residue is modeled as a single sphere

3D Moving: Rigid-body Transform

Homogeneous Coordinates

- *Affinity*: linear transformation A + translation **b**
- Homogeneous coordinates allow all affine transformations to be dealt with by a matrix operation.
- In our setting, linear transformations of interest are *rotations*
- **y** = A**x**+**b**
  - A : rotation matrix
  - **b** : translation vector
- In matrix notation:
  y = T**x**

$$\begin{pmatrix} y_i \\ y_j \\ y_k \\ 1 \end{pmatrix} = \begin{pmatrix} A_{11} & A_{12} & A_{13} & b_i \\ A_{21} & A_{22} & A_{23} & b_j \\ A_{31} & A_{32} & A_{33} & b_k \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_i \\ x_j \\ x_k \\ 1 \end{pmatrix}$$

- Properties of rotation matrices:

  $A^T = A^{-1}$

  $\det(A) = +1$ (axes' handness is preserved)
- Combination of two transformations T_1 and T_2:
  $T' = T_1 T_2$

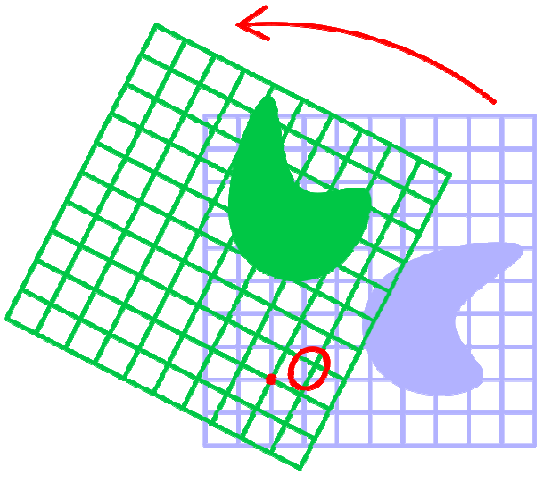
2D Rotation Matrices

- Because of constraints (properties) defined before, 2x2 rotation matrices have the following form:

$$\mathbf{A} = \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \quad \text{with } a^2+b^2=1$$

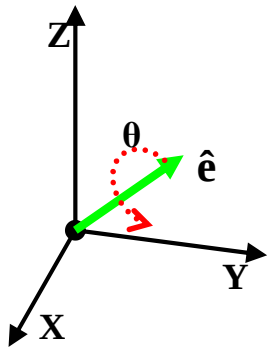
- Considering a counterclockwise rotation of θ around the origin,

$$\mathbf{A} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$



3D Rotation Matrices

- In 3D a rotation can be specified by:
  - The direction of rotation axis (Euler axis), via the unit vector $\hat{\mathbf{e}} = (\hat{e}_i\ \hat{e}_j\ \hat{e}_k)^T$
  - The counterclockwise rotation angle θ



- If $\hat{\mathbf{e}} \equiv \mathbf{k}$, then

$$\mathbf{A}_{\mathbf{k},\theta} = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

- If $\hat{\mathbf{e}} \equiv \mathbf{j}$, then

$$\mathbf{A}_{\mathbf{j},\theta} = \begin{bmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{bmatrix}$$

- If $\hat{\mathbf{e}} \equiv \mathbf{i}$, then

$$\mathbf{A}_{\mathbf{j},\theta} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix}$$

General case:

$$\mathbf{A}_{\hat{\mathbf{e}},\theta} = \begin{bmatrix} c\theta + (1-c\theta)\,\hat{e}_i^2 & (1-c\theta)\,\hat{e}_i\hat{e}_j - s\theta\,\hat{e}_k & (1-c\theta)\,\hat{e}_i\hat{e}_k + s\theta\,\hat{e}_j \\ (1-c\theta)\,\hat{e}_i\hat{e}_j + s\theta\,\hat{e}_k & c\theta + (1-c\theta)\,\hat{e}_j^2 & (1-c\theta)\,\hat{e}_j\hat{e}_k - s\theta\,\hat{e}_i \\ (1-c\theta)\,\hat{e}_i\hat{e}_j - s\theta\,\hat{e}_j & (1-c\theta)\,\hat{e}_k\hat{e}_j + s\theta\,\hat{e}_i & c\theta + (1-c\theta)\,\hat{e}_k^2 \end{bmatrix}$$

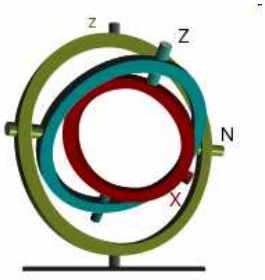
Eulerian Angles

- *Euler's rotation theorem*: Any generic rotation A can be obtained by the successive application of (some) 3 basic rotations:

  $A = A_1 A_2 A_3$

- A possible choice is the following:

$$\mathbf{A}_{\alpha,\beta,\gamma} = \begin{bmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\beta & -\sin\beta \\ 0 & \sin\beta & \cos\beta \end{bmatrix} \begin{bmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

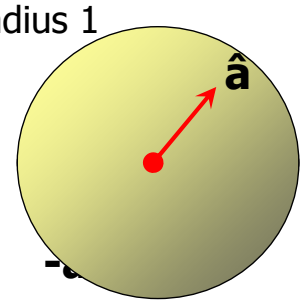


  - A_1 : around z-axis of the original reference frame
  - A_2 : around an *intermediate* x-axis (known as *line of nodes*)
  - A_3 : around z-axis of the final reference frame

- Rotation A can thus be specified by the *Eulerian angles* α, β, and γ. Given A, angles α, β, and γ can be usually be determined: in some singular cases, ambiguities are possible.

Quaternions

- A convenient representation of rotations: unit quaternions.
- A rotation $A_{\hat{\mathbf{e}},\theta}$ of angle θ about $\hat{\mathbf{e}}$ can be described by the 4-dimentional *unit* vector $\hat{\mathbf{a}} = (a_1\ a_2\ a_3\ a_4)^T$ with
  - $a_1 = \cos(\theta/2)$ known as ***scalar term***
  - $a_2 = \hat{e}_i \sin(\theta/2)$ $a_3 = \hat{e}_j \sin(\theta/2)$ $a_4 = \hat{e}_k \sin(\theta/2)$ collectively known as ***vector term***
  - Separating scalar and vector terms, $\hat{\mathbf{a}}$ is indicated also as $\hat{\mathbf{a}} = \cos(\theta/2) + \hat{\mathbf{e}} \sin(\theta/2)$
- Space of rotations → 4-dimentional sphere of radius 1
- Rotation $A_{\hat{\mathbf{e}},\theta}$ can be described both by $\hat{\mathbf{a}} = (a_1\ a_2\ a_3\ a_4)^T$ and by the opposite quaternion $\hat{\mathbf{a}}' = (-a_1\ -a_2\ -a_3\ -a_4)^T$
- A point (vector) $\mathbf{c}$ in 3D space → quaternion $\hat{\mathbf{c}} = (0\ c_i\ c_j\ c_k)^T$



Operations on Quaternions

Given two quaternions $p = p_s + \mathbf{p_v}$ and $q = q_s + \mathbf{q_v}$

- **Norm** of p is $\|p\| = (p_1^2+p_2^2+p_3^2+p_4^2)^{1/2} = (p_s^2+|\mathbf{p_v}|^2)^{1/2}$
- **Addition** – trivial, pairwise on components
- **Multiplication** – not commutative, based on the assumed rule
  $\mathbf{st} = \mathbf{s}\wedge\mathbf{t} - \mathbf{s}\cdot\mathbf{t}$

  $pq = (p_s + \mathbf{p_v})(q_s + \mathbf{q_v}) = (p_s q_s - \mathbf{p_v}\cdot\mathbf{q_v}) + (p_s\mathbf{q_v} + \mathbf{p_v}\wedge\mathbf{q_v} + q_s\mathbf{p_v})$
- **Conjugate** of p is $p^* = p_s - \mathbf{p_v}$
- The (left and right) multiplicative **inverse** (reciprocal):
  $p^{-1} = p^*/\|p\|^2 = (p_s - \mathbf{p_v}) / (p_s^2 + |\mathbf{p_v}|^2)$ (for unit quat., $p^{-1} = p^*$)

Using Quaternions

- To translate a vector **x** to the new position **x'** by the vector **b** :
  $(0+\mathbf{x}) + (0+\mathbf{b}) \rightarrow (0+\mathbf{x'})$ with $\mathbf{x'} = \mathbf{x}+\mathbf{b}$ (trivial)

- To rotate a vector **x** to the new position **x'** by the quaternion **â**:
  $\hat{\mathbf{a}}\ (0+\mathbf{x})\ \hat{\mathbf{a}}^{-1} \rightarrow (0+\mathbf{x'})$ this operation is named *conjugation*

- How to relate rotation matrix A
  and the corresponding quaternion $\hat{\mathbf{a}} = (a_1\ a_2\ a_3\ a_4)^T$?

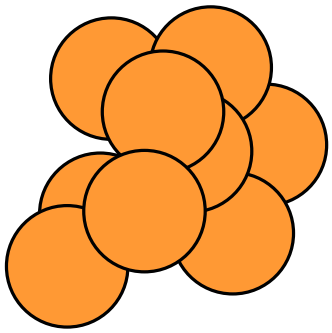
$$\mathbf{A}_{\hat{\mathbf{e}},\theta} = \begin{pmatrix} 2(a_1^2+a_4^2)-1 & 2(a_1a_2+a_3a_4) & 2(a_1a_3+a_2a_4) \\ 2(a_1a_2+a_3a_4) & 2(a_2^2+a_4^2)-1 & 2(a_2a_3+a_1a_4) \\ 2(a_1a_3+a_2a_4) & 2(a_2a_3+a_1a_4) & 2(a_3^2+a_4^2)-1 \end{pmatrix}$$

- Although the curvilinear internal coordinates can give a good description of the molecular potential, it is difficult to express the kinetic energy of nuclear vibrations in these coordinates.

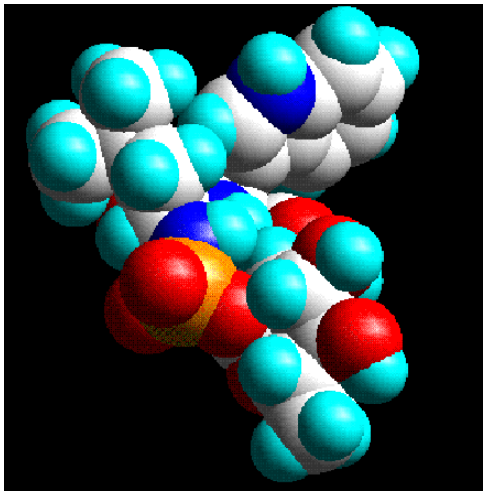
Surface Characterization

VdW Surface

- Hard-sphere model (van der Waals radii)
  - Van der Waals surface



H	C	N	O	F	P	S	Cl
1.2	1.7	1.5	1.4	1.35	1.9	1.85	1.8

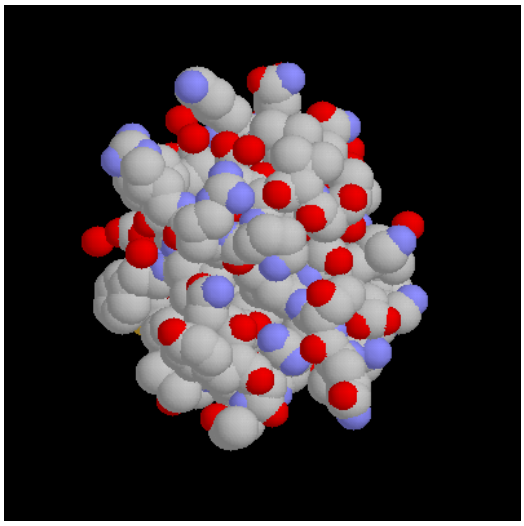


Surface and interior of globular proteins

- solvent accessible surface
- molecular surface
- residue fractional accessibility
- pockets and cavities
- "hydrophobic core"
- ordered waters in protein structures

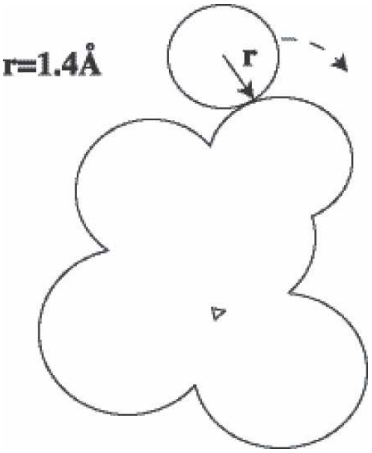
"Accessible Surface"

represent atoms as spheres w/appropriate radii and eliminate overlapping parts...



mathematically roll a sphere all around that surface...

the sphere's center traces out a surface as it rolls...

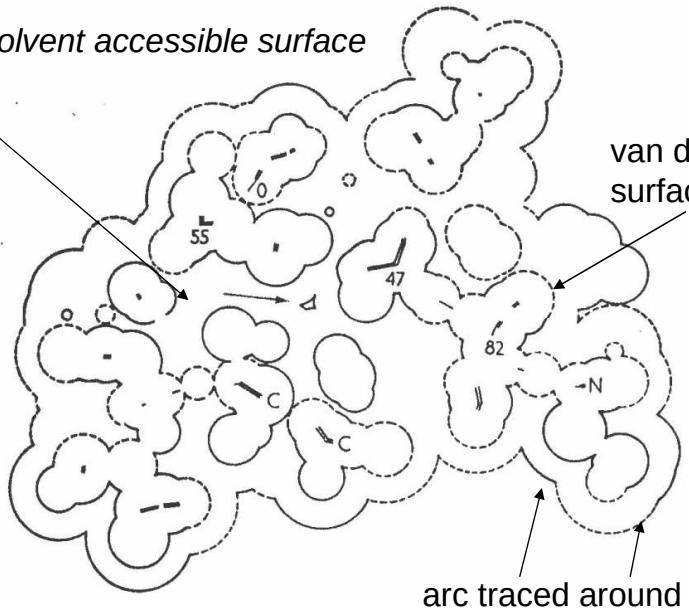


Lee & Richards, 1971
Shrake & Rupley, 1973

Now look at a cross-section (slice) of a protein structure:
Inner surfaces here are van der Waals. Outer surface is that *traced out by the center of the sphere* as it rolls around the van der Waals' surface. If any part of the arc around a given atom is traced out, that atom is accessible to solvent. The *solvent accessible surface* of the atom is defined as the sum the arcs traced around an atom.

there's not much solvent accessible surface in the middle

solvent accessible surface

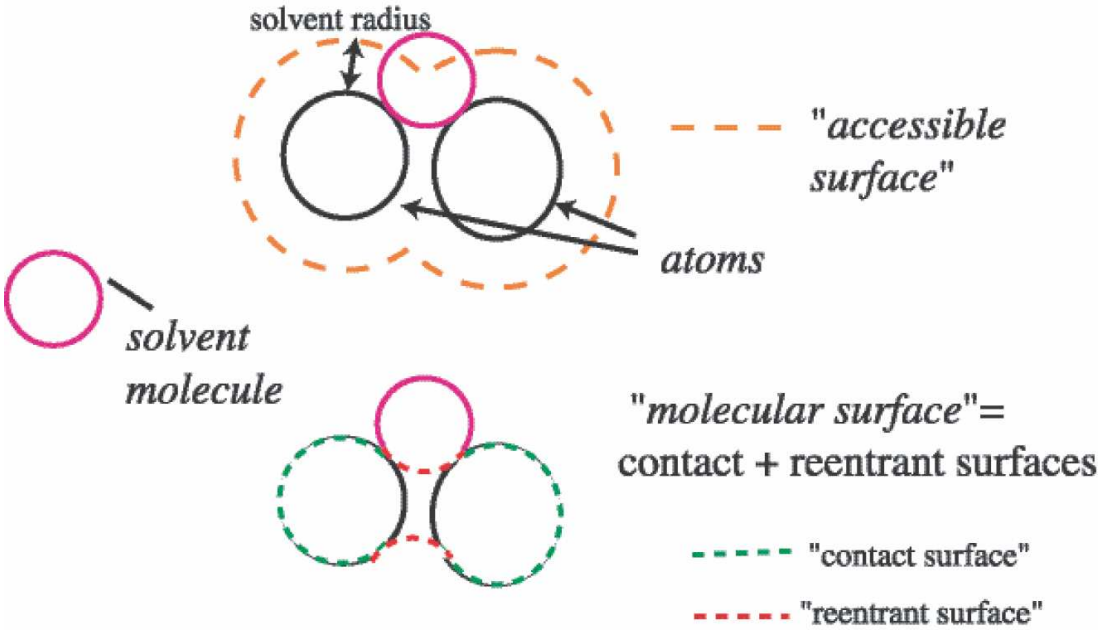


van der Waals surface

from Lee & Richards, 1971

arc traced around atom

"Accessible surface"/"Molecular surface"

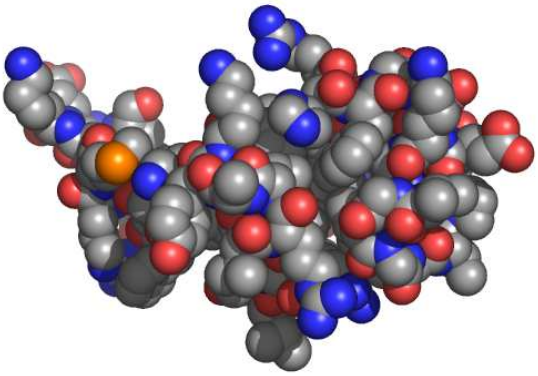


note: these are *alternative* ways of representing the same reality: the surface which is essentially in contact with solvent

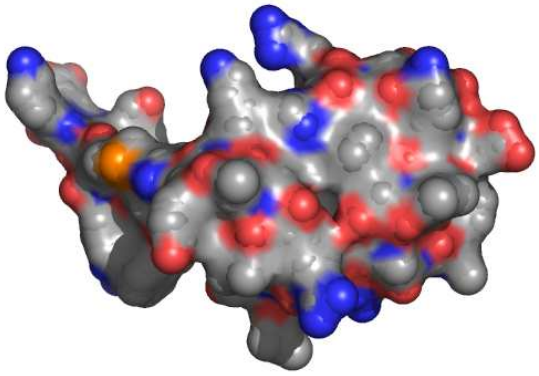
- ***molecular*** and ***accessible*** surfaces are both useful representations, but molecular surface is more closely related to the actual atomic surfaces. This makes it somewhat better for visualizing the texture of the outer surface, as well as for assessing the shape and volume of any internal cavities.

- you will hear the term ***Connolly surface*** used often, after Michael Connolly. A Connolly surface is a particular way of calculating the molecular surface. The accessible surface is also occasionally called the ***Richards surface***, after Fred Richards.

Molecular surface of proteins



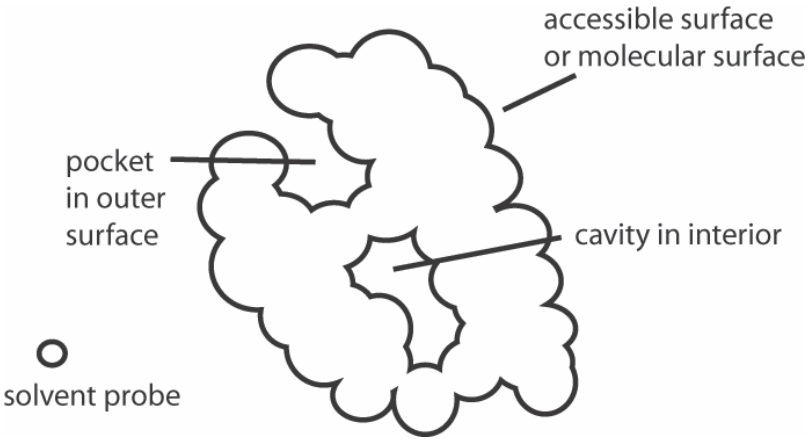
depiction of heavy atoms (O, N,C, S) in a protein as van der Waals spheres



depiction of the corresponding "molecular surface"--volume contained by this surface is vdW volume plus "interstitial volume"--spaces in between

The irregular surface of proteins: pockets and cavities

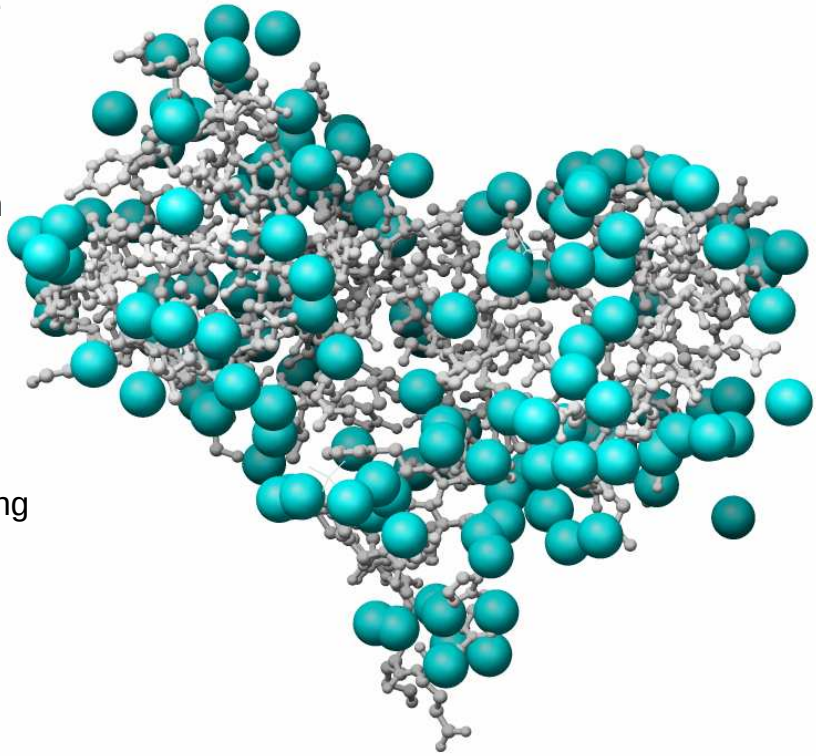
- a **pocket** is an empty concavity on a protein surface which is accessible to solvent from the outside.
- a **cavity** or **void** in a protein is a pocket which has no opening to the outside. It is an interior empty space inside the protein.



Pockets and cavities can be critical features of proteins in terms of their binding behavior, and identifying them is usually a first step in structure-based ligand design etc.

The outer surface: water in protein structures

Structures of water-soluble proteins determined at reasonably high resolution will be decorated on their outer surfaces with water molecules (cyan balls) with relatively well-defined positions, and waters may also occur internally



Water is not just surrounding the protein--it is *interacting* with it

Water interacts with protein surfaces

Most waters visible in crystal structures make hydrogen bonds to each other and/or to the protein, as donor/acceptor/both