Protein Geometry

Measuring
Geometric Properties
of
Proteins

Motivation

- One of the goals of structural bioinformatics is to aid the biochemist in modeling molecular functionality on a computer.
  - Notice the change in experimental setting:
    in vivo ⟹ in vitro ⟹ in silico
- Although reaction mechanisms are essentially the final statement in characterizing molecular interactions, there is often a need to track conformational changes and other geometric aspects of the molecules.

Introduction

- The geometry of molecules deals with computations related to bond length and inter-atomic distances in general, bond angles, and dihedral angles.
  - More complicated calculations deal with the construction of molecular surfaces and quantities such as charge densities.
  - These are calculations related to a static molecule.
- In more dynamic setting we may attempt to evaluate these quantities as they change with time due to flexibility of the molecule.
  - Modeling the flexibility is itself a big challenge.

Given Atomic Coordinates

- We start with the assumption that we know the positions of atoms in 3D space.
- We then develop the formulae that give us various other measurements.
  - bond length, bond angles, dihedral angles.
  - In some applications you are given the inter-atomic distances and need to derive the coordinates.

Distance Between Atoms

- If the position of atoms in 3-space is given by (x, y, z) coordinates then we can use the standard Pythagorean calculation of distance.
  - If atom a has coordinates $(a_x, a_y, a_z)^T$ and atom b has coordinates $(b_x, b_y, b_z)^T$ the distance between a and b is given by:

$$d(a,b) = \sqrt{(a_x - b_x)^2 + (a_y - b_y)^2 + (a_z - b_z)^2}$$

  - This is the same as the norm calculation:

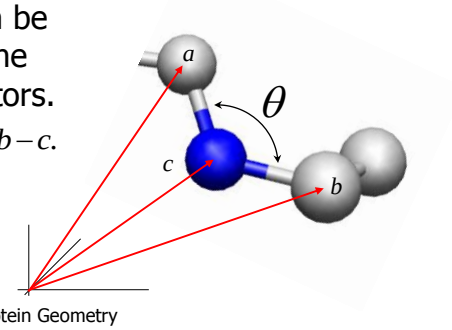
$$\|a - b\| = \sqrt{(a-b)^T (a-b)}.$$

Bond Angle

- Consider two atoms (vectors a and b) bonded to a third atom represented by vector c.
  - We calculate the bond angle at c by first calculating $\cos\theta$.
    - Recall that an inner product of normalized vectors can be viewed as the cosine of the angle between these vectors.
    - In this case: $u = a - c \quad v = b - c.$
    - So:

$$\cos\theta = \frac{(a-c)^T (b-c)}{\|a-c\|\|b-c\|}.$$

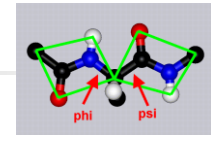
$$\cos\theta = \langle u, v \rangle / (\|u\|\|v\|)$$



Dihedral Angles (1)

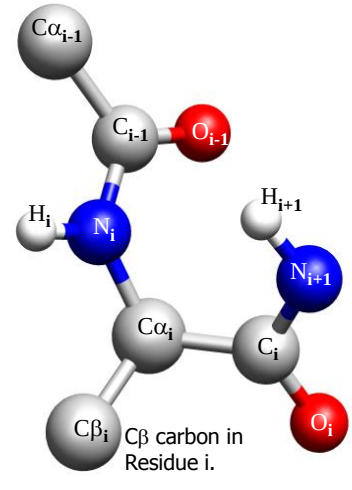
- Consider atoms in the protein backbone:
  - $N_1\ C\alpha_1\ C_1\ N_2\ C\alpha_2\ C_2\ N_3\ C\alpha_3\ C_3\ N_4\ \ldots$
  - Going from residue to residue:
    - Corresponding bond lengths do not change much.
      - For example: $N_i - C\alpha_i$ has approximately the same bond length as $N_j - C\alpha_j$.
    - Similarly, corresponding bond angles tend to be the same.
  - You cannot make this same statement for dihedral angles.

Dihedral Angles (2)



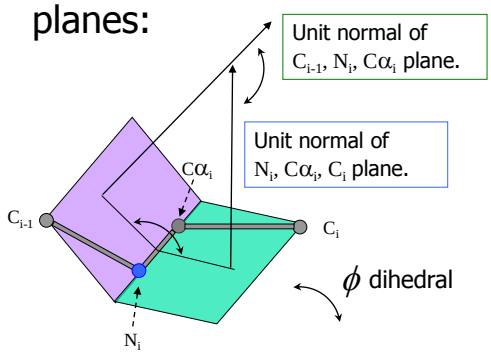
?

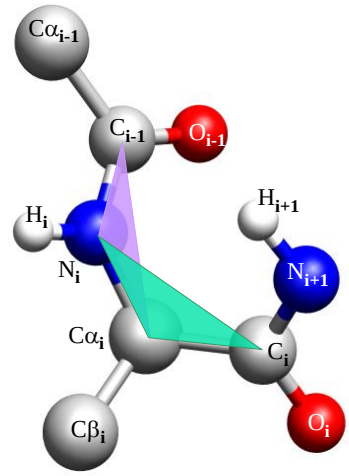
- Dihedral angles are due to a "swivel" action around a single bond.
  - For example, we could keep all bond angles constant while moving N_i relative to N_{i+1} swiveling around the C_i - $C\alpha_i$ bond.
  - So, while a bond angle is determined by 3 atoms a dihedral angle is defined by 4 atoms.



Dihedral Angles (3)

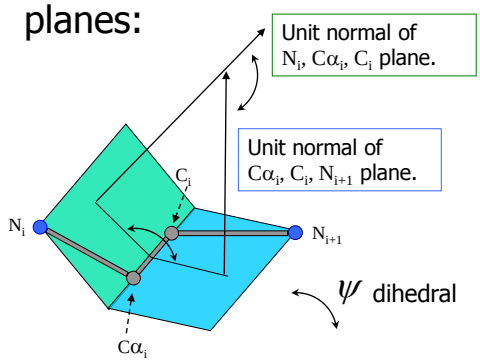
- The alpha carbon PHI dihedral angle is the angle between two planes:

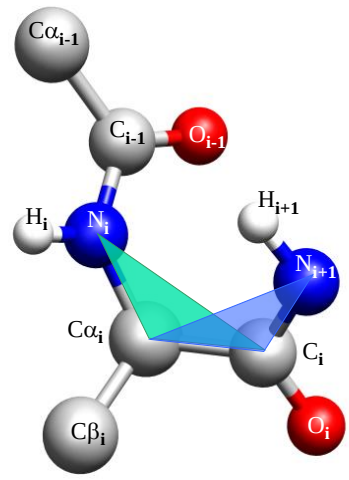




Dihedral Angles (4)

- The alpha carbon PSI dihedral angle is the angle between two planes:





Dihedral Angles (5)

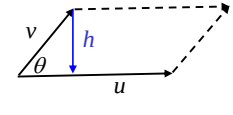
- So, given three points in a plane, we have to know how a unit normal to that plane is calculated.
  - This is done by calculating a *cross product*.
  - Suppose we have two vectors u and v.
  - The cross product is a vector that is perpendicular to both u and v and it has a magnitude that is equal to the area of the parallelogram spanned by the vectors.
    - It can be shown that this is:

$$u \times v = \det \begin{bmatrix} i & j & k \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{bmatrix} = (u_2 v_3 - u_3 v_2) i + (u_3 v_1 - u_1 v_3) j + (u_1 v_2 - u_2 v_1) k.$$

The next slides show this.

Dihedral Angles (6)

- If the magnitude of the normal is the area of the parallelogram defined by u and v then its magnitude is given by the formula: $\|u\|\|h\| = \|u\|\|v\|\sin\theta$ where θ is the angle between u and v.



  - Then we can write:

$$\begin{aligned} \|u \times v\|^2 &= \|u\|^2 \|v\|^2 \sin^2\theta = \|u\|^2 \|v\|^2 (1 - \cos^2\theta) \\ &= \|u\|^2 \|v\|^2 - (\|u\|\|v\|\cos\theta)^2 \\ &= \|u\|^2 \|v\|^2 - (u^{\mathrm{T}} v)^2 \end{aligned}$$

Dihedral Angles (7)

- Continuing:

$$\|u \times v\|^2 = \|u\|^2 \|v\|^2 - (u^{\mathrm{T}} v)^2 = \left(\sum_{k=1}^{3} u_k^2\right)\left(\sum_{k=1}^{3} v_k^2\right) - \left(\sum_{k=1}^{3} u_k v_k\right)^2$$

- We can now apply Lagrange's Identity to see that the last expression is actually:

$$\left(\sum_{i=1}^{2}\sum_{j=i+1}^{3}\left(u_i v_j - u_j v_i\right)^2\right).$$

To prove this you can simply expand both expressions (tedious).

(A proof of the general case can be found at: http://en.wikipedia.org/wiki/Lagrange's_identity)

- Now note that:

$$\left\| \det \begin{bmatrix} i & j & k \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{bmatrix} \right\|^2 = \left\| (u_2 v_3 - u_3 v_2) i + (u_3 v_1 - u_1 v_3) j + (u_1 v_2 - u_2 v_1) k \right\|^2$$

is the same as this last sum.

Dihedral Angles (8)

- Recall: The cross product is a vector that is perpendicular to both u and v and it has a magnitude that is equal to the area of the parallelogram spanned by the vectors.
- So far we have shown that working with this definition we get a vector that has a magnitude given by

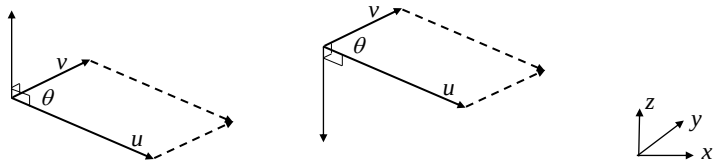
$$\left\| \det \begin{bmatrix} i & j & k \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{bmatrix} \right\|.$$

- Is the vector defined by this determinant perpendicular to both u and v?
  - Yes. If you calculate the inner product of this vector with u or v you get 0. For example:

$$\begin{bmatrix} u_2 v_3 - u_3 v_2 \\ u_3 v_1 - u_1 v_3 \\ u_1 v_2 - u_2 v_1 \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = 0.$$

Dihedral Angles (9)

- Recall: The cross product is a vector that is perpendicular to both u and v and it has a magnitude that is equal to the area of the parallelogram spanned by the vectors.
  - So far our determinant based calculation for the cross product has given the correct magnitude and we have the "perpendicular to both u and v" requirement. BUT which of these is correct? (They both show a vector perpendicular to both u and v).



  - For $u \times v$ we use the "right-hand rule" that is also seen in the usual 3D Euclidean coordinate system: If the fingers of the right hand curl around the $u \times v$ normal going in the direction from u to v then the thumb points in the direction assigned to $u \times v$.
    So the first diagram is for $u \times v$ while the next diagram is for $v \times u$.

Dihedral Angles (10)

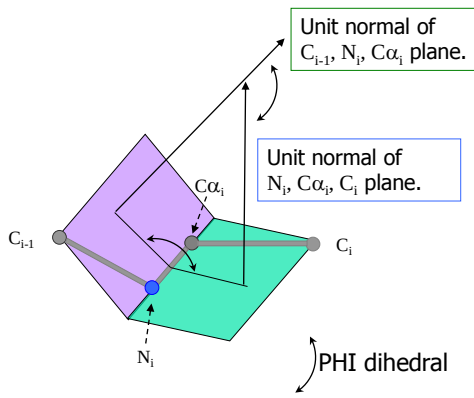
- One last vector algebra issue:
  - The cross product evaluated using the determinant does not necessarily have a unit length.
    - Recall that our dihedral angle calculation requires vectors that are unit normal.
  - In such a situation, we will have to normalize:
    - Given $u \times v$ we calculate:

$$\frac{u \times v}{\|u \times v\|}.$$

Back to proteins ⟶

Dihedral Angles (11)

- Suppose we are calculating the PHI dihedral.
  - Start with the calculation of



For simplicity, we will just use the atom names to label vectors that are their coordinates.

We let $u = N_i - C_{i-1}$

and $v = C\alpha_i - N_i$.

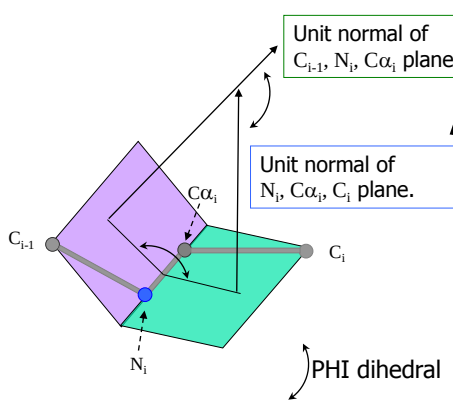
$$u \times v = \det \begin{bmatrix} i & j & k \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{bmatrix} =$$

$$(u_2v_3 - u_3v_2)i + (u_3v_1 - u_1v_3)j + (u_1v_2 - u_2v_1)k.$$

Dividing this by its norm will give the unit normal (call it $n(C_{i-1}, N_i, C\alpha_i)$).

Dihedral Angles (12)

- Calculating the PHI dihedral (continued):
  - Then calculate the other unit normal:



We let $v = C\alpha_i - N_i$

and $w = C_i - C\alpha_i$.

Then use:

$$v \times w = \det \begin{bmatrix} i & j & k \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{bmatrix} =$$

$$(v_2w_3 - v_3w_2)i + (v_3w_1 - v_1w_3)j + (v_1w_2 - v_2w_1)k.$$

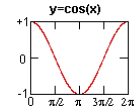
Dividing this by its norm will give the unit normal (call it $n(N_i, C\alpha_i, C_i)$).

Dihedral Angles (13)

- Finally with both unit normals computed we can get the value of PHI for this alpha carbon using the arccos() function:

$$\phi = \tau(C_{i-1}, N_i, C\alpha_i, C_i) = \arccos\left(n(C_{i-1}, N_i, C\alpha_i) \bullet n(N_i, C\alpha_i, C_i)\right).$$

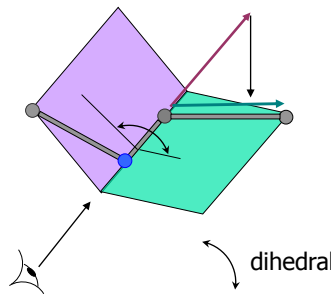
- Sign of the dihedral angle:
  - By convention, a dihedral angle is assumed to be in the range $[-\pi, \pi]$.
  - Since the calculation of arccos() may lead to an angle in the range $[0, \pi]$ we typically have to adjust the sign.



(See next slide)

Dihedral Angles (14)

- To derive the sign of a dihedral angle:
  - We look along the bond lying in the intersection of the two planes (be sure to look in the increasing i direction).



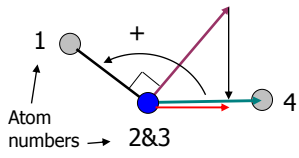
Consider the unit normal of the plane defined by the first three atoms. Compute its inner product with the vector going from the third atom to the last atom.

If this inner product is positive (as in the diagram) then the sign of the dihedral angle is positive, otherwise it is negative.
In the case of PHI this computation is:

$$n(C_{i-1}, N_i, C\alpha_i) \bullet (C_i - C\alpha_i).$$

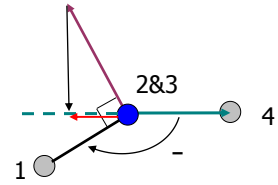
Dihedral Angles (15)

- Positive Dihedrals:



Projection of the unit normal is in the same direction as the last bond.

Negative Dihedrals:



Projection of the unit normal is in the direction opposite to the last bond.

In both cases we are looking down along the central bond in the increasing *i* direction.

Dihedral Angles (16)

- What about PSI dihedrals?
  - Follow the same analysis but this time compute:

$$\psi = \tau\left(N_i, C\alpha_i, C_i, N_{i+1}\right) = \arccos\left(n\left(N_i, C\alpha_i, C_i\right) \bullet n\left(C\alpha_i, C_i, N_{i+1}\right)\right).$$

  - In other words, all the previous computations work to get PSI if you do the following substitutions:

$$C_{i-1} \leftarrow N_i$$
$$N_i \leftarrow C\alpha_i$$
$$C\alpha_i \leftarrow C_i$$
$$C_i \leftarrow N_{i+1}.$$

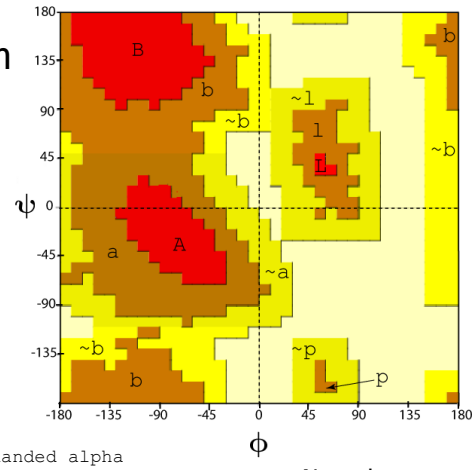
Ramachandran Plots (1)

- It was noted in the early 1960's that steric collisions prohibit certain combinations of the phi and psi angles on either side of a given alpha carbon atom.
  - An observed combination of phi/psi angles is often represented by a single point in a two dimensional plot that has a horizontal axis for phi ranging over values (i.e. -180 to +180 degrees) and a vertical axis for psi ranging over the same values.
  - A collection of many such points can be used to record the observed phi/psi combinations corresponding to a set of alpha carbon atoms.
  - This type of scatter plot is called a Ramachandran plot.

Ramachandran Plots (2)

- This figure is derived from a Ramachandran scatter plot of 121,870 phi/psi combinations using 463 protein structures.

Procheck "background" plot



A	- Core alpha	**L**	- Core left-handed alpha
a	- Allowed alpha	**l**	- Allowed left-handed alpha
~a	- Generous alpha	**~l**	- Generous left-handed alpha
B	- Core beta	**p**	- Allowed epsilon
b	- Allowed beta	**~p**	- Generous epsilon
~b	- Generous beta		

Note the "wrap-around" continuity.

Ramachandran Plots (3)

- Procedure for generating this pixilated figure:
  - The area of the original scatter plot was subdivided into square "pixels" each having a side of length 10°.
    - The entire plot is covered with $36^2 = 1296$ pixels.
    - A pixel was given a dark color and characterized as "core" if it covered at least 100 points in the scatter plot.
    - A pixel was given a color of medium intensity and characterized as "allowed" if it was not a core pixel but covered at least 8 points in the scatter plot.

Ramachandran Plots (4)

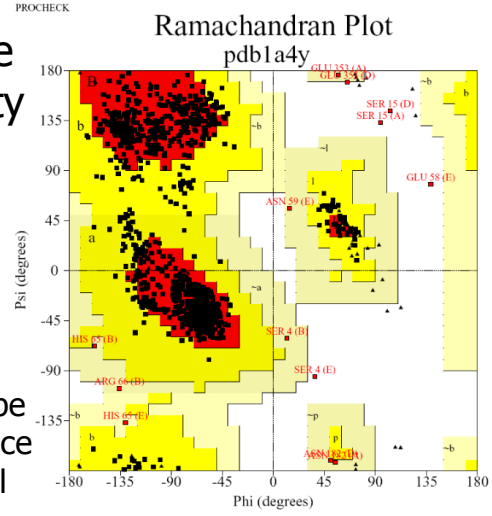
Procedure (continued):

- Pixels in the "generous" region are made by extending an allowed region with a border area of width 20°.
- The generous region is given a color that is lighter than an allowed region.
- Any pixel that is not core, allowed, or generous is given the lightest color or simply left as white.

- So, the darker areas correspond to regions of the plot where the density of points is highest and other colors represent regions of lower density.
  - White regions or those with the lightest color are used to denote areas of the plot where there is the least likelihood of observing a point, for example, $\phi = 90^\circ$ and $\psi = -90^\circ$.

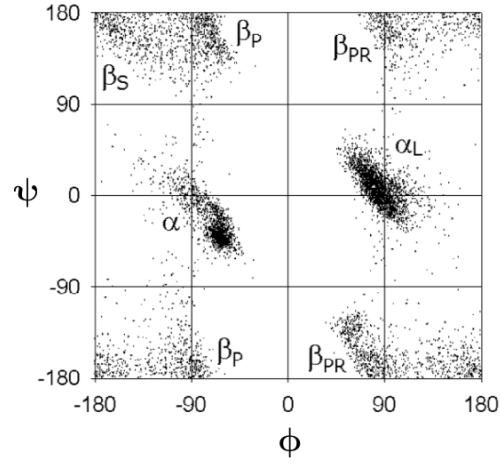
PROCHECK

- PROCHECK checks the stereo-chemical quality of a protein structure.
  - Here the Ramachandran scatter plot for 1A4Y is superimposed over the background distribution discussed earlier.
  - PROCHECK software can be used to detect the existence of any anomalous dihedral angles.



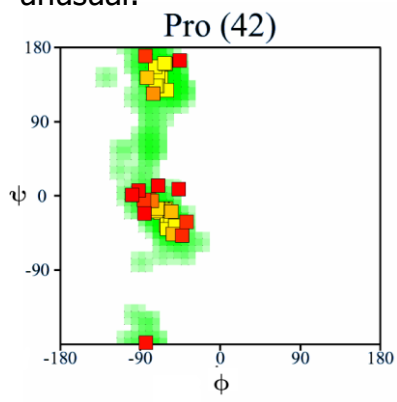
Plots for Individual Residues (1)

- Ramachandran plots may be restricted to the phi/psi angles corresponding to a particular amino acid.
  - Because glycine has less chance of a steric collision, its plot is different.



Plots for Individual Residues (2)

- Proline and the pre-proline alpha carbon are also unusual:



The 42 Proline AAs of 1A4Y.

Note how Phe is veering away from Pro.