5. Torsion angles and pdb files

In the study of space curves, the Frenet frame is used to define torsion and curvature, and these are used to describe the shape of the curve. A long molecule such as DNA or a protein can be thought of as a curve in space. Rather than being described by continuous functions, it is described by line segments which represent covalent bonds between atoms. The concept of curvature and torsion from differentiable curves can be adapted to study the structure of these molecules. Curvature corresponds to the angle between adjacent bonds, and torsion corresponds to the torsion angle discussed here.

5.1. **Torsion Angles.** In the study of molecular structure, torsion angles are frequently used to describe the shape of the molecule. In figure 1, we see four atoms $\mathbf{p}_1$, $\mathbf{p}_2$, $\mathbf{p}_3$, and $\mathbf{p}_4$. Think of the vectors $\mathbf{p}_j$ as vectors giving the coordinates of

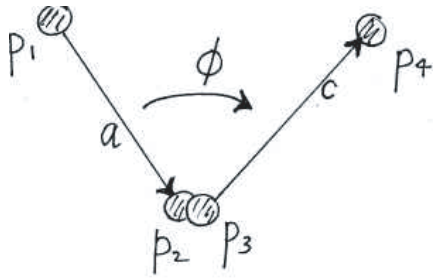


Figure 1. Torsion angle $\phi = \mathrm{Tor}(\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3, \mathbf{p}_4)$. The angle is measured in the plane perpendicular to $\mathbf{b} = \mathbf{p}_3 - \mathbf{p}_2$.

the centers of the atoms. Let

$$\begin{aligned} \mathbf{a} &= \mathbf{p}_2 - \mathbf{p}_1 \\ \mathbf{b} &= \mathbf{p}_3 - \mathbf{p}_2 \\ \mathbf{c} &= \mathbf{p}_4 - \mathbf{p}_3. \end{aligned} \tag{1}$$

and let P$\mathbf{a}$ and P$\mathbf{c}$ be the projections of $\mathbf{a}$ and $\mathbf{c}$ respectively onto the plane perpendicular to $\mathbf{b}$. The angle, ϕ from $-$P$\mathbf{a}$ to P$\mathbf{c}$, measured counterclockwise around $\mathbf{b}$, is the torsion angle. Denote this angle as

$$\phi = \mathrm{Tor}(\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3, \mathbf{p}_4).$$

It is important to note that this angle is measured not between the two vectors $-\mathbf{a}$ and $\mathbf{c}$, but between their projections onto the plane perpendicular to $\mathbf{b}$.

Since the torsion angle depends only on the vectors $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ also write

$$\phi = \tau(\mathbf{a}, \mathbf{b}, \mathbf{c}).$$

In this case the torsion angle is also called the *dihedral angle*. The angle is usually measured in degrees and chosen in the interval $(-180, 180]$.

The dihedral angle can be thought of as the angle between two planes (See figure 2). It is the angle counterclockwise from the normal vector $\mathbf{a} \times \mathbf{b}$ of the plane containing $\mathbf{a}$ and $\mathbf{b}$ to the normal vector $\mathbf{b} \times \mathbf{c}$ of the plane containing $\mathbf{b}$ and $\mathbf{c}$. Both $\mathbf{a} \times \mathbf{b}$ and $\mathbf{b} \times \mathbf{c}$ are in the plane perpendicular to $\mathbf{b}$

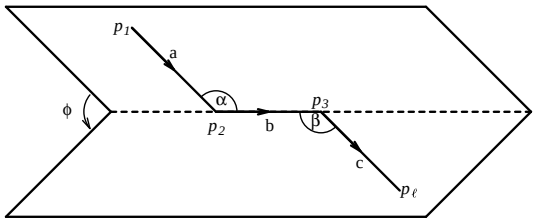


FIGURE 2. The torsion angle ϕ as the angle between planes.

$$\phi = \mathrm{Tor}\left(\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3, \mathbf{p}_4\right) = \tau\left(\mathbf{a}, \mathbf{b}, \mathbf{c}\right).$$

Angles α and β are bond angles.

5.2. **The arg function.** The torsion angle can be defined in terms of the argument of a vector or complex number. Define θ to be the *argument* of a vector $(x, y) \neq (0, 0)$, written $\theta = \arg(x, y)$, if $-180° < \theta \leq 180°$ and

$$\cos\theta = x/\sqrt{(x^2 + y^2)}$$
$$\sin\theta = y/\sqrt{(x^2 + y^2)}.$$

We can also write the argument in terms of complex numbers. The angle $\theta = \arg(x, y)$ if $x + iy$ is written in polar form

$$x + iy = re^{i\theta}.$$

In Maple the command to find the argument of a complex number is `argument`.

Note that if $\mathbf{i}, \mathbf{j}, \mathbf{k}$ is the standard basis then the argument can be written as a dihedral angle,

$$\arg(x, y) = \tau\left(-\mathbf{i}, \mathbf{k}, x\mathbf{i} + y\mathbf{j}\right). \tag{2}$$

The angle ϕ of the spherical coordinates of a 3D vector

$$\mathbf{p} = (\cos\theta, \sin\theta\cos\phi, \sin\theta\sin\phi)$$

can also be thought of as a dihedral angle,

$$\phi = \tau\left(-\mathbf{i}, \mathbf{k}, \mathbf{p}\right). \tag{3}$$

5.3. **The torsion angle formula.** We give a formula for computing the dihedral angle, hence the torsion angle, in terms of the argument.

The Dihedral Angle Formula. For vectors $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$ for which the torsion angle is defined,

$$\tau\left(\mathbf{a}, \mathbf{b}, \mathbf{c}\right) = \arg\left(-|\mathbf{b}|^2\mathbf{a}\cdot\mathbf{c} + (\mathbf{a}\cdot\mathbf{b})(\mathbf{b}\cdot\mathbf{c}), |\mathbf{b}|\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})\right). \tag{4}$$

Proof. Notice that both sides of (4) are unchanged if $\mathbf{a}, \mathbf{b}, \mathbf{c}$ are replaced by $\mathbf{Aa}$, $\mathbf{Ab}$, and $\mathbf{Ac}$ for a rotation $\mathbf{A}$. So we can assume $\mathbf{b}$ is in the direction of $\mathbf{e}_3$. Likewise the equation unchanged if $\mathbf{b}$ is replaced by $\lambda\mathbf{b}$. So we can assume $\mathbf{b} = \mathbf{e}_3$.

The equation is unchanged if $\mathbf{a}$ is replaced by its projection $\mathbf{a} - (\mathbf{a} \cdot \mathbf{b})\mathbf{b}$ perpendicular to $\mathbf{b}$. So we can assume $\mathbf{a}$ is perpendicular to $\mathbf{b}$. As above we can rotate a dilate so that $\mathbf{a} = \mathbf{e}_2$. Let $\mathbf{c} = (x, y, z)'$. Then (4) is equivalent to

$$\tau\left(\mathbf{e}_2, \mathbf{e}_3, (x, y, z)\right) = \arg(-y, x)$$

which is true because the left hand side is the angle from $\mathbf{e}_2 \times \mathbf{e}_3 = \mathbf{e}_1$ to

$$\mathbf{e}_3 \times (x, y, z) = (-y, x, 0).$$

Here is a Maple worksheet to compute torsion angles.

5.4. **Protein torsion angles.**

5.4.1. *Protein backbone torsion angles.* The atoms along a protein backbone are C_α-C-N-C_α-C-N-C_α... in a sequence repeating every third atom. If each atom has a set of coordinates, the torsion angles along the backbone of a protein are named as follows

- the angle $\mathrm{Tor}(C, N, C_\alpha, C)$ is the ϕ torsion angle
- the angle $\mathrm{Tor}(N, C_\alpha, C, N)$ is the ψ torsion angle
- the angle $\mathrm{Tor}(C_\alpha, C, N, C_\alpha)$ is the ω torsion angle

Moving along the backbone we get a sequence of ϕ, ψ and ω torsion angles that can be used to describe the structure of the backbone.

5.4.2. *Protein sidechain torsion angles.* We can also get torsion angles by moving along a side chain. The greek letter subscripts for the atoms along the side chain are indicated in figure 3. For example, the sequence of atoms C_α, C_β, C_γ, C_δ of Lysine determine the χ_2 torsion angle. The atoms C_α, C_β, C_γ, S of Methionine determine the χ_2 torsion angle.

For the χ_1 angle, the first atom used for the torsion angle is the N on the backbone. For example,

- for Leucine, the angle $\mathrm{Tor}(N, C_\alpha, C_\beta, C_\gamma)$ is the χ_1 torsion angle
- for Threonine, the angle $\mathrm{Tor}(N, C_\alpha, C_\beta, O)$ is the χ_{1_2} torsion angle (when there are two χ_1 angles, another subscript is added).

5.5. **Protein Data Bank files.** Structures of all known proteins are stored online at the Protein Data Bank. The files there are called *pdb files*. The structural information contained in the file is a list of three coordinates, (x, y, z), for the centers of every atom in the molecule (although hydrogen atoms are sometimes left out because they are small and their positions can be determined from the positions of the other atoms). For example, a file identified as 1E0P contains the coordinates for a protein called bacteriorhodopsin. Here is part of the file which can be downloaded from the RCSB Protein Data Bank:

```
ATOM   1557  CB  ILE A 205     -14.646  17.302  50.448  1.00 21.52 C
ATOM   1558  CG1 ILE A 205     -13.253  16.800  50.104  1.00 19.39 C
ATOM   1559  CG2 ILE A 205     -15.422  17.496  49.149  1.00 22.80 C
ATOM   1560  CD1 ILE A 205     -13.299  15.453  49.472  1.00 18.37 C
ATOM   1561  N   PHE A 206     -12.336  19.006  52.127  1.00 22.92 N
ATOM   1562  CA  PHE A 206     -11.262  18.905  53.109  1.00 23.61 C
```

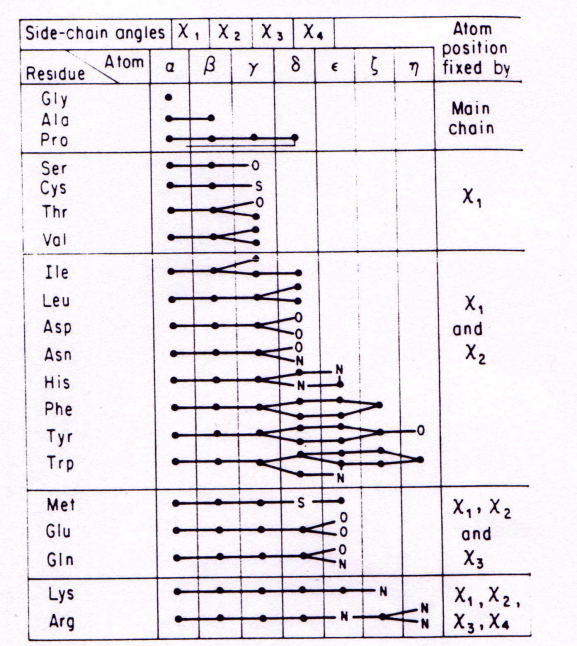


Figure 1. Flexibility of amino acid side-chains. The Figure shows the chi angle values required to fix the positions of side-chain atoms in each residue type.

FIGURE 3. Naming convention for torsion angles along the sidechains of a protein

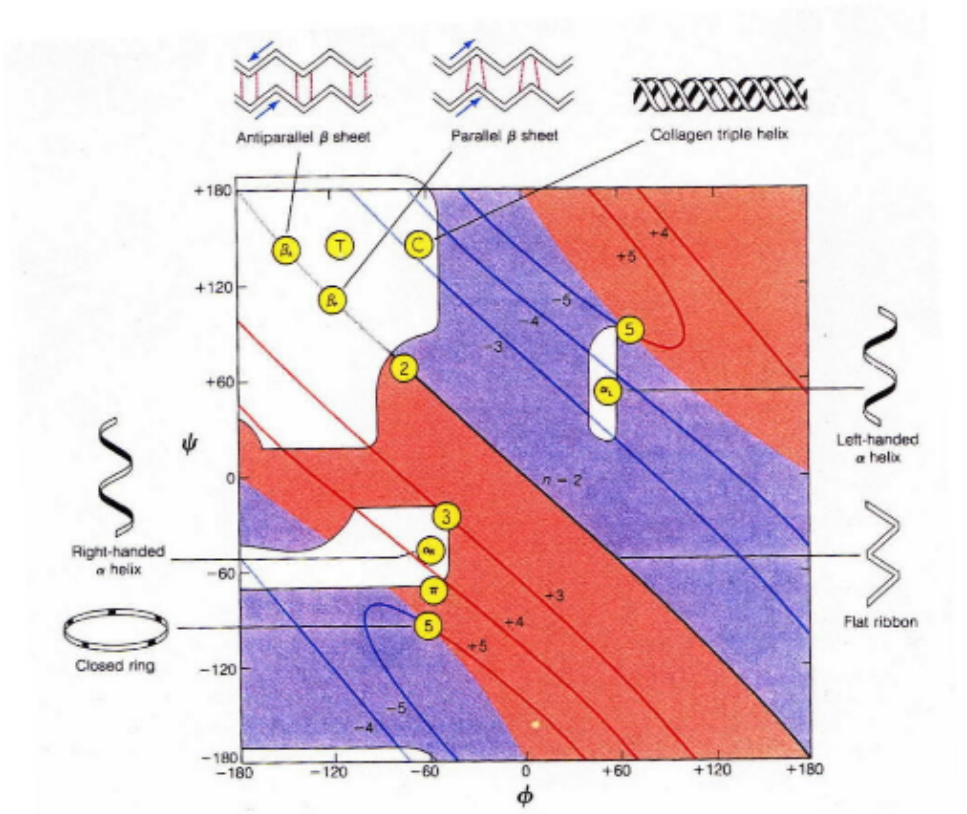
```
ATOM   1563  C   PHE A 206      -11.438  19.982  54.171  1.00 26.34 C
```

The x, y and z coordinates are contained in columns 7, 8, and 9 respectively. Other important information is

- column 2, the number of the atom in the list
- column 3, the position of the atom in the protein using the naming convention in table (3). Note that, for example, CB is written instead of C_β.
- column 4, the three letter code for the amino acid
- column 5, the number of the amino acid in the list of amino acids along the protein.

5.6. **Ramachandran diagram.** The ϕ, ψ pairs of torsion angles for each amino acid along the backbone of a protein can be plotted as points in a rectangle, and this plot is called a *Ramachandran plot*. The ω torsion angle in generally considered to be 180° since the peptide bond is planar, and so it is not plotted. The Ramachandran plot gives information about the secondary structure of the protein. (See figure 4.)

A regular protein backbone structure is one where all of the ω torsion angles are 180 degrees and all of the ϕ, ψ pairs at alpha carbons have the same value. The diagram below indicates the types of structures we get for different pairs. A regular alpha helix, for example, corresponds to $(\phi, \psi) = (-60°, -50°)$. The level curves indicate the number of residues per turn, the number of amino acids for each 180



ure 6.9

FIGURE 4. The Ramachandran plot giving the number of residues per turn of the helix.

degree turn about the axis of the helix, for the corresponding regular structure. All of the structures can be thought of as helices with various numbers of residues per turn.

5.7. **Torsion angles on the diamond packing.** The diamond packing is a set of points in space where the centers of carbons of a diamond crystal lie. The diamond packing is obtained from the face centered cubic lattice (the set of points with integer coordinates adding up to an even number) by adding to it points of the face centered cubic lattice moved over by the vector $(1/2, 1/2, 1/2)$. So

$$\text{Diamond packing} = \text{fcc} \cup \{\text{fcc} + (1/2, 1/2, 1/2)\}\,.$$

By moving on a path through the diamond packing you can get torsion angles of 180, 60, and -60 only (or undefined if two consecutive vectors are parallel). Since the -60 degree torsion angles are close to the ones for alpha helices, attempts have been made to model proteins by putting atoms in a protein on points in a diamond packing. Maple demo

5.8. **Appendix, properties of cross product.** Here are some useful formulas involving the cross product and the dot product

$$(\mathbf{a} \times \mathbf{b}) \cdot (\mathbf{c} \times \mathbf{d}) = (\mathbf{a} \cdot \mathbf{c})(\mathbf{b} \cdot \mathbf{d}) - (\mathbf{a} \cdot \mathbf{d})(\mathbf{b} \cdot \mathbf{c})$$

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = (\mathbf{a} \cdot \mathbf{c})\mathbf{b} - (\mathbf{a} \cdot \mathbf{b})\mathbf{c}$$