Nuclear Magnetic Resonance (NMR) is another method besides crystallography that can be used to find structures of proteins. NMR spectroscopy is the observation of spins of atoms and electrons in a molecule that is placed in a magnetic field. The spins precess at a frequency in the radio frequency range and the frequency can be detected by the electrical signal that it generates.

We briefly discuss the physics of NMR, and then describe distance geometry, a mathematical theory that can be used to find the protein structure from some types of NMR data.

7.1. Larmor frequency. Spins placed in a magnetic field precess; they wobble like a spinning top. Only certain isotopes of molecules found in organic compounds have spins that react to the magnetic field; the most common ones used in proteins are $^1$H, $^{13}$C, $^{15}$N. The isotopes $^{13}$C and $^{15}$N are not in common abundance, so specially prepared protein samples must be used.

![Figure 1](image-url)

**Figure 1.** An NMR experiment is the observation of the precession of nuclear spins in the presence of a magnetic field. The large arrow represents the magnetic field of the magnet, the small arrow represents magnetic field of the nucleus which is precessing like a spinning top as the tip of the arrow moves on the indicated circle.

The frequency of precession is called the *Larmor frequency* and it is determined mainly by the type of atom and the strength of the magnetic field. The basic NMR
equation is,

\[ \omega = \gamma B_0 \]  

where

- \( \omega \) = the Larmor frequency, i.e., the angular frequency of the precession in radians per second
- \( \gamma \) = gyromagnetic ratio, a constant depending on the type of the atom
- \( B_0 \) = the intensity of the magnetic field

7.2. Splitting and chemical shift. Equation (1) assumes that the detected frequency depends only on the atom and the intensity of the magnetic field. There is another factor, however. The magnetic field intensity \( B_0 \) is not the same everywhere in the molecule; it is affected by neighboring atoms and electrons. Neighboring spins have their own magnetic field and this perturbs the field of the magnet and changes the frequency of precession.

This is illustrated by the NMR spectrum of the hydrogen atoms in the molecule Toluene (figure 2). The spectrum is a Fourier transform of the electrical signal showing the intensities (vertical axis) of certain frequencies (horizontal axis). The spectrum can be thought of as the absolute values of Fourier coefficients for the function giving the signal as a function of time. If the signal is the real part of

\[ s(t) = \sum_{j=1}^{n} a_j e^{2\pi i \omega_j t} \]  

then the spectrum gives absolute values \( |a_j| \), \( j = 1 \ldots n \) at frequencies \( \omega_j \) cycles per second.

The spectrum of Toluene shows that the peak frequencies cluster around two values. Also a reference signal is shown for hydrogens which are not part of any molecule. The frequencies of the hydrogens in the molecule are different (shifted) from the reference signal. The observed difference is divided by the frequency of the reference signal times \( 10^6 \), and the change in frequency is reported in parts per million or ppm. The change is called a chemical shift. The spectrum shows that the 3 methyl hydrogens (to the right) are shifted less that the other 5 hydrogens. From the symmetry of the molecule it is easy to see that the peak on the right comes from the methyl hydrogens, the hydrogens on the CH\(_3\) group at the end of the molecule. This is because they all have the same relation to the rest of the molecule which causes the shift.

The spectrum of Toluene shows that we can infer facts about the shape of a molecule by looking at the spectrum. This suggests that we can find chemical structures of larger molecules by NMR. However the spectrum of a protein is much more difficult to interpret. Below is the NMR spectrum of all the hydrogens in the protein thioredoxin indicating which part of the molecule the hydrogen signals are coming from. Although some aspects of the structure can be deduced from the spectrum, it would be difficult to find coordinates of the atoms from this spectrum.
The 300 MHz $^1$H NMR spectrum of Toluene

**Figure 2.** The spectrum of Toluene. In a magnetic field, the hydrogens spins in the molecule precess at different frequencies depending on the surrounding atoms and electrons. This is because the neighboring atoms and electrons have spins which contribute to the magnetic field. In Toluene the frequencies are seen in two separate ranges shifted from the reference signal.

There is another type of spectrum (figure [3]) called a 2D NOESY spectrum. This experiment observes two frequencies from an atom, so the intensity is a function of two variables. The figure shows level curves for high intensity, which looks like a set of points. These can be used to estimate distances between atoms. Such estimates are called *distance constraints*. Distance constraints can be used to find atomic coordinates using techniques of distance geometry. Similar spectra can be used to find *orientational constraints* which measure angles rather than distances. Orientational constraint measure the angle between covalent bond vectors and the vector giving the direction of the magnetic field.

**7.3. Distance geometry.** Determination of the structure of a protein from distance constraints caused renewed interest in an old branch of mathematics called *distance geometry*. A protein structure is a list of coordinates for the atoms. Rather than coordinates, we could consider a list of distances between atoms. Distances, unlike coordinates, are invariant under rotation and translation, so they are also useful in shape analysis.
**Figure 3.** The hydrogen NMR spectrum of the protein thioredoxin. Indicated on the spectrum are the parts of the protein responsible for each part of the signal. The signal is shifted according to the unique chemical environment of each part of the protein.

Consider a sequence of points in 3D space. Information about distances between a sequence of points can be put into a matrix called a distance matrix. The atoms in a structure are numbered 1 to \( n \). The distance matrix is an \( n \times n \) matrix. The entry in row \( i \) and column \( j \) is the square of the distance between points \( i \) and \( j \). The object of distance geometry is to find coordinates of the points from the distance matrix.

There is no unique list of coordinates since a rotated and translated set of coordinates gives the same distance matrix, however, coordinates can be found which are unique up to a Euclidean motion. The situation is analogous to viewing a pdb file. Protein viewers change all the coordinates in the structure through rotations and translations, but the shape of the protein and the distances between atoms stays fixed.

7.3.1. **Distance matrix example.** Label the rows and columns below by the letters a through h. The corresponding entry gives the square of the distance between the
Figure 4. NOESY spectrum of thioredoxin. A labeled point in the spectrum indicates that the signal comes from the interaction between a certain pair of hydrogen atoms. That the two hydrogen atoms interact indicates that they are not far apart. The distance between the atoms can be estimated by the intensity of the signal.

two corresponding points on the cube with sides of length 1 in figure 5

\[
\begin{bmatrix}
0 & 1 & 2 & 1 & 1 & 2 & 3 & 2 \\
1 & 0 & 1 & 2 & 2 & 1 & 2 & 3 \\
2 & 1 & 0 & 1 & 3 & 2 & 1 & 2 \\
1 & 2 & 1 & 0 & 2 & 3 & 2 & 1 \\
1 & 2 & 3 & 2 & 0 & 1 & 2 & 1 \\
2 & 1 & 2 & 3 & 1 & 0 & 1 & 2 \\
3 & 2 & 1 & 2 & 2 & 1 & 0 & 1 \\
2 & 3 & 2 & 1 & 1 & 2 & 1 & 0
\end{bmatrix}
\]

7.4. Obtaining coordinates from a distance matrix. The central problem in distance geometry is to recover the coordinates of a sequence of points, up to a Euclidean motion of space, from the distance matrix. The solution is presented in this section.

Suppose there are \( n + 1 \) vectors \( \mathbf{v}_0 \ldots \mathbf{v}_n \). For simplicity, translate the points so that one vector \( \mathbf{v}_0 = 0 \). Consider the vectors as columns of a \( 3 \times n \) matrix

\[
\mathbf{M} = (\mathbf{v}_1, \ldots, \mathbf{v}_n)
\]

Finding the vectors \( \mathbf{v}_j \) from the distance matrix follows in two steps
7.4.1. **Gram matrix from distance matrix.** The entries in the distance matrix $D$ are $|v_j - v_k|^2$, $j, k = 0, 1, \ldots, n$,

\[
D = \begin{pmatrix}
0 & |v_1|^2 & |v_2|^2 & \cdots & |v_n|^2 \\
|v_1|^2 & 0 & |v_1 - v_2|^2 & \cdots & |v_1 - v_n|^2 \\
|v_2|^2 & |v_2 - v_1|^2 & 0 & \cdots & |v_2 - v_n|^2 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
|v_n|^2 & |v_n - v_1|^2 & |v_n - v_2|^2 & \cdots & 0
\end{pmatrix}
\]

The gram matrix $G$ is given by

\[
G = \begin{pmatrix}
v_1 \cdot v_1 & v_1 \cdot v_2 & \cdots & v_1 \cdot v_n \\
v_2 \cdot v_1 & v_2 \cdot v_2 & \cdots & v_2 \cdot v_n \\
\vdots & \vdots & \ddots & \vdots \\
v_n \cdot v_1 & v_n \cdot v_2 & \cdots & v_n \cdot v_n
\end{pmatrix}
\]

The gram matrix can be found from the distance matrix using the vector identity

\[
-2 v_i \cdot v_j = |v_i - v_j|^2 - |v_i|^2 - |v_j|^2.
\]

All the information on the right hand side of this equation is in the distance matrix $D$. Using (4) we can get the gram matrix from the distance matrix by row and column operations.
(1) subtract the first row of \( D \) from each row
(2) subtract the first column from each column.

The result is \(-2G\) bordered by zeros the first row and column.

7.5. **Coordinates from the gram matrix.** Given \( G \) obtained from a distance matrix of vectors in 3D space, the problem is to find a square root of \( G \), a \( 3 \times n \) matrix \( M \) such that

\[
G = M'M.
\]

This can be done by finding eigenvectors and eigenvalues. Since \( G \) can be written in the form \( [5] \), it has real, non-negative eigenvalues. Since \( M \) is \( 3 \times n \) there are at most 3 non-zero eigenvalues. (These statements are left as an exercise.)

Write the eigenvectors of \( G \) in a matrix \( V \) and the eigenvalues as diagonal entries in a diagonal matrix \( E \). Then

\[
GV = VE.
\]

Arrange the eigenvalues so that

\[
E = \begin{pmatrix}
 E_1 & 0 \\
 0 & 0
\end{pmatrix}
\]

where \( E_1 \) is a diagonal matrix with 3 non-negative entries. Eigenvectors can always be found which are an orthonormal set,

\[
V'V = VV' = I.
\]

Write \( V = (V_1, V_2) \) where \( V_1 \) is a \( 3 \times n \) matrix whose columns are eigenvectors corresponding to the diagonal elements of \( E_1 \), and where the columns of the \( 3 \times (3 - n) \) matrix \( V_2 \) have eigenvalue 0. The columns of \( V_2 \) are in the kernel of \( G \).

Now construct a diagonal matrix whose entries on the diagonal are the square roots of the entries of \( E \). Since the entries on the diagonal are non-negative, the square roots are also non-negative real numbers. Call this matrix \( \sqrt{E} \). From \( [6] \) and \( [8] \) it follows that

\[
G = V\sqrt{E}\sqrt{E}' = M'M.
\]

Where

\[
M = \sqrt{E}V' = \sqrt{E} \begin{pmatrix}
 V_1' \\
 V_2'
\end{pmatrix} = \begin{pmatrix}
 \sqrt{E_1}V_1' \\
 0
\end{pmatrix}.
\]

The first three rows of the matrix \( M \) give coordinates of points with the desired gram and distance matrices.

This procedure for finding coordinates is best illustrated using 4 points \( (n = 4) \).
7.5.1. Example. Find coordinates of points in 3D space giving the distance matrix

\[
\begin{bmatrix}
0 & 2 & 1 & 1 \\
2 & 0 & 3 & 1 \\
1 & 3 & 0 & 2 \\
1 & 1 & 2 & 0 \\
\end{bmatrix}
\]

Maple demo

7.5.2. Example. Find coordinates of points in 3D space giving the distance matrix

\[
\begin{bmatrix}
0 & 2 & 1 & 1 & 2 \\
2 & 0 & 3 & 1 & 2 \\
1 & 3 & 0 & 2 & 1 \\
1 & 1 & 2 & 0 & 1 \\
2 & 2 & 1 & 1 & 0 \\
\end{bmatrix}
\]

Maple demo