Molecules in 3D: theory and visualization

# Structural properties of molecules

## Bond lengths

A bond length is the distance between two-covalently bound atoms. The unit of bond lengths is typically Ångström (Å) where 1 Å = 0.1 nm. However, all units of distance can in principle be used.

Factors influencing bond lengths include:

* Size of atoms and electronegativity, e.g. C-F < C-Cl < C-Br:  
  A diagram of a molecule

  AI-generated content may be incorrect.
* Hybridization, e.g. C(sp2)-C(sp2) < C(sp3)-C(sp3):  
  A close-up of several molecules

  AI-generated content may be incorrect.
* Charge, e.g. C-NH2 < C-NH3+:  
  A close-up of several molecules

  AI-generated content may be incorrect.
* Molecular environment, e.g. F3C-CH3 < H3C-CH3:  
  A close-up of a white and blue molecule

  AI-generated content may be incorrect.

## Bond angles

A bond angle is an angle formed between three covalently bound atoms (two bonds) and is typically shown in degrees (°). Bond angles are influenced by factors similar to bond lengths so I will simply provide a few examples below:

A close-up of several molecules

AI-generated content may be incorrect.

A close-up of several molecules

AI-generated content may be incorrect.

## Dihedral angles (also known as torsional angles)

A dihedral angle is an angle formed by four covalently bound atoms (three bonds) and is typically shown in degrees (°). The energy profile of a dihedral angle is rather intricate and is also influenced by factors similar to bond lengths and bond angles. The concept will be illustrated through one of the easiest cases to understand, i.e. butane.

Afbeelding met diagram, cirkel, lijn, symbool

Door AI gegenereerde inhoud is mogelijk onjuist.Afbeelding met diagram, cirkel, clipart, Lettertype

Door AI gegenereerde inhoud is mogelijk onjuist.Afbeelding met diagram, cirkel, schets, clipart

Door AI gegenereerde inhoud is mogelijk onjuist.Afbeelding met diagram, cirkel, schets, clipart

Door AI gegenereerde inhoud is mogelijk onjuist.Afbeelding met diagram, cirkel, symbool, lijn

Door AI gegenereerde inhoud is mogelijk onjuist.Afbeelding met diagram, cirkel, schets, clipart

Door AI gegenereerde inhoud is mogelijk onjuist.Afbeelding met diagram, cirkel, lijn, symbool

Door AI gegenereerde inhoud is mogelijk onjuist.Afbeelding met tekst, Perceel, lijn, diagram

Door AI gegenereerde inhoud is mogelijk onjuist.

## Characterizing non-covalent interactions

The use of these molecular desriptors is not limited to covalently bound atoms. We can also define distances, angles or even dihedrals between non-covalent atoms. This is specially useful in the characterization of non-covalent interactions such as hydrogen bonds. The orientation of the atoms participating in the non-covalent interaction will determine the strength of the interaction:

* Distance between water molecules in a dimer
* Angle between water molecules in a dimerAfbeelding met tekst, schermopname

  Door AI gegenereerde inhoud is mogelijk onjuist.

# Molecules in 3D space

For visualizing molecules, we have two options: we either describe the entire molecule through individual bond lengths, angles and dihedrals between specified atoms – this is called “internal coordinates” – or use a simple representation in Cartesian coordinates in 3D space. For large biomolecules, the latter is typically more easy to interpret and hence every atom will simply have their own specific x-, y-, and z-coordinate. Relevant distances and angles can still be obtained through geometric principles. We will briefly discuss 2 file formats that use Cartesian coordinates for visualizing molecules: the simple .xyz format and the .pdb format which is the most common format for biomolecular molecules.

## The .xyz format

.xyz files are a very simple format but has strict rules. The first line contains the number of atoms and the second line is a title card which you can use to describe the molecule. The following lines are used for the atoms, one line per atom, and each line has four columns (atom element, x coordinate, y coordinate, z coordinate). The total number of lines in a .xyz file will always be the *n*+2 (with *n* the number of atoms in the molecule).

Afbeelding met tekst, schermopname, Lettertype, nummer

Door AI gegenereerde inhoud is mogelijk onjuist.

## The .pdb format

PDB can refer to both a file format as well as the Protein DataBase (which uses the .pdb format for its biomolecular strutcures).

A screenshot of a computer code

AI-generated content may be incorrect.The .pdb format is a lot different from the .xyz format. Lines in in a .pdb start with a particular keyword that indicates what data is present in that particular row of the file.

On the right hand side, you can see an example of the beginning of a PDB file (in this case PDB 6ZUV). The first two rows are HEADER and TITLE, respectively, followed by a lot of additional information of the structure. This information is typically a large block and actual coordinates only start from ATOM or HETATM records which is indicated in the second screenshot.

A screenshot of a computer code

AI-generated content may be incorrect.

Each ATOM record has again a lot more information than the .xyz format. In order you have: atom number, atom name, residue name, chain identifier, residue number, x coordinate, y coordinate, z coordinate, occupancy, B value and element. Please note that this is not all necessary information. There can, for example, also be .pdb files without chain identifier. Additionally, the columns occupancy and B value are related to crystallographic data and are not of interested for us so we will not elaborate on this.

### The Protein DataBase

The Protein Database <https://www.rcsb.org> is probably the single most important database related to biomolecular structures and it houses a vast colection of protein, nucleic acid, carbohydrate or lipid structures. It is a very easy-to-navigate website and important information is directly available for all entries. More information will be provided by hands-on examples.

A screenshot of a computer

AI-generated content may be incorrect.