## **Amino Acids in the PDB**

**Authors**: Alexandra Pettit, Carleton University, Ottawa, Ontario, Canada and Shuchismita Dutta, Rutgers University, NJ

**ASBMB Learning Objectives**

(<https://www.asbmb.org/education/core-concept-teaching-strategies/foundational-concepts/structure-function>)

1. Biological macromolecules are large and complex

* Students should be able to **describe the basic units of the macromolecules** and the types of linkages between them *(Introductory)*.

2. Structure is determined by several factors

* Students should be able to **recognize the repeating units in biological macromolecules** and be able to discuss the structural impacts of the covalent and noncovalent interactions involved *(Introductory)*.

### Amino acids:

#### Role in biology:

* Amino acids are the building blocks of proteins/peptides. When the amino acid polymer is short 20-50 amino acids in most scientific literature (2-25 amino acids in PDB related searches etc.) they are sometimes referred to as peptides, while longer polymers are called proteins.
* Amino acids to proteins: Amino acids form covalent linkages,known as peptide bonds,with each other through a dehydration reaction to form polymers. This reaction is carried out in cells at the ribosome (protein synthesis). Amino acids bound to their cognate tRNA molecules are transported to the ribosome to form proteins.
* Proteins to amino acids: Specific enzymes called proteases can break the peptide bonds at specific sites or from one end of the protein/peptide to yield amino acids.

|  |  |
| --- | --- |

*Figure 1: A generic amino acid showing the amino, carboxyl, and R-groups. Left- Chemical drawing and Right- 3D structure of the amino acid found in the Chemical Component Dictionary (maintained by the wwPDB).*

#### Chemical structure of amino acids:

* All amino acids have an amino group and a carboxylic acid group attached to a central carbon.
* The central carbon has a hydrogen atom and another group (ranging from a single hydrogen atom to many different types of groups, called a side chain or R group).
* It is the properties of this R group that gives the amino acid specific properties (aromatic, aliphatic, charged, polar etc.)
* Charge: All amino acids have two ionizable groups (the amino and carboxyl groups). Some of them have a third ionizable group in the side chains.
  + The **α-carboxyl (-COOH) group** is acidic and can lose a proton at the pKa (between 1.8 and 2.6). So at the physiological pH it is negatively charged.
  + The **α-amino (-NH 3+ ) group** is basic and can accept protons at pH below its pKa (between 8.8 and 10.6). So at the physiological pH it is positively charged.
  + The side chains of **acidic amino acids (Asp and Glu)** have a carboxyl group with pKa ~4. At this pH the group loses a proton and becomes negatively charged.
  + The side chains of **basic amino acids (Arg and Lys)** gain a proton on their side chain and become positively charged. The pKa of **His** side chain is closer to the physiological pH (~6.1) so it is often involved in sensing small pH changes around the physiological pH.
  + Did you know that the side chains of **Tyr and Cys** also have pKa values at which ionized amino and carboxyl groups of each amino acid are charged at pH 10 and 8 respectively? Learn more about pKa values and other [properties of amino acids](https://www.sigmaaldrich.com/US/en/technical-documents/technical-article/protein-biology/protein-structural-analysis/amino-acid-reference-chart).

Note: The **amino acids in the** [**Chemical Component Dictionary**](https://www.wwpdb.org/data/ccd) **are not charged**. They are built as building blocks of proteins which lose specific atoms with the formation of the peptide bond - i.e., the carboxyl (COOH) group loses one H atom and the amino (NH2) group loses one of its two H atoms when an amino acid covalently links to another amino acid. Most structures in the PDB do not include hydrogen atom coordinates (since they can not be seen in X-ray or EM structures). So explicit definitions of protonated and deprotonated states can not be defined based on experimental data. The charged states of the side chains are also implied.

#### Chirality:

The central carbon is chiral so the way in which the various atoms and groups are attached to it matter.

* + In **L-amino acids**, when looking down the hydrogen towards the central carbon, the carbonyl group is on the left, the sidechain or R group is in the middle and the N or amino group is on the right. Traveling from the carboxyl to amino group clockwise, you can spell the word CO-R-N. Sometimes this is referred to as the *“CORN rule”*.
  + In **D-amino acids**, the word CO-R-N can be spelled traveling in an anticlockwise manner.

|  |  |
| --- | --- |
| ALA (L-amino acid) | DAL (D-amino acid) |

Did you know that the definitions, chemical formulae, and representative coordinates for all amino acids are available in the Chemical Component Dictionary (maintained by the [worldwide PDB](https://www.wwpdb.org/) (wwPDB)? You can search for these components or ligands using the [Ligand Expo](http://ligand-expo.rcsb.org/index.html) tool or directly from the [RCSB PDB](https://www.rcsb.org/) website.

### Explore:

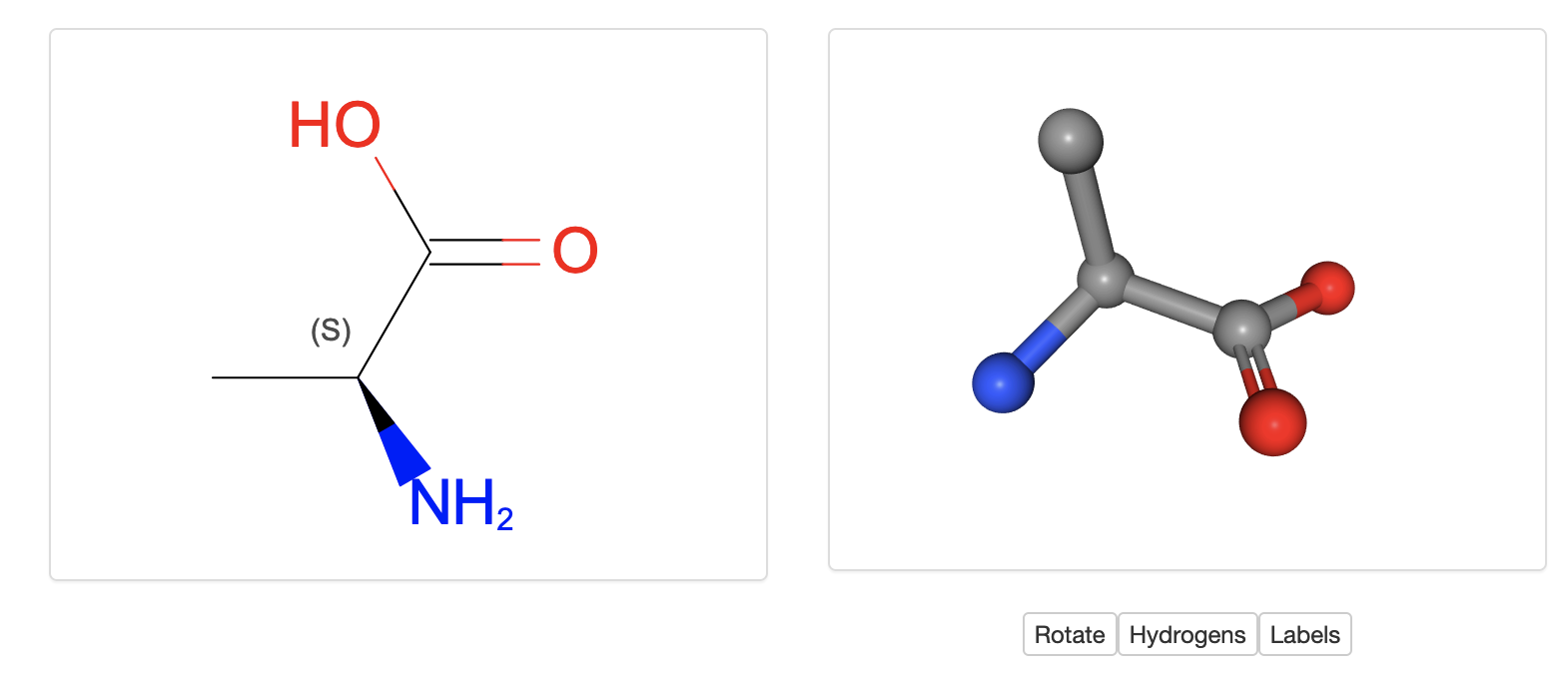
* Explore the [20 standard amino acids](https://www.rcsb.org/search?request=%7B%22query%22%3A%7B%22type%22%3A%22group%22%2C%22logical_operator%22%3A%22and%22%2C%22nodes%22%3A%5B%7B%22type%22%3A%22group%22%2C%22logical_operator%22%3A%22and%22%2C%22nodes%22%3A%5B%7B%22type%22%3A%22group%22%2C%22nodes%22%3A%5B%7B%22type%22%3A%22terminal%22%2C%22service%22%3A%22text_chem%22%2C%22parameters%22%3A%7B%22attribute%22%3A%22rcsb_chem_comp_container_identifiers.comp_id%22%2C%22operator%22%3A%22in%22%2C%22negation%22%3Afalse%2C%22value%22%3A%5B%22gly%22%2C%22ala%22%2C%22ser%22%2C%22thr%22%2C%22val%22%2C%22leu%22%2C%22ile%22%2C%22tyr%22%2C%22trp%22%2C%22phe%22%2C%22his%22%2C%22asp%22%2C%22asn%22%2C%22glu%22%2C%22gln%22%2C%22pro%22%2C%22arg%22%2C%22lys%22%2C%22met%22%2C%22cys%22%5D%7D%7D%5D%2C%22logical_operator%22%3A%22and%22%7D%5D%2C%22label%22%3A%22text_chem%22%7D%5D%7D%2C%22return_type%22%3A%22mol_definition%22%2C%22request_options%22%3A%7B%22paginate%22%3A%7B%22start%22%3A0%2C%22rows%22%3A25%7D%2C%22scoring_strategy%22%3A%22combined%22%2C%22sort%22%3A%5B%7B%22sort_by%22%3A%22score%22%2C%22direction%22%3A%22desc%22%7D%5D%7D%2C%22request_info%22%3A%7B%22query_id%22%3A%2222bbd5c22f2ad12a74a6e733045968ba%22%7D%7D). Note that these are L-amino acids. The D-amino acids are also available in the dictionary.
* Besides these common amino acids, PDB structures have many modified amino acids that are either present in nature or designed. Explore all [L-peptide linking](https://www.rcsb.org/search?request=%7B%22query%22%3A%7B%22type%22%3A%22group%22%2C%22logical_operator%22%3A%22and%22%2C%22nodes%22%3A%5B%7B%22type%22%3A%22group%22%2C%22logical_operator%22%3A%22and%22%2C%22nodes%22%3A%5B%7B%22type%22%3A%22group%22%2C%22nodes%22%3A%5B%7B%22type%22%3A%22terminal%22%2C%22service%22%3A%22text_chem%22%2C%22parameters%22%3A%7B%22attribute%22%3A%22chem_comp.type%22%2C%22operator%22%3A%22exact_match%22%2C%22negation%22%3Afalse%2C%22value%22%3A%22L-peptide%20linking%22%7D%7D%5D%2C%22logical_operator%22%3A%22and%22%7D%5D%2C%22label%22%3A%22text_chem%22%7D%5D%7D%2C%22return_type%22%3A%22mol_definition%22%2C%22request_options%22%3A%7B%22paginate%22%3A%7B%22start%22%3A0%2C%22rows%22%3A25%7D%2C%22scoring_strategy%22%3A%22combined%22%2C%22sort%22%3A%5B%7B%22sort_by%22%3A%22score%22%2C%22direction%22%3A%22desc%22%7D%5D%7D%2C%22request_info%22%3A%7B%22query_id%22%3A%22c1a76b4b1572b4cd57ddeb4aaa5fc81c%22%7D%7D) and [D-peptide linking](https://www.rcsb.org/search?request=%7B%22query%22%3A%7B%22type%22%3A%22group%22%2C%22logical_operator%22%3A%22and%22%2C%22nodes%22%3A%5B%7B%22type%22%3A%22group%22%2C%22logical_operator%22%3A%22and%22%2C%22nodes%22%3A%5B%7B%22type%22%3A%22group%22%2C%22nodes%22%3A%5B%7B%22type%22%3A%22terminal%22%2C%22service%22%3A%22text_chem%22%2C%22parameters%22%3A%7B%22attribute%22%3A%22chem_comp.type%22%2C%22operator%22%3A%22exact_match%22%2C%22negation%22%3Afalse%2C%22value%22%3A%22D-peptide%20linking%22%7D%7D%5D%2C%22logical_operator%22%3A%22and%22%7D%5D%2C%22label%22%3A%22text_chem%22%7D%5D%7D%2C%22return_type%22%3A%22mol_definition%22%2C%22request_options%22%3A%7B%22paginate%22%3A%7B%22start%22%3A0%2C%22rows%22%3A25%7D%2C%22scoring_strategy%22%3A%22combined%22%2C%22sort%22%3A%5B%7B%22sort_by%22%3A%22score%22%2C%22direction%22%3A%22desc%22%7D%5D%7D%2C%22request_info%22%3A%7B%22query_id%22%3A%2245a1ba6ca5a129908c4aa0b0abcde264%22%7D%7D) amino acids in PDB’s Chemical Component Dictionary to learn more. Make a gallery of these amino acids by clicking on the  icon at the top of the list of amino acids listed on the page for your reference.

### Examine:

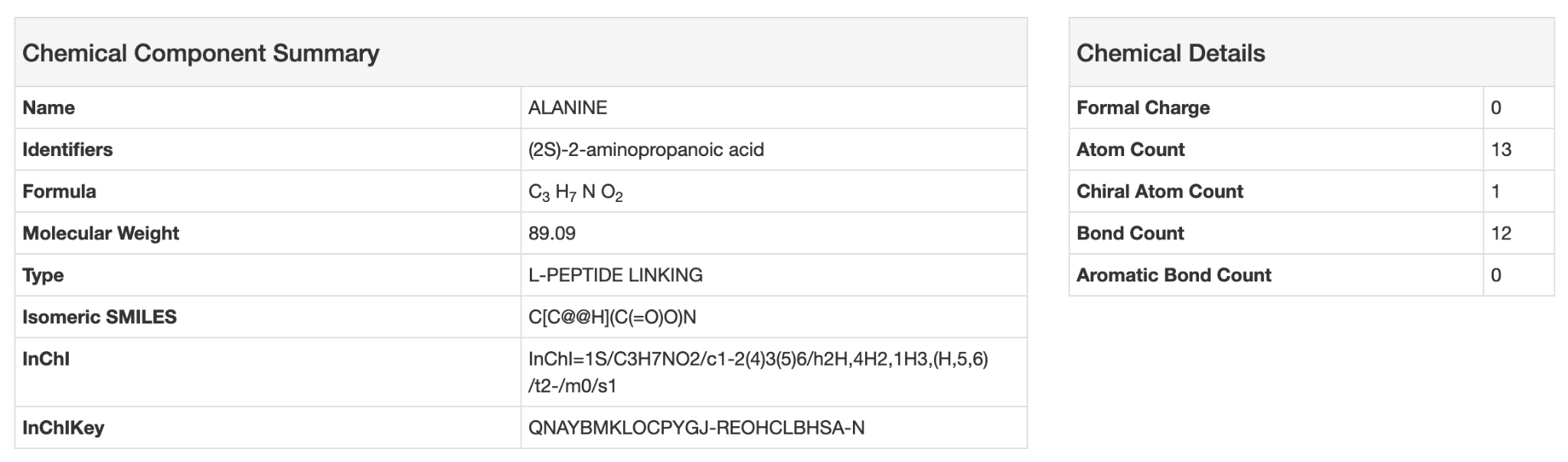
To learn more about each amino acid you can examine its ligand summary page.

For example for [ALA](https://www.rcsb.org/ligand/ALA), you can

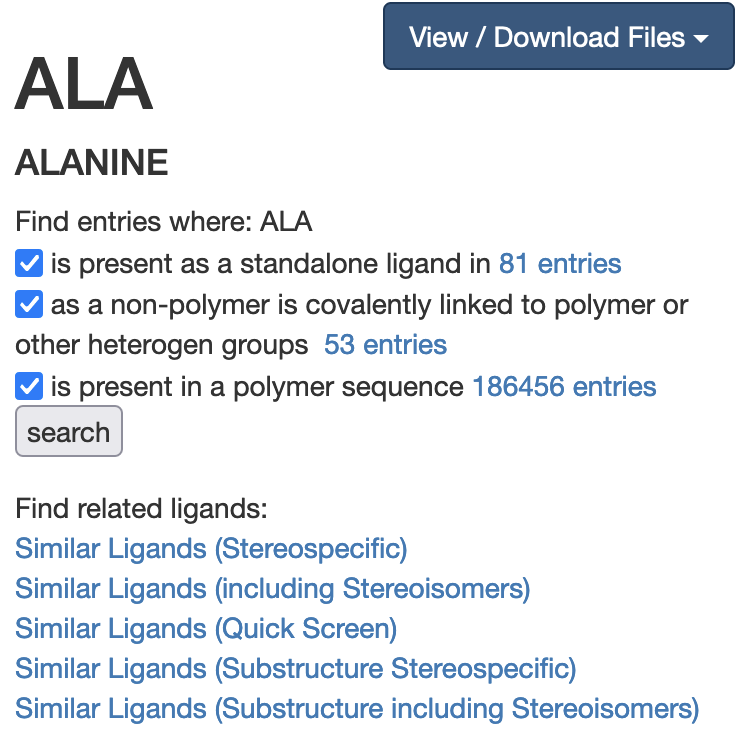
* Examine the 2D and 3D structures of the molecule



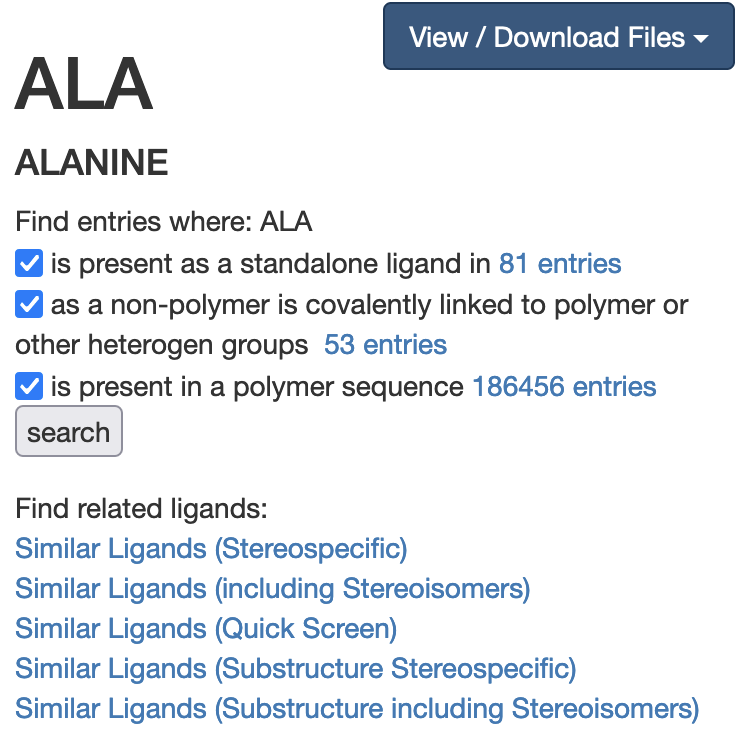
* Access the complete chemical name, formula, and other chemical identifiers



* Download files, identify PDB entries where the molecule is present as part of a polymer, covalently bound, or interacting through non-covalent interactions.



* Find other ligands in the dictionary with the same or similar chemical properties.



* Where available, information about the molecule being a drug or nutraceutical and links to the DrugBank to access additional information.

