



## Appendix

### Table of Contents

<b>A2</b>	Ethics Background
<b>A3</b>	Creating Discussion Ground Rules
<b>A4</b>	Amino Acid Abbreviations and Chemistry Resources
<b>A5</b>	Codons and Amino Acid Chemistry
<b>A6</b>	Behind the Scenes with the NCBI Databases and the Entrez Search Engine
<b>A7</b>	Understanding BLAST
<b>A9</b>	Finding Structures in the NCBI Structure Database

## Ethics Background

### Principles – Respect, Maximize Benefits/Minimize Harms, and Justice

#### Summary

The focus of this perspective is on the four **principles** supported by or compromised by the question or issue at hand.

Philosophers Tom Beauchamp and Jim Childress identify four principles that form a commonly held set of pillars for moral life.

Respect for Persons	Acknowledge a person's autonomy: the right to make choices, to hold views, and to take actions based on personal values and beliefs. Value the dignity and inherent worth of the individual.
Maximize Benefits	Provide benefits to persons and contribute to their welfare. Refers to an action done for the benefit of others. Also known as <b>beneficence</b> .
Minimize Harms	Inflict no harm intentionally. In medical ethics, the physician's guiding maxim is, "First, do no harm." Also known as <b>nonmaleficence</b> .
Justice	Be fair, give what is "owed" or "due." Treat others equitably, distribute benefits/burdens fairly across groups of individuals.

#### Contributions

- Draws on principles or pillars that are a part of American life—familiar to most people, although not by their philosophical terms.
- Provides useful and fairly specific action guidelines.
- Offers an approach that is appropriate for general bioethics and clinical ethics.
- Requires weighing and balance—flexible, responsive to particular situations.

#### Challenges

- Lacks a unifying moral theory that ties the principles together to provide guidelines.
- Principles can conflict and the theory provides no decision-making procedure to resolve these conflicts.
- Difficult to weigh and balance various principles.
- Autonomy in some cultures refers to individual autonomy, while in others it refers to group/family/community autonomy.

#### Additional Information

Additional information about ethical theories and perspectives can be found in *An Ethics Primer: Lesson Ideas and Ethics Background* by Jeanne Ting Chowning and Paula Fraser, produced through the Northwest Association for Biomedical Research. The complete *Ethics Primer* is available free for download from <http://www.NWABR.org>.

## Creating Discussion Ground Rules

### Introduction

The study of ethics involves consideration of conflicting moral choices and dilemmas about which reasonable people may disagree. Since a wide range of positions is likely to be found among students in most classrooms, it is especially important to foster a safe classroom atmosphere by creating some discussion ground rules. These ground rules are often referred to as “norms.” An agreed-upon set of ground rules should be in place before beginning this curriculum.

### Procedure

Ask the students, “What can we do to make this a safe and comfortable group for discussing issues that might be controversial or difficult? What ground rules should we set up?” Allow students some quiet reflection time, and then gather ideas from the group in a brainstorming session. One method is to ask students to generate a list of ground rules in small groups and then ask each group to share one rule until all have been listed. Clarify and consolidate the ground rules as necessary.

Post norms where they can be seen by all and revisit them often. If a discussion gets overly contentious at any time, it is helpful to stop and refer to the ground rules as a class to determine whether they have been upheld.

Some possible student ground rules/norms could include:

- A bioethics discussion is not a competition or a debate with a winner and a loser.
- Everyone will respect the different viewpoints expressed.
- If conflicts arise during discussion, they must be resolved in a manner that retains everyone’s dignity.
- Everyone has an equal voice.
- Interruptions are not allowed and no one person is allowed to dominate the discussion.
- All are responsible for following and enforcing the rules.
- Critique ideas, not people.
- Assume good intent.

### Objective

*Students will be able to:*

- Create and agree to classroom discussion norms.



# Amino Acid Abbreviations and Chemistry Resources

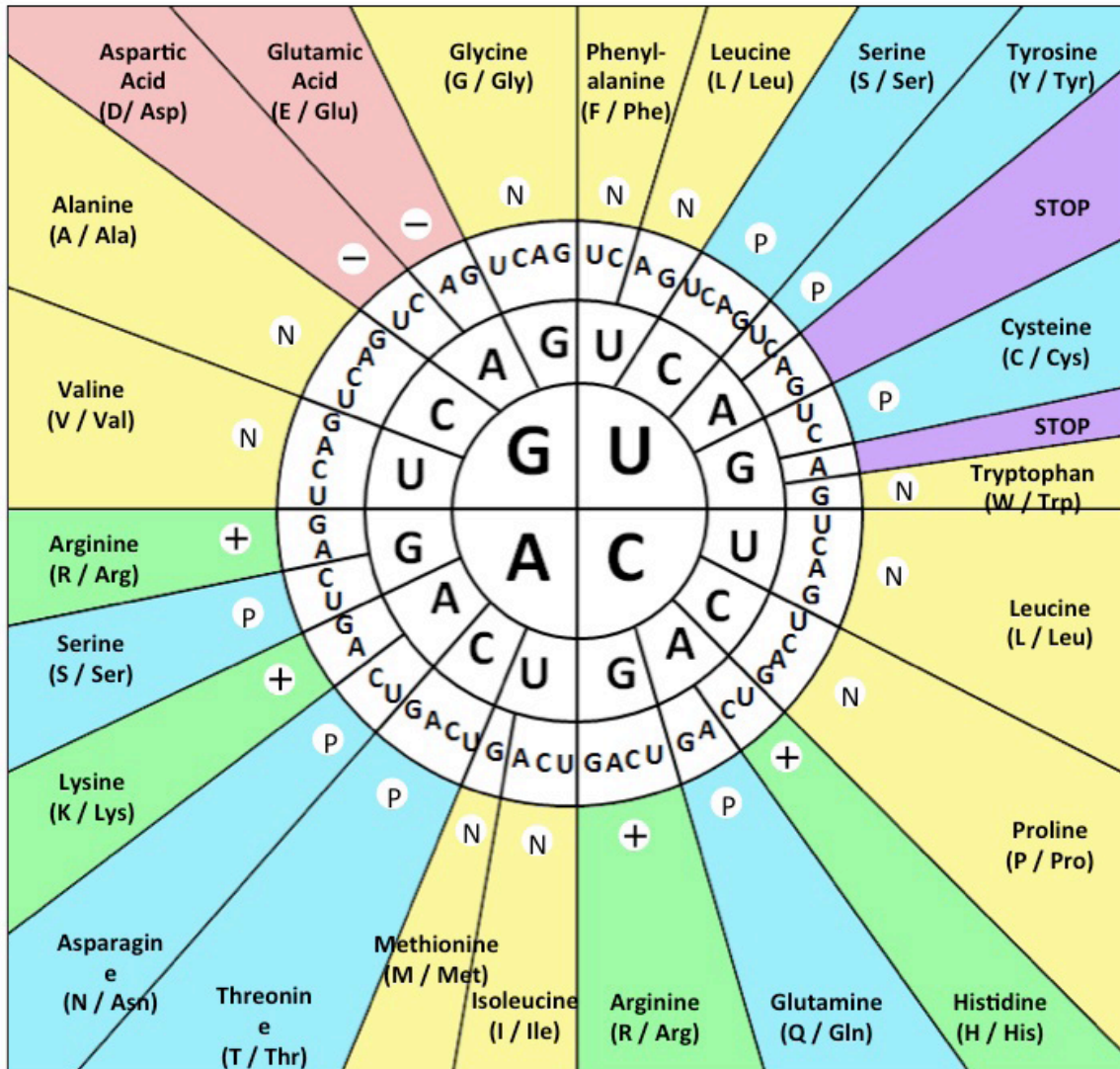
## Single-letter Amino Acid Abbreviations

- A – Alanine
- C – Cysteine
- D – Aspartic Acid
- E – Glutamic Acid
- F – Phenylalanine
- G – Glycine
- H – Histidine
- I – Isoleucine
- K – Lysine
- L – Leucine
- M – Methionine
- N – Asparagine
- P – Proline
- Q – Glutamine
- R – Arginine
- S – Serine
- T – Threonine
- V – Valine
- W – Tryptophan
- Y – Tyrosine

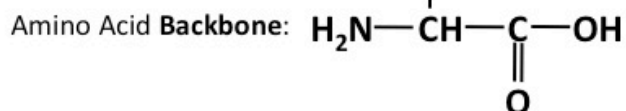
## Amino Acid Abbreviations and Categorization by Chemistry

	Uncharged	Positive	Negative
Hydrophilic	Asparagine (Asn – N) Cysteine (Cys – C) Glutamine (Gln – Q) Serine (Ser – S) Threonine (Thr – T)	Arginine (Arg – R) Histidine (His – H) Lysine (Lys – K)	Aspartic Acid (Asp – D) Glutamic Acid (Glu – E)
Hydrophobic	Alanine (Ala – A) Glycine (Gly – G) Isoleucine (Iso – I) Leucine (Leu – L) Methionine (Met – M) Phenylalanine (Phe – F) Proline (Pro – P) Tryptophan (Trp – W) Tyrosine (Tyr – Y) Valine (Val – V)		

## Codons and Amino Acid Chemistry



Amino Acid **Side Chain (R-Group):**



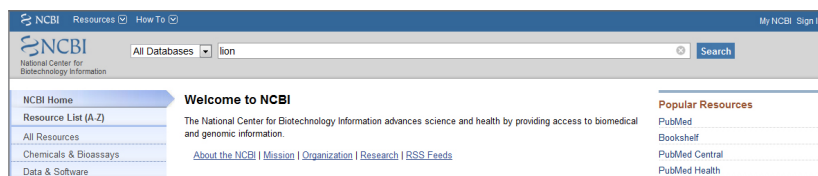
**Side Chain (R-Group) Chemistry:**

N	Nonpolar / Hydrophobic
P	Polar / Hydrophilic
—	Negative / Acidic
+	Positive / Basic
STOP	STOP

## Behind the Scenes with the NCBI Databases and the Entrez Search Engine

We have already discussed the similarity between the NCBI databases and iTunes® [in the Introductory curriculum, *Using Bioinformatics: Genetic Testing*]. Now, we're going to go a little bit farther and consider what happens when data are submitted to NCBI and when we use **Entrez** to do a database search.

When researchers submit data to the NCBI, they do so by filling in a form at the NCBI website. The sections in the form where information gets entered are called **fields**. Different data types have different kinds of fields. For example, the nucleotide database (GenBank) has fields for the **gene name**, **organism**, **sequence length**, and other information related to DNA or RNA sequences. The taxonomy database entry form includes fields for information about the **common name**, the **scientific name**, and the **rank**. Field names are used to help organize and find information.



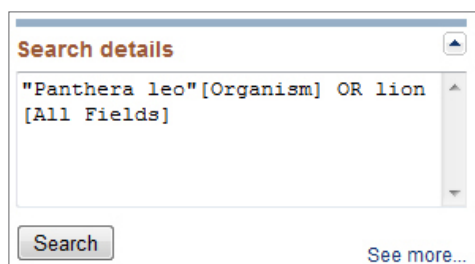
**Figure 1:** Entering terms in the NCBI Search box. Source: NCBI.

Entrez is the software system that searches NCBI databases. When you type terms into the NCBI **Search** box (**Figure 1**), Entrez takes those terms and searches all the fields, in all the database records, to see if those terms can be found. Sometimes searching with Entrez can lead to some puzzling results. For example, searching the nucleotide database with the word “lion” returns several records that come from *Sus scrofa*. *Sus scrofa* is the scientific name for “pig.” While some lions might act like pigs, their DNA sequences should be different.

To explain why searching with “lion” gives us results for pigs, we can select the link to one of the *Sus scrofa* records and look at the results. If we search the record for the word “lion,” we see that the journal is published from an address at Lion Mountain 1 Street.

What if we were searching for something from lions but instead found thousands of records from pigs? What could we do to make our search more specific and filter out the pig results?

We can get ideas for filtering by looking at the way Entrez did the search. Selecting the **Details** tab from our search results shows us that Entrez searched the organism field with the scientific name for lion (*Panthera leo*) and Entrez searched all the fields with the word “lion” (**Figure 2**).



**Figure 2:** Entrez Search Details. Source: NCBI.

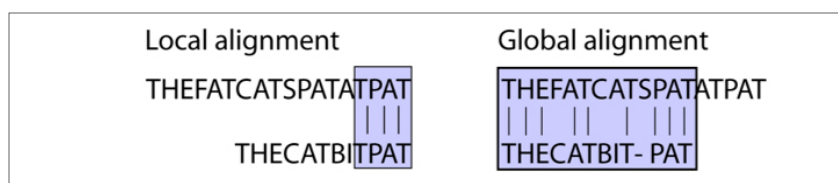
Consequently, our results included all the records where *Panthera leo* could be found in the organism field plus all the records that included “lion” anywhere in the record. We can use this information to help guide our quest for more specific results.

**Discussion:** What do you think would happen if you used “*Panthera leo*”[Organism] as a query instead of “lion?”

You can delete the words “OR lion” [All Fields] and click the **Search** button to do the experiment and find out.

## Understanding Blast

**BLAST** stands for Basic Local Alignment Search Tool. An alignment is a way of lining sequences up in rows so they are easier to compare. A local alignment is one where short regions of the sequence are aligned preferentially over long regions (**Figure 1**).



**Figure 1:** Local vs. global alignments.

Although the name “BLAST” sounds like one program, BLAST is really a family of programs that is used by biologists all over the world to compare sequences from DNA, RNA, and proteins. Nucleotide blast (**blastn**) is used to compare nucleotide sequences. Protein blast (**blastp**) is used to compare protein sequences. Other kinds of blast programs add a step in which nucleotide sequences are translated to protein sequences before searching. **Blastx** for example, translates a nucleotide query in all six reading frames and compares the predicted amino acid sequences to a protein database. **Tblast** compares a protein query to a translated nucleotide database; and **tblastx** translates both a nucleotide query sequence and the nucleotide database sequences before doing a comparison.

### How Does BLAST Work?

BLAST begins the process of comparing sequences and aligning matching regions by breaking both the query sequence (the one we added) and all the database sequences into shorter strings of text, called **words**. A typical “word” might be 11 letters long, with each letter representing a base, if the sequence is from DNA or RNA, or an amino acid, if the sequence is from a protein. Next, every word from the query sequence is compared to every word from the database until words are found that match perfectly. Once BLAST has found a word from the query that perfectly matches a database word, the program evaluates the letters at each end of the word to determine whether additional letters match. If letters on the ends match, the matching region is extended until the letters no longer match.

### BLAST Scores and Statistics

When the BLAST programs were first written in 1990, their major function was to determine whether two sequences were similar enough to infer that they most likely evolved from a common ancestor. Since the original goal for BLAST was to find matching sequences and measure the significance of the match, BLAST provides many statistics for each search and assigns different scores that can be used to evaluate the results. BLAST scores from protein comparisons are based on evolution. If a mutation occurs in a nucleic acid sequence that changes an amino acid, the altered protein experiences natural selection. If the change has a beneficial

or neutral effect, the change can persist and be inherited. If an amino acid change is harmful, negative selection will make it less likely to persist in a population. In general, amino acid replacements are tolerated better when the new amino acid is either chemically similar or located in a less important part of a protein.

When researchers wrote the scoring system for BLAST, they looked at all the changes that took place between amino acid sequences from the same protein in different organisms and used that data to calculate the log of the probability that any one amino acid would be replaced by any one of the other 19. These values are contained in the BLAST **scoring table**. To calculate the BLAST score for an alignment between two protein sequences, blastp compares the amino acids at each position, looks up the value for replacing that amino acid with another at every position, and adds all the values together.

For example, say we had this pair of aligned sequences:

E	L	V	I	S
E	L	V	E	S

The BLAST scoring table gives the value for replacing an E with an E is 5, the probability for finding an L matching an L is 4, for V replacing V is 4, for E replacing an I is -3, and for S replacing S is 4. Notice that the probability of replacing an E (Glutamic acid) with an I (Isoleucine) is negative. Amino acid changes like this, that affect the chemistry of the protein, occur less frequently and so have a negative score. To determine the BLAST score for this match, we add the scores for each position together [ 5 + 4 + 4 + (-3) + 4 ] to give a BLAST score of 14.

For nucleotide sequences, BLAST calculates a score based on identity. BLAST assigns two points for each position where a pair of nucleotides match and subtracts points for each position where they do not.

Once BLAST has calculated a score, the program applies corrections based on the size of the database and the length of the sequence to arrive at a value called the **E** or **Expect** value. The E value corresponds to the number of sequences that one would expect to find, with an equivalent number of matching residues, in a database of certain size, containing random sequences. If a BLAST result has an E value of 5, it means we would expect to find 5 sequences that align as well as our aligned query and database sequence in a random set of sequences. If a BLAST result has an E value so low that BLAST rounds it off to zero, we can conclude that the match isn't random and that the pair of sequences are indeed very similar.

### Other Applications Where BLAST is Used

Although BLAST was written with the goal of finding homologous sequences, scientists use BLAST for many other tasks. BLAST can be used to:

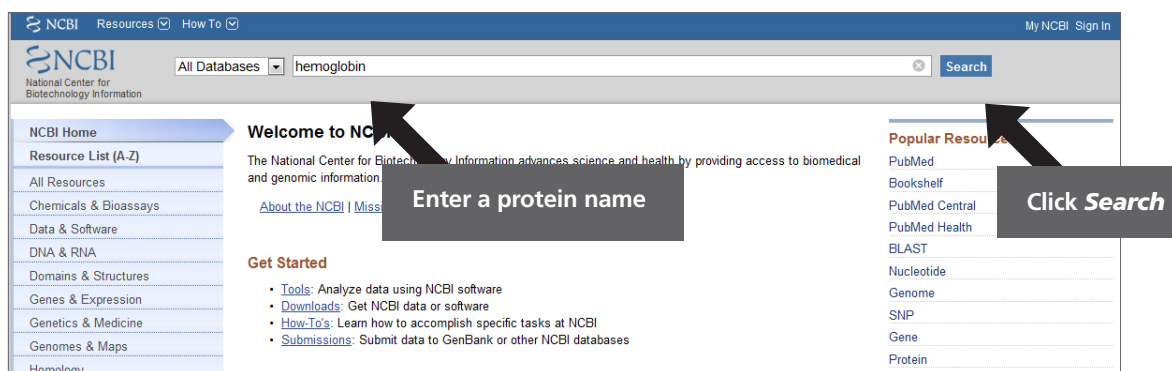
- Determine where sequences with matching regions are positioned relative to one another.
- View the relationship between mRNA and genomic DNA.
- Design and test polymerase chain reaction (PCR) primers.
- Distinguish between different species and identify genetic variation and mutation sites.

The NCBI even uses BLAST as a step in producing phylogenetic trees. Over the years, BLAST has become one of the most commonly used programs in biology.



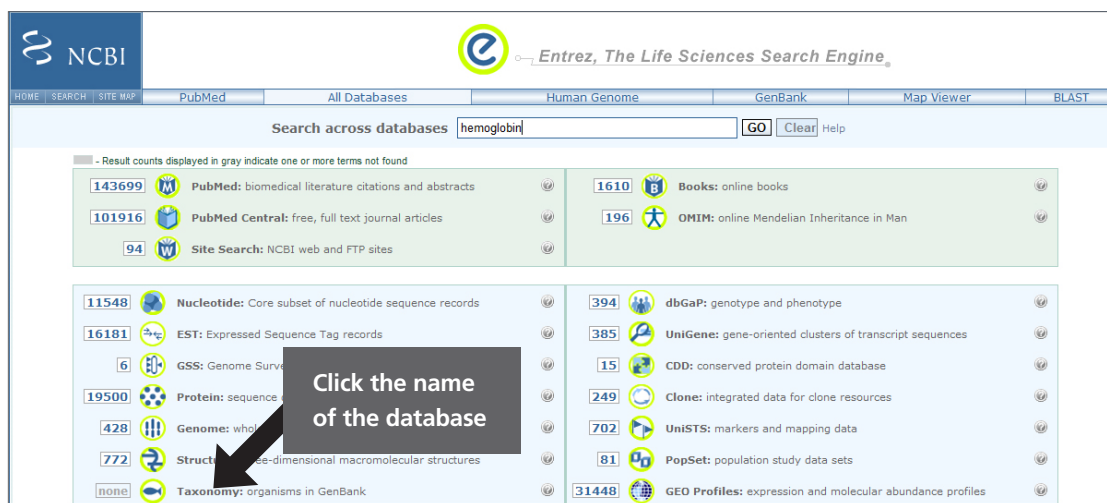
## Finding Structures in the NCBI Structure Database

1. Go to the NCBI website (<http://www.ncbi.nih.gov>).
2. Enter the name of a protein or gene in the text box and click the **Search** button (**Figure 1**).



**Figure 1:** Searching NCBI databases. Source: NCBI.

3. Searches that begin at the NCBI home page scan the contents of all the NCBI databases and provide the results on a page like the one below (**Figure 2**). The number of matching records appears next to the name for each database.



**Figure 2:** Choosing the Structure Database. Source: NCBI.

4. Click **Structure** to obtain the search results from the Structure database (**Figure 2**).

5. If you find too many search results, you may narrow the search by using the word “AND” in combination with other search terms. For example, if you wish to find structures for human proteins, you may wish to add the term “homo” to your search (**Figure 3**). “Homo” is the genus name for the human species (*Homo sapiens*).

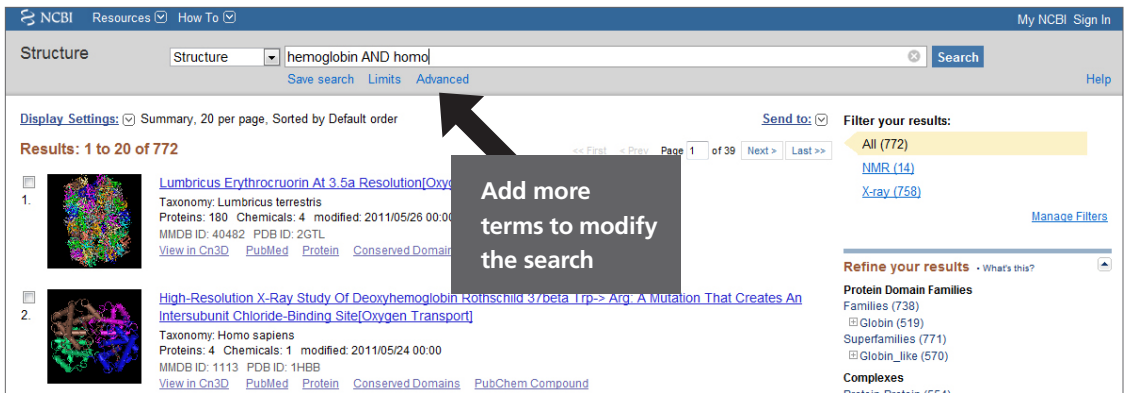


Figure 3: Using Additional Terms to Filter the Search Results. Source: NCBI.

6. The search results consist of a list of records from the database (**Figure 4**). Each record includes the title and thumbnail image of the structure. In addition, every structure contains unique identifiers that can be used to access the record, such as the **PDB ID** (Protein Data Bank Identifier). The PDB ID can be used as a search term in Entrez to retrieve a specific structure record.

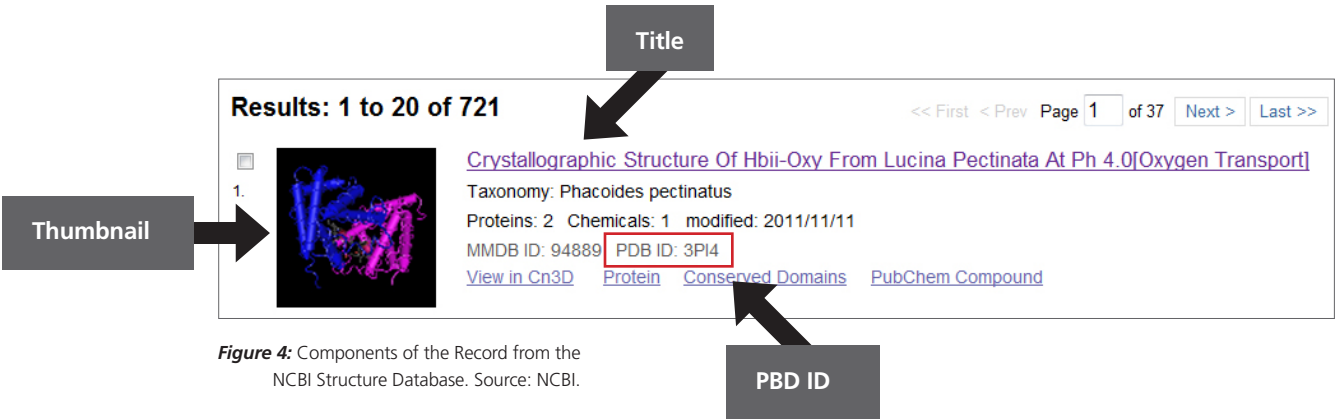


Figure 4: Components of the Record from the NCBI Structure Database. Source: NCBI.

- 7. Click the title of the structure record or the structure thumbnail image to view the structure summary page and access the download link for an individual structure.
- 8. Click the **View Structure** button to open the structure directly or to download the structure file to your computer. If you have not already done so, you can also download the Cn3D molecular viewing program from the link below the **View Structure** button.

**NCBI**  
National Center for Biotechnology Information

## Structure Summary

MMDB

PDB or MMDB ID

HOME | SEARCH | GUIDE | Structure Home | 3D Macromolecular Structures | Conserved Domains | Pubchem | BioSystems | Help

### Crystallographic Structure Of Hbii-Oxy From Lucina Pectinata At Ph 4.0

**Citation:**   
Ph-dependence crystallographic studies of the oxygen carrier hemoglobin ii from lucina pectinata.  
Nieves-Marrero CA, Ruiz-Martinez CR, Estremera-Andujar RA, Lopez-Garriga J, Garcia-Ruiz JM, Gavira JA  
No recorded citation in PubMed

MMDB ID: 94889 | PDB ID: 3P14   
PDB Deposition Date: 2010/11/5   
Updated in MMDB: 11/2011   
Experimental Method: X-Ray Diffraction   
Resolution: 3.17 Å  
Source Organism: Phacoides pectinatus   
Similar Structures:    
Inferred Interactions:

☒ First Biological Unit | ☐ All Biological Units (4) | ☐ Asymmetric Unit

Biological Unit: dimeric; determined by: author, software (PISA)

Interactions

☐ Protein ☐ Chemical

Molecular Graphic

View or Save 3D Structure

File Format:   
Display As:   
Data Set:   
  
  
**NOTICE**  
In order to view this biological unit properly, please upgrade to Cn3D 4.3.

**Figure 5:** A Structure Summary Page.  
Source: NCBI.

Click **View Structure** to download the structure.

9. If you saved the file on your computer, open it in Cn3D.