

7.003 Spring 2022 Day 19 Protein Alignment Worksheet

The instructions below outline how to perform a protein alignment using either SnapGene software or NCBI COBALT. Note that to do alignments in SnapGene, you will need to use the fully-licensed version (not just the free SnapGene Viewer). If you are unable to access the fully-licensed version of SnapGene, you may use the free online NCBI COBALT software.

Instructions for performing a multiple protein alignment (SnapGene):

- 1) From the top menubar, go to Tools → Align Sequences → Align Multiple Protein Sequences... This will open up a new alignment window.
- 2) Under the “Align individual sequences” option, go to Import Sequences To Align → Import Pasted Sequences...
- 3) Enter in the protein amino acid sequences you wish to align using FASTA format (you can copy-paste the 1-letter amino acid sequence from any SnapGene file) and click “Import.”
 - a. You can copy-paste multiple protein sequences simultaneously into the input window if they are separated with a return.
 - b. You can use the “>” FASTA header to label each sequence when you enter it for easier organization of the final alignment. For example:

```
>Human Protein X  
MTCGHWERILKKDGFHCGWE
```

```
>Mouse Protein X  
MTGGWPRIVKLEEGFHVCQAW
```

- 4) Make sure all your imported protein sequences are selected and click “Align” in the bottom right. This will produce a graphic aligning all the selected sequences.
- 5) The default setting highlights any conserved residues between the aligned protein sequences. You can adjust the alignment parameters using the control panel on the right. For instance, you can change the threshold of consensus (ranging from >50% to >95%) so that only the more highly conserved residues are highlighted.
- 6) You can save the protein alignment file (as a .praln file) in case you want to refer back to it later.

Instructions for performing a multiple protein alignment (NCBI Cobalt):

If you are unable to obtain the licensed version of SnapGene to use, you can perform protein alignments using the free online COBALT software from NCBI at

<https://www.ncbi.nlm.nih.gov/tools/cobalt/>

- 1) Enter in the protein amino acid sequences you wish to align using FASTA format (you can copy-paste the 1-letter amino acid sequence from any SnapGene file) and click “Align.”
 - a. You can copy-paste multiple protein sequences simultaneously into the input window if they are separated with a return.
 - b. You can use the “>” FASTA header to label each sequence when you enter it for easier organization of the final alignment. For example:

```
>Human Protein X  
MTCGHWERILKKDGFHCGWE
```

```
>Mouse Protein X  
MTGGWPRIVKLEEGFHVCQAW
```

- 2) The COBALT results page will show a graphical depiction of the protein sequence alignment at the top and a text alignment of the amino acid sequence at the bottom of the page.
- 3) The text alignment shows highly conserved residues in red and less conserved residues in blue. You can adjust the “Conservation Setting” to set the stringency for determining what level of conservation will color a set of residues red vs. blue vs. gray.
- 4) The COBALT alignment can be downloaded (as a FASTA file) to be viewed in SnapGene or uploaded back into COBALT if you want to refer back to it later.

MIT OpenCourseWare
<https://ocw.mit.edu/>

7.003 Applied Molecular Biology Lab
Spring 2022

For information about citing these materials or our Terms of Use, visit: <https://ocw.mit.edu/terms>.