

**MIC653: Advanced Molecular Genetics Bioinformatics & Computational Genomics**  
**Projects: Alignments of viral proteins**  
 (uses [ViroBIKE](#))

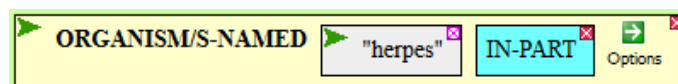
Over long periods of time, regions of a protein that are essential for function tend to remain unchanged while other regions accumulate mutations. The differences can be made plain by collecting from different organisms orthologous proteins (those related by common descent) and comparing their sequences. The comparison commonly takes the form of an alignment, where similar regions of the collected proteins are set forth one on top of the other. Alignments may also serve as the basis of the construction of phylogenetic trees, which if accurate, represent the evolutionary relatedness of the proteins.

The difficult part of making alignments is usually collecting the desired sequences. One of the major motivations of BioBIKE is to remove the pain from this step as much as possible. Freed from the burden of finding sequences and massaging them into the proper format for various tools, you can devote more of your brain to imagining what interesting things you can do with the sequences. We'll illustrate this by playing with orthologs of terminase protein from the family of herpesviruses, including human cytomegalovirus. Terminase proteins are common targets for drugs directed against certain viruses.

First, however, if you haven't used BioBIKE before, it might be a good idea to invest some time learning about how it works. You can do this through a short on-line tutorial available [here](#) or through the portal.\*

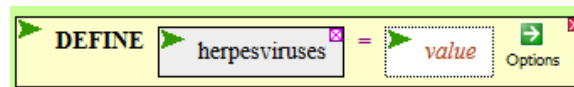
**A. Exploration of herpesviruses in ViroBIKE**

Go to [ViroBIKE](#), an instance of BioBIKE devoted to prokaryotic and eukaryotic viruses. You can see what herpesviruses ViroBIKE offers by going to the **Genome** button, **Genome Description** tab, and clicking ORGANISM/S-NAMED, thereby bringing this function into your workspace. In the *name* field, type "herpes" (with the word between double quotes, to indicate you are referring to those six letters and not the contents of a variable named "herpes"). Press Tab or Enter to close the entry field. If you execute the function now, you'll retrieve only viruses whose complete name is "herpes" (there aren't any). To specify you want matches to any part of the name, mouse over the **Option** icon and click the IN-PART option.



Now execute the function. Scroll down the popup window that will result until you find a favorite virus. I've chosen for this tour human herpesvirus 5 strain ad169 (nicknamed ad169), but you can choose whatever you like.

Since we'll be using this collection of viruses for a few operations, give it a name, perhaps:



To complete the definition, drag the ORGANISM/S-NAMED box into the *value* box and execute the function.

\* Once at the portal (<http://biobike.csbc.vcu.edu>), click the link to the *guided tours*, and choose the tour called *BioBIKE syntax and conventions*.

To get a quick idea of how similar are these viruses to each other, list their lengths. To do this, go to the **ALL** button and find on the alphabetical menu the **LENGTHS-OF** function. Once you've brought it down to the workspace, fill the *entity* box with the `herpesviruses` variable you just defined, either by bringing it down from the **Variables** menu or just typing it in. Then execute the function and take in the range of virus sizes.

## B. Exploration of the proteins of herpesviruses

What kind of genes does ad169 have? You can list them by executing the **DESCRIPTION-OF** function, available either from the alphabetical list (**ALL**) or from the **Genes-Proteins** menu. Click the *entity* box and fill it with the **GENES-OF** function (from the **Genome** menu). Finally, fill in its *entity* box with ad169 and execute the completed function.

Unless you are a keen student of herpesviruses, the list won't be very edifying. On the other hand, if you're an expert, you may recognize that UL89 is the name of terminase protein. Verify this by clicking the name of the gene (`hhv5gp084`) next to UL89. That will bring you to a page that describes the gene. Mousing over the Note field reveals that the gene indeed encodes a subunit of terminase. There's a way of finding this gene without prior knowledge about UL89, but I won't go into that here.

What similar proteins are in other herpesviruses? To find out, bring down the **SEQUENCE-SIMILAR-TO** function and type in the name of the terminase gene from ad169. Executing the function now will yield similar proteins from all viruses that have them. To confine the search to herpesviruses, use the **IN** option (from the **Option**) menu and put `herpesviruses` in the *value* box. Specify that the search is Protein-vs-Protein by selecting that option. Then execute the function.

You'll see hits that range from perfect to not so good. You can get a more selective set with the **ORTHOLOG/S-OF** function, which proteins that are bidirectional best hits of the protein you put in the *gene-or-protein* box. You might want to **DEFINE** a variable that contains this set of putative orthologs.

To find out what herpesvirus proteins are most conserved, you can compare all of them at once, by putting **PROTEINS-OF** ad169 as the *query*.

## C. Align terminase sequences

Now that you have in hand a set of orthologous proteins, align their sequences by bringing down the **ALIGNMENT-OF** function and putting in the *sequence-list* box the variable you defined to contain the orthologs. Executing the function gives you the alignment.\* However, the names of the proteins will probably not be recognizable. You can label the proteins by the names of their viruses by using the **LABEL-WITH-ORGANISM** option.

The alignment you produced can be used as the basis of a quick phylogenetic tree. To produce one, bring down the **TREE-OF** function and drag the alignment in the Result pane into the *alignment* box.

## D. Play!

---

\* Alignments are time consuming. If it runs over 40 seconds (which it might), you'll run past the default time limit. To give yourself more time, mouse over the black **Session** button and click Preferences.