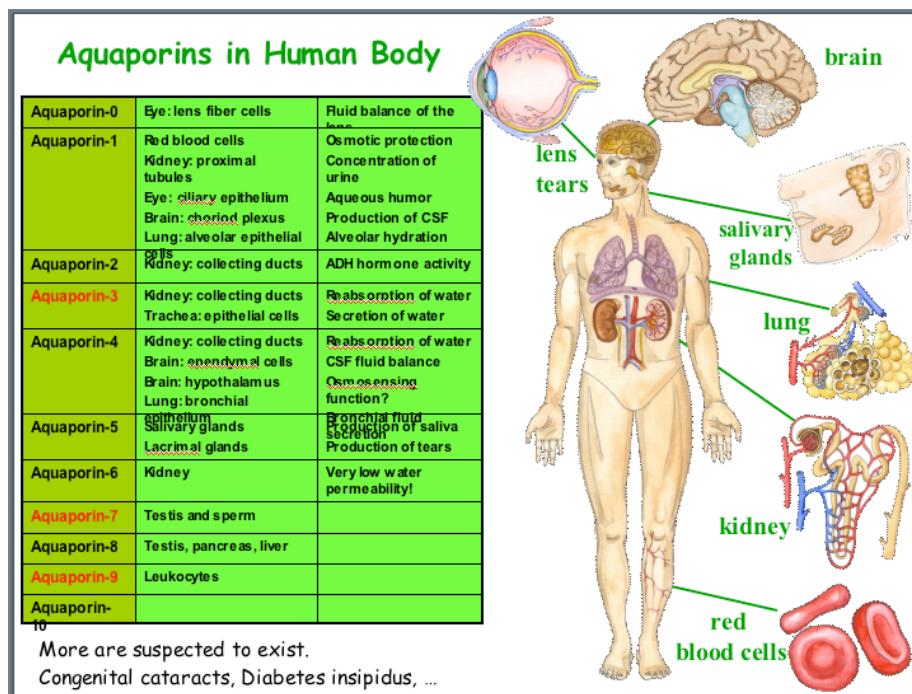


Aquaporins



VMD Developers:

John Stone

Dan Wright

John Eargle

Fatemeh Khalili

Elizabeth Villa

Emad Tajkhorshid

Brijet Dhaliwal

Zan Luthey-Schulten

May 26, 2004 (9h 40min)

Contents

1	Introduction to Aquaporin Structure	5
1.1	Structure and Function of Aquaporins	5
1.2	Loading other aquaporin structures	11
2	Structural Alignment of Aquaporins	13
2.1	Starting Multiple Sequence Alignment	13
2.2	Setting parameters for Multiple Structure Alignments	13
2.3	Aligning the molecules	14
3	Comparing Protein Sequence and Structure	16
3.1	Protein Structure	16
3.2	Protein Sequence	17
4	Bulk Residue Selection	19
4.1	Structure conservation	19
4.2	Sequence conservation	21
5	Investigating Structural Alignment	24
5.1	RMSD per Residue	24
5.2	Pairwise RMSD between the molecules	27
6	Examining the Aquaporin Tetramer	28
6.1	Loading Tetramer	28
6.2	Examining channel lining and tetrameric contacts	28
7	Phylogenetic Tree	30
8	Exporting File Formats	31
8.1	Exporting a PDB file from user-specified selections	31
8.2	Exporting a multiple alignment in FASTA format	32

Introduction

The MultiSeq extension to VMD

The MultiSeq extension to VMD links protein structures to protein sequences and permits comparison of proteins in terms of structure and sequence. MultiSeq is designed to study protein mechanism through available genetic information, e.g., evolutionary conservation, offering biomedical researchers a new tool to examine protein structure and function. This tutorial can be used by both new and previous users of VMD. However, it is recommended that new users go through the "VMD Molecular Graphics" tutorial, in order to gain further knowledge about the overall program¹

The present tutorial has been designed specifically for VMD with MultiSeq and should take about an hour to complete in its entirety.

Aquaporins

We will use the family of aquaporins for a case study in the applications of the MultiSeq tool. Aquaporins are membrane channel proteins found in a wide range of species, from bacteria to plants, to human. They facilitate rapid transport of water across the cell membrane playing an important role in the control of cell volume and transcellular water traffic.

In human there are at least 11 different aquaporin types (see cover). The kidney alone contains AQP1, AQP2 and AQP3, using these proteins to filter hundreds of liter of water a day. In addition to water transport some members of the aquaporin family also transport selectively small molecules such as glycerol.

Here we use MultiSeq a tool in VMD, to conduct a comparative study of the structure and sequence of four aquaporins from different species: human AQP1, bovine AQP1, AqpZ from *E.coli*, and GlpF (*E.coli* glycerol facilitator).

¹URL: <http://www.ks.uiuc.edu/Training/SumSchool/materials/sources/tutorials/01-vmd-tutorial/html/vmd-tutorial.html>

Getting Started

MultiSeq, in its current release, is operable on the following platforms:

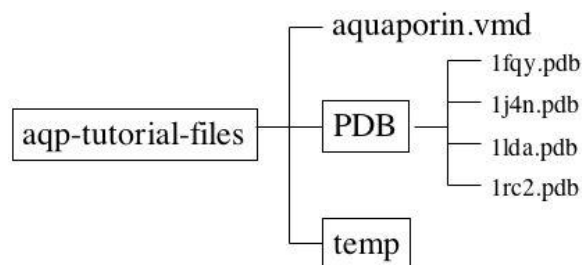
- Macintosh OS X
- Solaris
- Linux

This tutorial requires certain files, available in compressed format (tar.gz format) for download at:

<http://www.ks.uiuc.edu/Proposals/NCBC2004/Supplement/> .

- 1 Download the file `aquaporin.tar.gz`.
- 2 Once the zip file has been downloaded to your desktop, you need to unzip the multiseq package. On a Mac OS X this is done simply by double-clicking on `aquaporin.tar.gz`. If you are using VMD on a UNIX or Linux machine, you will need to unzip the file using the appropriate command(s) at a prompt.

Once you decompress the file, you will have a directory called `tutorial_aqp_files`. In this directory you will find all the files needed for this tutorial. The file structure of this directory is the following:



- 3 To start using the tutorial, just start VMD by typing `vmd` at a UNIX terminal, or double-click on the VMD application icon in a Macintosh.

1 Introduction to Aquaporin Structure

In this unit you will use the conventional molecular graphics tools of VMD to become familiar with the key structural features of aquaporins, and how these are related to aquaporin function.

1.1 Structure and Function of Aquaporins

You will first consider bovine aquaporin (PDB code 1J4N).

In order to learn about the structural features of aquaporins, you will create several graphical representations of the molecule. Here, we will only guide you through the steps to create the necessary representations. If you want to know more about graphical representations, please look at the main VMD tutorial².



Using a VMD saved state. If you are familiar with VMD and don't want to go through all the steps of creating representations, you can use the VMD saved state `aquaporin.vmd` located on your tutorial directory to load the molecule with all the necessary representation. If you choose to use the saved state, you still need to read through this section to learn about important structural features of aquaporins, that will be referred to in the remainder of the tutorial. We also note that in case you need to interrupt the tutorial, you can save the current state of the VMD session (see VMD tutorial) and resume your study at a later time.

- 1 Choose the File → New Molecule... menu item in the VMD Main window. Another window, the Molecule File Browser, will appear on your screen.
- 2 Use the Browse... button to find the file `1j4n.pdb` in the directory `aqp_tutorial_files` → PDB in the tutorial directory. Note that when you select the file, you will be back in the Molecule File Browser window. In order to actually load the file you have to press Load. Do not forget to do this!



Getting a PDB file over the Internet. In case your computer is connected to the Internet, you can use VMD to get the file `1j4n.pdb` automatically from the Protein Data Bank³(see VMD tutorial).

Now, the aquaporin is shown on your screen in the OpenGL Display window. You may close the Molecule File Browser window at any time.

²URL: <http://www.ks.uiuc.edu/Training/SumSchool/materials/sources/tutorials/01-vmd-tutorial/html/vmd-tutorial.html>

³URL: <http://www.pdb.org>

- 3 Choose the Graphics → Representations... menu item. A window called Graphical Representations will appear and you will see the current graphical representation used to display your molecule highlighted in yellow.
- 4 In the Draw Style tab you can change the style and color of the representation. In the Selected Atoms text entry of the Graphical Representations window delete the word `all`, type `protein` and press the Apply button or hit the Enter key (do this every time you type something). From Drawing Method, choose the `tube` menu item. The representation draws a tube along the backbone of the protein (Fig. 1). Now, change the color of the molecule by choosing the Molecule menu item from the Coloring Method menu.

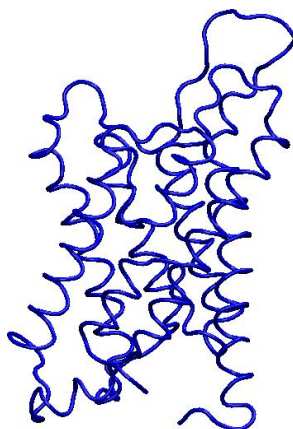


Figure 1: Tube representation of aquaporin structure.

The tube representation you created will be used throughout the tutorial to look at the structural alignment of aquaporins, but before you start using MultiSeq, you will create other representations in order to learn some of the important structural features of aquaporins.

- 5 In the Graphical Representations window, press the Create Rep button. This will create a new representation, identical to the one you had before. This time, choose the drawing method `Cartoon`. This drawing method shows the secondary structure of aquaporins (Figure 2).
- 6 In the Graphical Representations window, double click on the first representation. This will make it appear in light color in this menu, and will hide the corresponding representation in the OpenGL window.

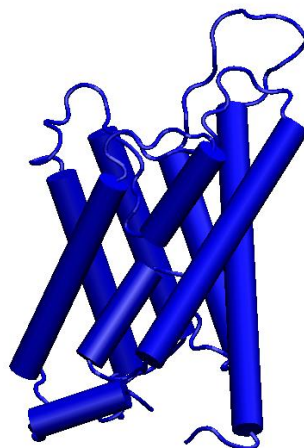


Figure 2: Cartoon representation of aquaporin structure.



Secondary Structure of Aquaporins. Aquaporins are composed of six transmembrane α -helices and two reentrant loops, formed by a short helix and an extended polypeptide that returns to the surface of the protein. The two reentrant loops, which are surrounded by the six long helices, meet each other in the middle of the channel. These loops line the channel and are of great functional importance.

- 7 Click on the OpenGL window and use your mouse to rotate the molecule and look at the structure of aquaporins.

For the rest of this section, you will continue to look at the aquaporin structure in tube representation. You can always go back and look at the cartoon representation if you need.

- 8 In the Graphical Representations window, double click on the first representation. This will show it again in the OpenGL window. Double click on the second representation (cartoon) to hide it.

You will now look closer at the structural details of the aquaporin structural elements.

- 9 Create a new representation as explained above. This time, you will focus on the helices of the reentrant loops mentioned above (see box). In the Selected Atoms text entry of the Graphical Representations window type `resid 79 to 88 196 to 204`. These residues correspond to two helices of aquaporin that face each other in the middle of the channel. Choose the coloring method `ColorID → 1` and the drawing method `tube`.

- 10 Repeat step 9. This time, use for selected atoms `resid 74 to 78 191 to 195`, and for coloring method `ColorID → 7`. These residues correspond to the extended polypeptide regions of the reentrant loops of aquaporins (see box).

Now that you have localized the reentrant loops, you will dim the rest of the protein.

- 11 Click (once) on the first representation. In the **Material** menu, choose **Transparent**. Your OpenGL window should look like Fig. 3.

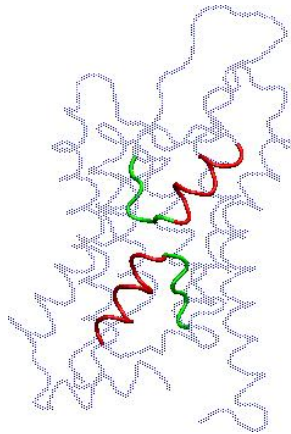


Figure 3: The reentrant loops are a key structural feature of aquaporins. Each reentrant loop is formed by a short helix, shown in red, and an extended polypeptide, shown in green.

Now that you can locate the reentrant loops, you will look at a site of particular interest inside the reentrant loops, the NPA motif.



NPA motif. The NPA motifs, as the name implies, are formed by amino acids NPA, i.e., asparagine, proline and alanine. They are highly conserved signature motifs in all aquaporins. They stabilize the reentrant loops through multiple hydrogen bonds, as shown in figure 4.

- 12 Create a new representation. In the **Selected Atoms** text entry of the **Graphical Representations** window, type `resid 78 to 80 194 to 196`. These residues correspond to the two NPA motifs present in aquaporin. Choose the drawing method **Licorice** and coloring method **Type**. Look at the NPA motifs, can you see how they would be hydrogen bonded?

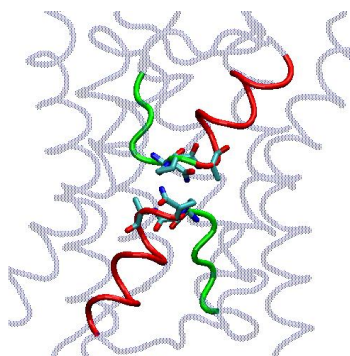


Figure 4: The NPA motifs in aquaporins.

- 13** Create another representation by pressing the Create Rep. With the same coloring and drawing methods of the previous selection, type `resid 197` in the Selected Atoms text window. This will draw an important arginine residue in aquaporins.

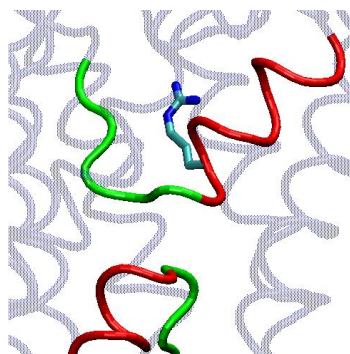


Figure 5: The conserved arginine provides H-bond sites to water molecules inside the channel.



Conserved arginine. This arginine is the only charged residue in the channel, and is conserved in all aquaporins. This arginine lines the pore and provides additional H-bond sites for water molecules permeating the channel (See Fig. 5).

- 14** Repeat step 13, but this time type `resname GLU` in the Selected Atoms text window. This will select all the glutamates in the protein. Notice that there are only two glutamate side chains in the transmembrane region, i.e., in the helical core of the protein. If you look closely (c.f. Fig. 6) you will see that each one of these glutamates sits behind a reentrant loop.

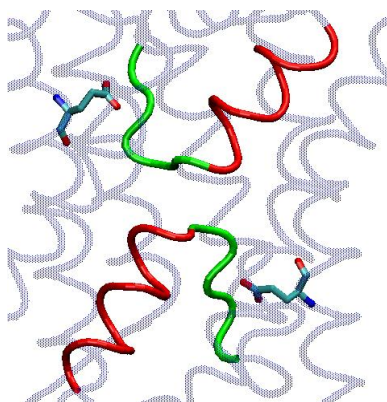
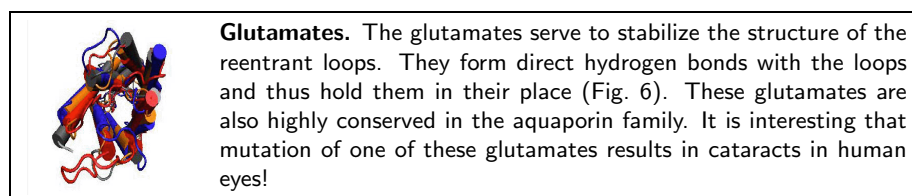


Figure 6: The glutamates help stabilize the reentrant loops.

Finally, you are going to look into the aquaporin pore.

- 15 Create a representation. In the Selected Atoms text window type `protein`. Set the drawing method to VDW (van der Waals). Each atom is now represented by a sphere. In this way you can view the volumetric distribution of the protein.
- 16 Rotate the molecule so that you can see it from the top. Look down the pore. Can you see the pore? How many water molecules you think fit there?

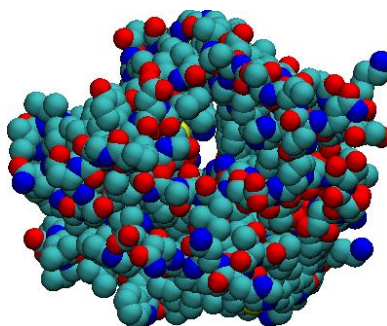


Figure 7: The pore of aquaporins can only accommodate water molecules in single file.



Pore. The pore is so narrow (Fig. 7) that it can fit water molecules only in single file. This feature is very important for the selectivity of the channel.

- 17** Create a new representation. In the Selected Atoms text window type **water**. Set the drawing method to VDW and the coloring method to **type**. This will create a representation with all the water molecules the crystallographers resolved in the crystal structure. You can now see that the pore is filled with water molecules forming a single file.

Now that you have learned about the structural features of aquaporins, you will load the other three aquaporins mentioned above and get ready to align them. Before that, you should turn off all representations created so far, leaving only the first. *Note that the coloring tools in MultiSeq work only on the first representation for each molecule.*

- 18** In the Graphical Representations window, double click on each of the graphical representations, except the first. This will make them appear in light color in this menu, and will hide the corresponding views in the OpenGL window. Select the first representation and set the material to Opaque.

1.2 Loading other aquaporin structures

In this tutorial you will be comparing four aquaporin structures listed in Table 1.2. You have already loaded the **1j4n** molecule. Now, you will load the other three aquaporins.

PDB code	Description
1j4n	Bovine AQP1
1fqy	Human AQP1
1lda	E. coli Glycerol Facilitator (GlpF)
1rc2	E. coli AqpZ

Table 1: Aquaporin structures

- 1** For each of the remaining aquaporins, i.e., **1fqy**, **1lda**, **1rc2**, repeat steps 1 through 4 of the previous section.

You should now see the four molecules loaded in the OpenGL Display (Figure 8). Note that the structures are not aligned, but rather placed according to the coordinates specified in their PDB files.

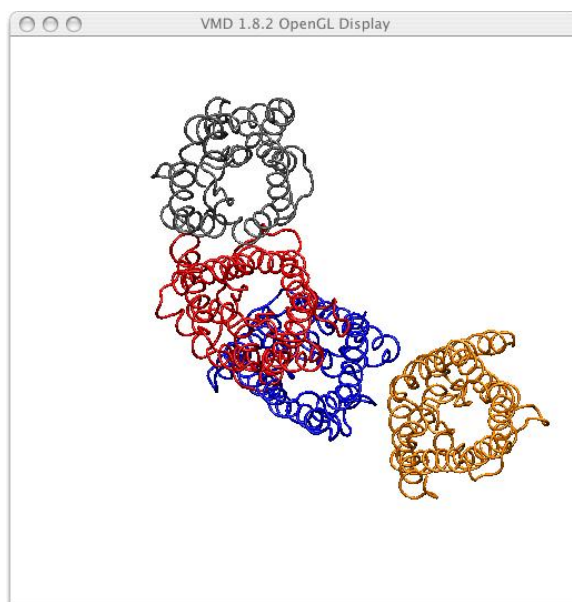


Figure 8: Aquaporin structures loaded into VMD.

2 Close the Graphical Representations window.



1rc2. The original 1rc2 file contains two copies of the aquaporin molecule. This is the way the crystallographers reported this structure to the PDB since the crystal used for analysis contained two independent copies of AqpZ. For this tutorial, we include only one copy of the molecule in the provided pdb file.

You are now ready to start using the Multiple Sequence Alignment tool, referred to above and below as the MultiSeqtool.

2 Structural Alignment of Aquaporins

In this section, you will align the aquaporin molecules loaded in VMD using the *MultiSeq* tool.

2.1 Starting Multiple Sequence Alignment

Now that you have loaded the four proteins in Table 1.2, you will use the Multiple Sequence Alignment tool to align them and analyze their structural relationships.

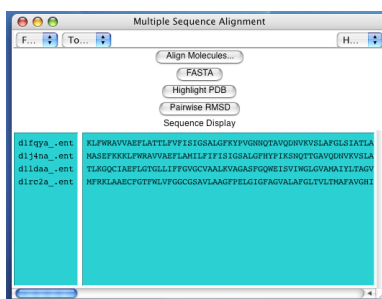


Figure 9: Multiple Sequence Alignment Window

- 1 Within the main menu window of VMD, choose the Extensions → multiseq menu. A window entitled Multiple Sequence Alignment (Figure 9), will appear on your screen. This is the main MultiSeq extension window. The rest of the tutorial and exercises will use features from this window, unless specified otherwise.

2.2 Setting parameters for Multiple Structure Alignments

MultiSeq uses the program STAMP to align protein molecules.



STAMP. The program STAMP (Structural Alignment of Multiple Proteins) is a tool for aligning protein sequences based on a three-dimensional structure. It utilizes the Smith Waterman algorithm for this process. The alignment involves the C_{α} atoms of the molecules. The alignments should be performed between proteins which are structurally similar. STAMP was developed by Robert B. Russell and Geoffrey J. Barton (PROTEINS, v14, 309-323, 1992). Information on STAMP can be found at <http://www.rfcgr.mrc.ac.uk/Registered/Help/stamp/stamp.html>

Before you align the molecules, you need to set certain parameters needed by STAMP.

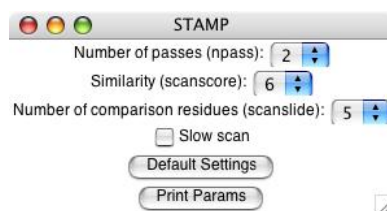


Figure 10: STAMP parameter window

- 1 Choose the Tools → STAMP parameters tool. The window shows four parameters that can be set (Figure 10). For this tutorial, you will use the default parameters.
- 2 Close the STAMP parameters window.

You can look at the STAMP user guide (see box) for information on how to optimize the STAMP parameters to obtain a better alignment.

2.3 Aligning the molecules

Now that you have opened MultiSeq and made sure STAMP parameters are correct, you can align the molecules you loaded into VMD.

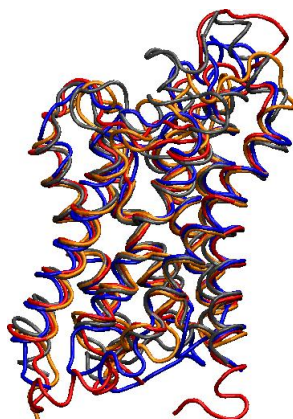


Figure 11: Structural alignment of four aquaporin molecules.

- 1 Click on the Align Molecules... button. A file browser window will pop up prompting you to select a temporary directory to store STAMP output files.
- 2 The tutorial package contains a directory called `temp`. Select this directory and click the Choose button. The file browser window will automatically close.

The molecules have been aligned. You can see this both in the OpenGL window (Fig. 11) and in the MultiSeq window (Fig. 12). Take some time to look at the alignments.

- 3 Click on the OpenGL window (Fig. 11). Move your mouse around, rotating the molecules. Do you think this is a good alignment? Look at the top of the molecules: Can you see the pore?

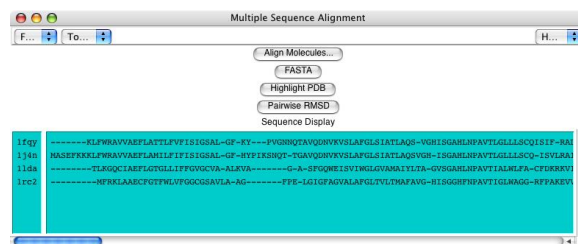


Figure 12: Alignment of four aquaporin molecules based on sequence similarity.

- 4 Now, click on the Multiple Sequence Alignment window. Using the slider at the bottom, take a look at the alignment. Observe that there are gaps in the sequences when they are aligned, so that they fit structurally.


In the next section, you will start using the MultiSeq features for the alignment you just made.

3 Comparing Protein Sequence and Structure

In this unit you will study the structural conservation of the different aquaporins and investigate their sequence similarity.

3.1 Protein Structure

In order to better understand the structural conservation within the aligned molecules, MultiSeq provides tools that allow you to color the molecules according to their Q value, a measure of structural conservation.



Q per residue. Q is a parameter that indicates *structural identity*. Q accounts for the fraction of similar native contacts between the aligned residues in two proteins⁴. Q=1 implies that structures are identical. When Q has a low score (0.1-0.3), structures are not aligned well, i.e., only a small fraction of the C_{α} atoms superimpose. Q per residue is the contribution from each residue to the overall Q value of aligned structures.

You will now color the molecules according to the value of Q per residue obtained in the alignment.

- 1 In the Multiple Sequence Alignment window, choose the Tools → Molecule Coloring menu item.

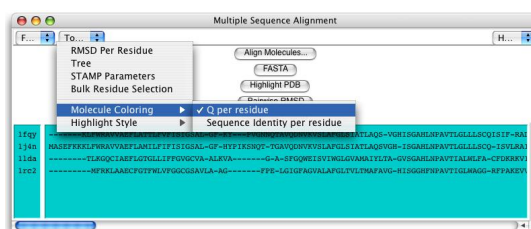


Figure 13: Molecule Coloring

- 2 Select Q per residue to visualize structural similarity.
- 3 Look at the OpenGL window to see the impact this selection has made on the coloring of the molecular graphics (Figure 14).

Notice that the transmembrane helices of the aligned molecules have turned blue. Rotate the molecule to see how much of it has turned blue. The blue areas indicate that the molecules are structurally similar at those points. If

⁴Eastwood, M.P., C. Hardin, Z. Luthey-Schulten, and P.G. Wolynes. "Evaluating the protein structure-prediction schemes using energy landscape theory." IBM J. Res. Dev. 45: 475-497, 2001. URL: <http://www.research.ibm.com/journal/rd/453/eastwood.pdf>

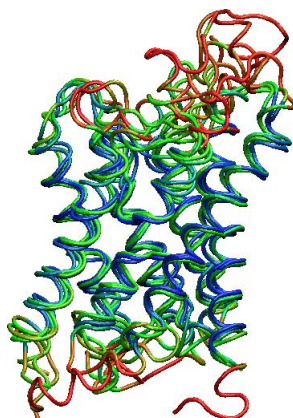


Figure 14: Molecules colored by Q value

there is no correspondence in structural proximities at these points, the points appear red. The red, i.e., structurally least similar segments, tend to be on the periphery of the aligned molecule. Note that the loops tend to be red, while the helices are blue, i.e., highly conserved.

3.2 Protein Sequence

You have examined the structural similarity between the molecules. Now you will look at the sequence similarity, i.e., when at a certain position of the aligned molecules, the amino acid of each molecule is the same. One usually refers to such amino acids as “conserved” amino acid, a nomenclature we will employ in the following. MultiSeq has a feature to color the molecules according to how much the sequence is conserved within the molecules aligned. This tool, Sequence Identity per Residue, colors each amino acid according to the degree of conservation within the alignment: blue means highly conserved and red means very low or no conservation.

- 1 Choose the Tools → Molecule Coloring → Sequence identity per residue coloring method.

Before you look at the viewer window, can you anticipate what will happen to the coloring of the molecules? Will the molecules still be blue in the transmembrane region, as they were when Q per residue was used to determine structure conservation?

- 2 Now take a look at the viewer window. As you can see (Fig. 15), a fair portion of the molecules has turned red, indicating less sequence conservation than structural similarity.

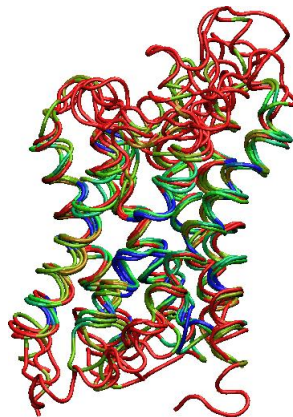


Figure 15: Molecules colored by sequence identity

- 3** Note that the residues at the site where the two short helices encounter are blue. Look at the molecules from the top (c.f. Fig. 16). Can you notice the blue residues tend to be on the inside of the pore?

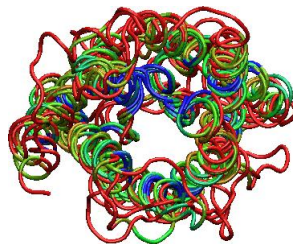


Figure 16: Conserved residues (shown in blue) are located inside the pore.

Note that the coloring of the molecules using Sequence Identity per Residue indicates that the sequence conservation is much less in comparison to the structural conservation. Sometimes structures from the same family have less than 10% of sequence identity, yet are structurally similar. A well known example is the protein myoglobin that we recommend for self-study.

To examine the relationship between sequence and structure in more detail, in the next section you will use the Bulk Residue Selection feature.

4 Bulk Residue Selection

In this section, you will use the Bulk Residue Selection tool for finding structural features of aquaporins that have been conserved throughout species, and are crucial for its function.

The Bulk Residue Selection feature permits a more precise analysis of structural similarity and sequence conservation. It makes use of different measures to highlight residues in the Sequence Display and Structure Display simultaneously. Bulk Residue Selection allows you to examine the conservation and similarity on a per residue basis, using two different measures, Q measuring structure similarity and Sequence Identity measuring sequence conservation.

4.1 Structure conservation

You can use MultiSeq to look at the residues that are structurally similar above or below a certain value of Q.

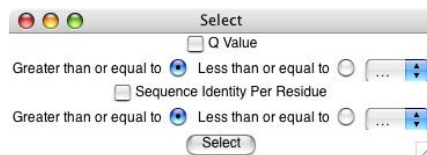


Figure 17: Bulk Residue Selection

- 1 Choose the Tools → Bulk Residue Selection tool. A new window entitled Bulk Residue Selection will appear (Fig. 17). Within the window you will notice two measures: Q value and Sequence Identity per Residue.
- 2 Select the Q value option, and greater or equal to → 0.5.
- 3 Click on the Select button.

Now new views will appear, one in the Open GL (structure) Display window (Fig. 18, left) and one in the Sequence Display window (Fig. 19).

- 4 Look in the Sequence Display window (Fig. 19). Since you selected the Q value to be greater or equal to 0.5, most of the structural elements of the aligned molecules will be shown as conserved (similar), and will be displayed in yellow.
- 5 Now, look at the OpenGL Display window (Fig. 18, left). In this case also the most parts of the molecules are highlighted, in particular, the transmembrane segments.

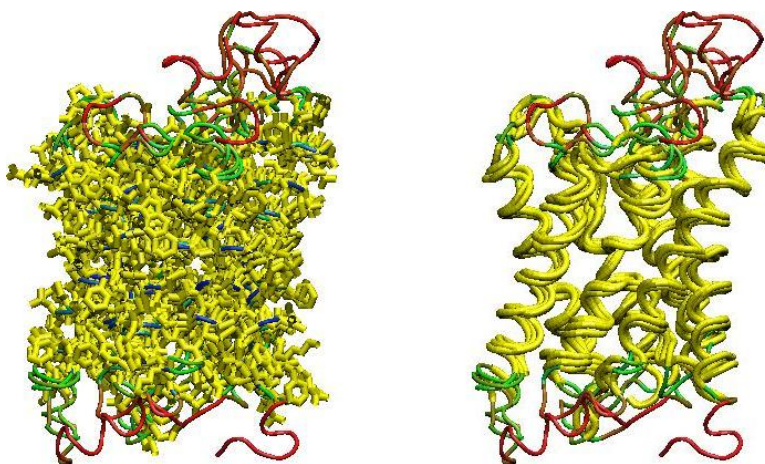


Figure 18: View of the bulk residue selection for $Q=0.5$ in the Open GL Display window. Highlight style set to bond (left) and tube (right)

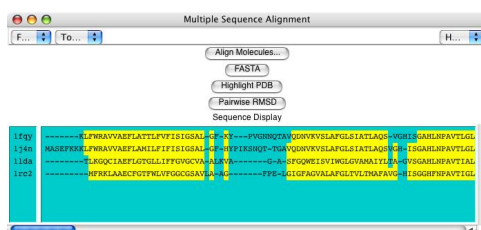


Figure 19: View of the bulk residue selection for $Q=0.5$ in the Sequence Display window.

- 6 Due to the high level of conservation, the Bond highlighting style is not very informative. To improve the view, chose the Tools \rightarrow Highlight Style menu item, and click on the Tube style (Fig. 18, right). You can recognize immediately that high Q values indicating conservation of neighborhood relationships for amino acids are seen in the transmembrane region throughout.

Now you will investigate what happens for higher values of Q .

- 7 Inside the Bulk Residue Selection window (Fig. 17), select the Q value option, and **greater or equal to** $\rightarrow 0.7$. Click on the Select button.
- 8 Look at the OpenGL Display window (Fig. 20). You notice that most of the structurally similar regions are still in the transmembrane region, but the highlighted regions are more confined than before.
- 9 Look in the Sequence Display window (Fig. 21). You can still distinguish blocks of structurally similar residues, which correspond to the helices.

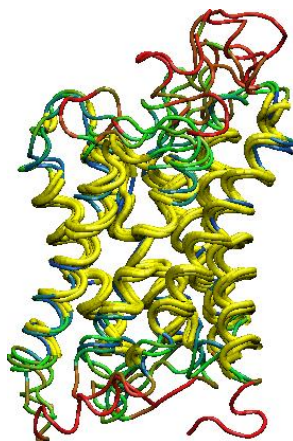


Figure 20: View of the bulk residue selection for $Q=0.7$ in the OpenGL Display window

They are not as well defined as before, but the structural similarity is still noticeable when probing for high values of Q .

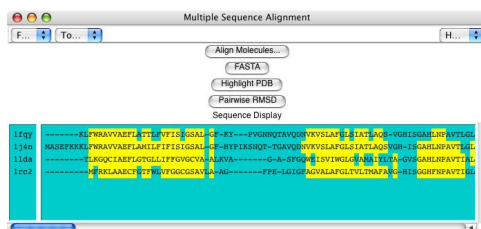


Figure 21: View of the bulk residue selection for $Q=0.7$ in the Sequence Display window

- 10 Repeat steps 7 to 9 and set the value of Q to **greater or equal to** \rightarrow 0.9. You will notice that almost all the sequence conservation is lost.

What does this mean? The aligned aquaporins have a moderate to high structure conservation. Structurally, the most similar areas are the transmembrane helices.

4.2 Sequence conservation

You can also use MultiSeq to look at the residues that are conserved on the sequence level.

- 1 Inside the Bulk Residue Selection window (Fig. 17), de-select the Q value option and select the Sequence Identity per Residue option.

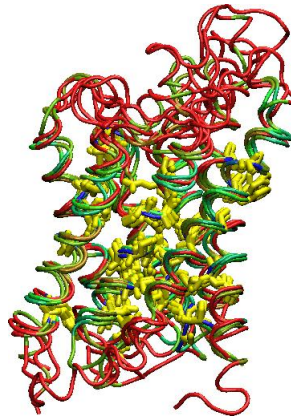


Figure 22: View of the bulk residue selection by Sequence Identity in the OpenGL Display window.

- 2 Below Sequence Identity per Residue you can select Greater than or equal to or Less than or equal to. Select Greater than or equal to.
- 3 To the right, you can select a value between 0.0 to 1.0. Select 0.7 for a value. This will then select residues that were assigned a sequence identity measure of 0.7 or more.
- 4 Click on the Select button.

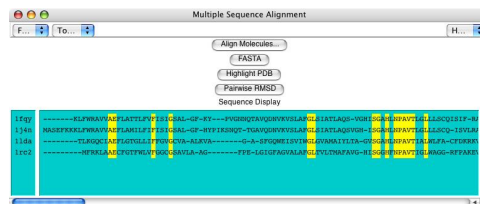


Figure 23: View of the bulk residue selection by Sequence Identity in the Sequence Display window

- 5 Look at the Multiple Sequence Alignment window. You will notice very few residues are selected. If you look carefully, you will locate some of the residues that correspond to key structural features of aquaporins, for which you made representations in section 1 of this tutorial. Search for the NPA motifs, highlighted in yellow (Fig. 23).
- 6 In the OpenGL window, look at the areas highlighted in yellow. Notice that many residues in the reentrant loops (Fig. 3) conserved their sequence over the course of evolution. This is a clear indication that the reentrant loops are functionally significant.

The conserved amino acids in the sequence are mostly located in the pore, and in the reentrant loops. You will now compare these residues to the representations you created in section 1 of this tutorial, which correspond to residues that are relevant to the function of aquaporins.


- 7 Change the Highlight Style to Licorice.
- 8 In the Graphical Representations window, in the Selected Molecules box, select the molecule 1j4n.
- 9 In the Representations window, double click on the representations you created for the NPA motif, the glutamates and the arginine. What is the sequence identity for these amino acids? Why do you think the NPA motif, the lysine and the glutamates are conserved throughout species?

5 Investigating Structural Alignment

In this section you will use RMSD values between molecules as a measure to see how close structures of molecules are after alignment.

5.1 RMSD per Residue

RMSD per residue values compare two proteins and show how well they align. Here we use RMSD values that compare solely the C_α atoms of the proteins.



RMSD values. The Root Mean Squared Deviation (RMSD) is a numerical measure of the difference between two structures. It is defined as :

$$RMSD = \sqrt{\frac{\sum_{i=1}^{N_{atoms}} (r_i(1) - r_i(2))^2}{N_{atoms}}} \quad (1)$$

where N_{atoms} is the number of atoms whose positions are being compared, and $r_i(1)$ and $r_i(2)$ are the position of atom i in each molecule.

- 1 In the MultiSeq window choose Tools → RMSD per residue. A window RMSD Dialog with the names of the molecules will appear on your screen. (Fig. 24)
- 2 Highlight the first and second molecules (0, 1) by clicking on them.

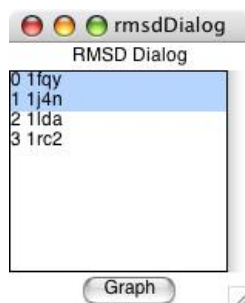


Figure 24: rmsd Dialog window

- 3 By clicking on the **Graph** button VMD produces a plot of the RMSD values between each pair of residues in the two aligned molecules (Fig. 25). The RMSD value is shown along the sequence of the protein, exhibiting regions of close alignment (small RMSD values) and of poor alignment (large RMSD values).

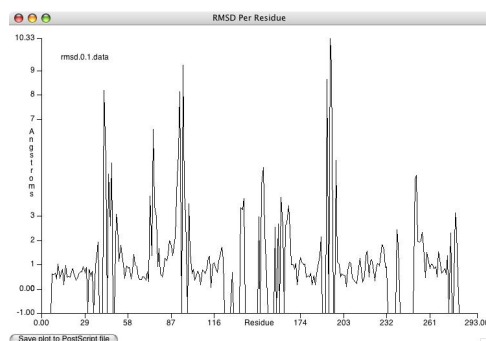


Figure 25: RMSD values between human AQP1 and *E.coli* AqpZ.

- 4 Go back to the Multiple Sequence Alignment window. Compare the RMSD graph with the sequence alignment, and identify the residues with large RMSD values. What structural elements do they correspond to? As you will see later, they correspond to the loops.
- 5 Select these residues in the Multiple Sequence Alignment window, by clicking on them. They will be highlighted as you click on them. (Fig. 26) If you want to select residues in different regions, hold down the shift key while selecting them.

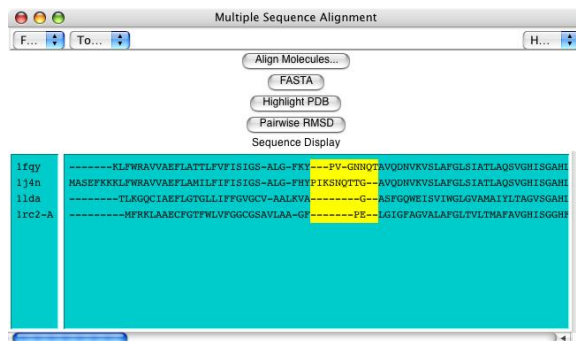


Figure 26: Multiple Sequence Alignment Window

- 6 As before, the selection will be highlighted in bond representation in the OpenGL Display window (Fig. 27).
- 7 Go back to the VMD Main window. Choose Graphics → Representation. In the Graphical Representation window (Fig. 28) you should find that a new representation has been created for each molecule, with the residues you have selected as shown in Fig. 26. Look at the representation for molecule 1fqy and compare the selection with the RMSD graph (Fig. 25)

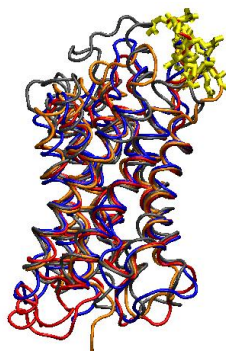


Figure 27: Highlighted residues shown in bond representation

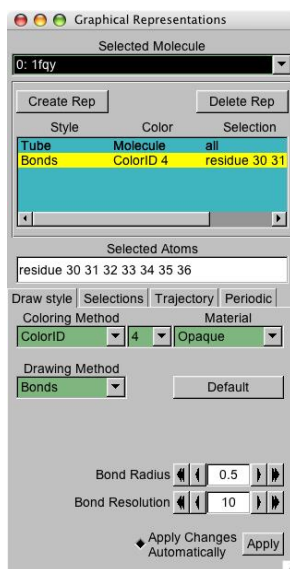
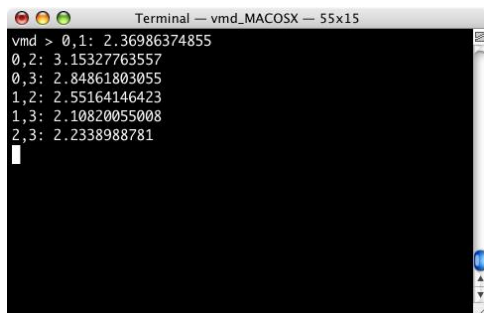


Figure 28: Graphical Representation Window

- 8 You can also choose more than one molecule and plot the RMSD values between them in the same graph. Go back to the RMSD Dialog window (Fig.24), choose all the molecules shown there, and click on the Graph button.
- 9 Now you see the RMSD between each pair of the molecules shown in a different color. The molecules are called according to their molid's in VMD (the number on the left hand side of the loaded molecules in the VMD Main window).

5.2 Pairwise RMSD between the molecules

You can obtain the overall average root mean square deviation between each two molecules, simply by clicking on the **Pairwise RMSD** button in the Multiple Sequence Alignment window (Fig. 26). Clicking on this button writes the average RMSD values between the molecules to the terminal window.

A screenshot of a terminal window titled "Terminal - vmd_MACOSX - 55x15". The terminal displays the following text:

```
vmd > 0,1: 2.36986374855  
0,2: 3.15327763557  
0,3: 2.84861803055  
1,2: 2.55164146423  
1,3: 2.10820055008  
2,3: 2.2338988781
```

Figure 29: Pairwise RMSD between the aligned molecules written to the terminal window

6 Examining the Aquaporin Tetramer

Aquaporins form homotetramers in cell membranes. The monomers are arranged side by side in a tight cluster. Each monomer conducts water. A fifth pore is formed in the center of the tetramer, with a yet unknown function. In the following we have a closer look at the aqua-glyceloporin GlpF, (PDB code: 1lda), an aquaporin that conducts both water and glycerol.

6.1 Loading Tetramer

In this section we will load the tetramer structure of GlpF and align it with the other four AQP molecules. A pdb file with the atom coordinates of the tetramer is provided with the tutorial (see Getting Started Section in the Introduction).

- 1 Go to the VMD Main Window
- 2 Choose File → New Molecule. Another window, the Molecule File Browser, will appear.
- 3 Use the Browse button to find the file `glpf_tetramer.pdb` in the directory `aqp_tutorial_files` → PDB, in the tutorial directory.
- 4 Once you have selected the pdb file, press the Load button in the Molecule File Browser window to load the molecule.

Now that you have the tetramer loaded, go back to the Multiple Sequence Alignment window and click on the Align Molecule button to align this molecule with the other four. Note that for this section you have to have the aquaporins 1j4n, 1fqy, 1lda and 1rc2 (Table 1.2) loaded. If you don't have them, go back and load the four pdb files, as explained in section 1.

6.2 Examining channel lining and tetrameric contacts

In the previous section, you have identified the conserved residues among the aquaporin molecules. Here we look where these residues are located in the tetrameric structure.

- 1 Go to the Tools → Bulk Residue Selection.
- 2 Choose Sequence Identity per Residue Greater than or equal to 1.0
- 3 Click on the Select button. Now you will see the conserved residues highlighted in the OpenGL Display window as well as in the Multiple Sequence Alignment window.
- 4 Take some time to examine the highlighted parts as shown in Figure 30. You recognize that most conserved residues line the channel interior. Locate the conserved residues that are outside the water conducting pore. Note that these residues are indeed located at the subunit-subunit interfaces of the tetrameric GlpF.

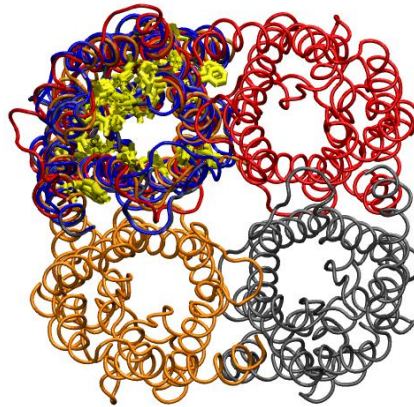


Figure 30: GlpF tetramer aligned with other aquaporin molecules. Conserved residues are highlighted in bond representataion (yellow).

- 5 Before starting the next section, go to the VMD Main window, choose the tetramer molecule by clicking on it (It will turn yellow), and delete the molecule by choosing `Molecule → Delete Molecule`.

7 Phylogenetic Tree

The Phylogenetic Tree feature in the MultiSeq tool elucidates the structure-based relationships between the four aquaporins considered.

Phylogenetic Trees can be constructed according to the RMSD or Q values between the molecules after alignment. Here we use the Q values to obtain the phylogenetic tree of four aquaporins.



Phylogenetic Tree. This is a tree diagram that depicts the line of descent in the evolution of the proteins analyzed. The tree is based on a metric that measures the evolutionary distances between the proteins. Various metrics can be employed. For example a pure sequence-based metric may be used. This is actually too simple of a metric for many purposes. Here we employ a metric based on comparing the structures of the proteins. The metrics used here are either derived from comparing Q values or RMSD values.

To utilize this feature:

- 1 In the MultiSeq window choose Tools → Tree. A window entitled treeWindow will open.
- 2 Here you can choose to construct a phylogenetic tree according to the Q_H values or RMSD. Q_H is a variation of Q (see below) that accounts for both gapped and aligned regions.



What is Q_H ? Q_H is an adaptation of the Q value, and is essentially a *metric for structural homology*. Q_H is comprised of two terms: Q_{align} and Q_{gap} . Each time you align two structures, unless they align perfectly, the structure can be divided into two parts: the part that aligns (Q_{align}) and the part that due to insertion and deletion of the protein sequence, does *not* align (Q_{gap}). Q_{gap} accounts for how much the insertion perturbs the core structure of the protein and for the size of the insertion in sequence and structure.

- 3 Click on the Q_H button. A graph will appear in the same window (Fig. 31). Here you can see the relationship between the four aquaporins, e.g. how the *E.coli* AqpZ (1rc2) is related to human AQP1 (1fqy). Identify bovine and human aquaporins in the Tree.
- 4 You can choose the information you want to be shown in the tree in the bottom part of this window. By default the filename and the scale of the Q_H values are shown.

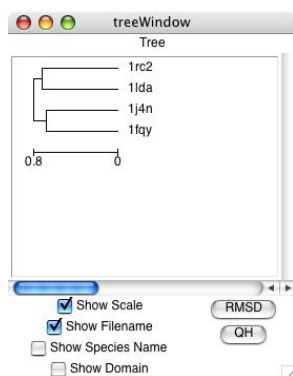


Figure 31: Structure-based phylogenetic tree, for the aquaporins in Table 1.2.

8 Exporting File Formats

As you work through this tutorial, you will notice that you are creating many new images, by highlighting various portions of the aligned protein sequences. If you would like to use this information further, you can generate your own PDB files, containing the coordinates of the selected residues. The aligned sequence information can also be saved in FASTA format, which is widely used in bioinformatics.

8.1 Exporting a PDB file from user-specified selections

In bulk residue selection you could select residues that have structural similarity (Q value) or amino acid sequence similarity. You can also select and deselect residues in the Multiple Sequence Alignment window, simply by clicking on them. All these selections can be written to separate PDB files, containing atom coordinates of the selected residues.

- 1 Go to the sequence alignment window and select the first aligned helix in all four molecules by clicking on them. This corresponds to the first part that is well aligned in the four aquaporins considered (Table 1.2). The selection will be highlighted.(Fig. 32)
- 2 As seen before, the selection will be shown in bond representation in the OpenGL Display window (Fig. 33).
- 3 Go back to the Multiple Sequence Alignment window (Fig. 32) and click on the Highlight PDB button.
- 4 You will be prompted to choose a directory to save the PDB file(s). The tutorial package contains a directory called PDB. Select this directory and click the Choose button. Once you choose this directory it will be used for further selections, unless you change it from File → Choose PDB Directory.

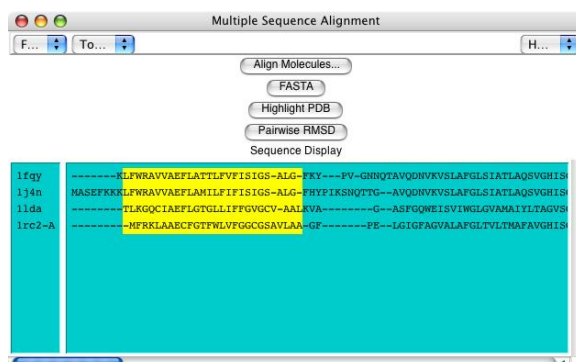


Figure 32: Highlighted residues in the Multiple Sequence Alignment window

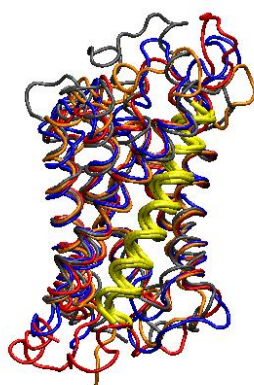


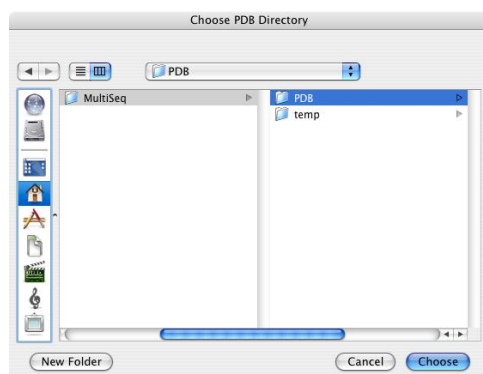
Figure 33: The first aligned helix highlighted.

- 5 The selections will be written to four pdb files, named: 1fqy.0.pdb, 1j4n.0.pdb, 1lda.0.pdb, 1rc2.0.pdb . Any other selection will be written to a file called accordingly, starting with the molecule name, and followed by a number showing the corresponding selection.
- 6 If you want to save the entire alignment, you can do so by going to the Bulk Residue Selection and selecting Q value, Greater than or equal to, and 0.0. The four sets of sequences will be highlighted. Click on Highlight PDB and four PDB files will be generated.

These segments can then be loaded into VMD and Multiseq for further analysis.

8.2 Exporting a multiple alignment in FASTA format

You can also create a FASTA format file from the aligned molecules by simply clicking on "FASTA" in the Multiple Sequence Alignment window. The



Generated PDB files:

1fqy.0.pdb
1j4n2.0.pdb
1lda.0.pdb
1rc2.0.pdb

Figure 34: Directories, where the generated PDB files will appear.

file, multiseq.fasta is saved in the temp directory that you have chosen before alignment. This file contains the sequence information of the aligned molecules, written in FASTA format, which is often required when searching molecular databases.