MOLECULAR MODELING OF D. VIVIPARUS ACHE

Using ClustalW to Align D. viviparus and T. californica ACHE Sequences

Lisa Speake

1. Before loading the protein sequences on to the ClustalW website, you must first create a Word document containing the sequences of both D. viviparus and T. californica in the correct format. First, open a new Word Document and type ‘>D. viviparus ACHE’ on the first line.

2. Next, find the sequence of D. viviparus ACHE on Genbank and paste it into the Word document.
   a. Open a new Internet Explorer or Mozilla Firefox page, and google ‘Genbank.’
   b. The first result of the search should be the NCBI HomePage. There will be a Genbank subheading under this result. Click on ‘Genbank.’
   c. The Genbank accession number for D. viviparus ACHE is DQ375489. Enter this number in the field at the top so that the Search toolbar reads, “Search Entrez for DQ375489.” Press Enter or click ‘Go.’
   d. In the column on the left side of the blue box, click on the one result for Nucleotide. On the next page, click on ‘DQ375489.’
   e. Scroll down towards the bottom of the next page. Before the section titled ‘ORIGIN,’ there will be a translation of the amino acid sequence of D. viviparus ACHE. Highlight the sequence, then right-click and select ‘Copy.’
   f. Paste the sequence below the ‘>D.vivparus ACHE’ heading you typed in the word document earlier. Delete the first 16 amino acids of this sequence (M-T). The first 16 amino acids are the signal sequence, and are cleaved before the protein is mature.

3. Skip a few lines, and then type ‘>T. californica ACHE’ in the Word document.

4. Now find the sequence of T. californica ACHE on Genbank and paste it under the heading you just typed in the Word document.
   a. In the Genbank window that you copied the amino acid sequence of D. viviparus from, search for the amino acid sequence of T. californica ACHE. The Genbank accession number for T. californica ACHE is CAA27169. The search toolbar at the top of the screen should read, “Search Protein for CAA27169.” Click ‘Go.’
   b. On the next page, click on ‘CAA27169.’
   e. On the next page, make sure the display option is ‘FASTA.’ Highlight the sequence, then right-click and select ‘Copy.’
   f. Paste the sequence below the ‘>T.californica ACHE’ heading you typed in the word document earlier. Delete the first 21 amino acids of this sequence (M-A). The first 21 amino acids are the signal sequence, and are cleaved before the protein is mature.
5. Now that you have created a correctly formatted Word document, highlight the entire text of the document, right-click, and select ‘Copy.’

6. Google ‘ClustalW.’ Click on the first link of the results.

7. In the second green box on the page, right-click in the text field and select ‘Paste.’
8. Click the ‘Run’ button. The internet window will change to a page that says that your results will appear in that window when the results are complete. It should take less than a minute or so for your results to appear.

9. At the bottom of the results page, click on the ‘View Alignment File’ button.

10. In the sequence alignment, asterisks (*) indicate residues that are completely conserved, while colons (:) represent residues that are very similar between the two sequences. The aromatic residues in the catalytic gorge of \textit{T. californica} are known, and are shown in the table on the following page. Use the ClustalW sequence alignment to find the corresponding aromatic residues in \textit{D. viviparus}. If the residue is conserved, circle it.

11. Now that you have completed the table above, you will make a figure with the results from ClustalW. First, copy the results from the ClustalW results page and paste it into a Word document. Set the font to Courier New and the font size to 9. You may have to change the margins to wide so that all the information fits on the correct lines as it does in ClustalW.

12. Change the heading of each line to \textit{D. viviparus} and \textit{T. californica}. Make sure you have the scientific names in italics.

13. Place ▼\textsuperscript{s} above the members of the catalytic triad. You can find the ▼ symbol under Insert>Symbol>More Symbols. The members of the catalytic triad for \textit{T. californica} are Ser200, Glu327, and His440.

14. Place ↓\textsuperscript{s} above conserved residues of the catalytic gorge.

15. Next, place ↑\textsuperscript{s} above the residues that are not conserved between the two sequences.

16. Highlight the residues in the acyl pocket, and make them bold.

17. Place a ■ above regions of the T-peptide that are conserved and a □ over regions of the T-peptide that are not conserved.
<table>
<thead>
<tr>
<th>Subsite</th>
<th><em>T. californica</em> ACHE</th>
<th><em>D. viviparus</em> ACHE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peripheral site</td>
<td>Tyr70</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tyr121</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Trp279</td>
<td></td>
</tr>
<tr>
<td>Choline binding site and hydrophobic site</td>
<td>Trp84</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tyr130</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Phe330</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Phe331</td>
<td></td>
</tr>
<tr>
<td>Acyl pocket</td>
<td>Phe288</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Phe290</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Val400</td>
<td></td>
</tr>
<tr>
<td>Wall of gorge</td>
<td>Trp114</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Trp233</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tyr334</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Trp432</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tyr442</td>
<td></td>
</tr>
<tr>
<td>T-peptide</td>
<td>Trp545</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Phe549</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Trp552</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tyr555</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Trp559</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Phe563</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tyr566</td>
<td></td>
</tr>
</tbody>
</table>
Using MOE to Create a Homology Model of D. viviparus AChE


2. Load the amino acid sequence of D. viviparus AChE from Genbank into the Sequence Editor in MOE.
   a. Open an Internet Explorer or Mozilla Firefox window and google 'genbank.'
   b. The first result of the search should be the NCBI HomePage. There will be a Genbank subheading under this result. Click on ‘Genbank.’
   c. The Genbank accession number for D. viviparus AChE is DQ375489. Enter this number in the field at the top so that the Search toolbar reads, “Search Entrez for DQ375489.” Press Enter or click ‘Go.’
   d. In the column on the left side of the blue box, click on the one result for Nucleotide. On the next page, click on 'DQ375489.'
   e. Scroll down towards the bottom of the next page. Before the section titled ‘ORIGIN,’ there will be a translation of the amino acid sequence of D. viviparus AChE. Highlight the sequence, then right-click and select ‘Copy.’
   f. Open a new document in Microsoft Word. Paste the sequence into this new document. Delete the first 16 amino acids of this sequence (M-T). The first 16 amino acids are the coding sequence, and are cleaved before the protein is mature. Highlight the sequence, right-click, and select ‘Copy.’
   g. Now open the Sequence Editor in MOE by clicking on the ‘Window’ drop down menu in the MOE main screen and then clicking on ‘Sequence Editor.’
   h. In the Sequence Editor, click on the ‘Homology’ drop down menu and then ‘PDB Search.’
   i. In the MOE-Search PDB window, click ‘Paste’ in the upper, right corner. The sequence that you copied from the word document should appear in the top screen of the window.
   j. Click the ‘Search’ button in the bottom, left corner. Select the hydrolase that appears at the top of the bottom screen by left-clicking on it, then click on the ‘Load Alignment…’ button on the right of the window between the upper and lower screens.
   k. When the MOE-SearchPDB: Load Alignment window appears, select the first chain in the window and click on the toggle box next to ‘Load Query Sequence’ at the bottom of the window. Click ‘Load Selected.’

3. The sequences for the D. viviparus AChE and a sequence that MOE found to be a good template for homology modeling for this protein should now appear in the Sequence Editor as Chain 1 and Chain 2, respectively. However, we want to use Human AChE as a template for homology modeling this protein instead of the hydrolase that MOE found.
4. In the ‘Display’ drop down menu of the Sequence Editor, select ‘Compound Name.’ Also select ‘Single Letter Residue.’

5. Chain 1 (the sequence of *D. viviparus* AChE) will not have a name because you pasted its sequence into MOE. Chain 2 will be the hydrolase that we are not interested in. Select Chain 2 by clicking on the box with the number 2 in it on the left side of the Sequence Editor. Next, right-click on the box and select ‘Delete Chain.’ A confirmation window will appear; click ‘Ok.’

6. Now, you need to load the template sequence that we want to use for this homology model, Human AChE, into the sequence editor.
   b. In the search field of the RCSB Browse window, enter ‘1b41.’ 1b41 is the PDB (Protein Data Bank) code for Human AChE complexed with fasciculin-II, glycosylated protein. After typing the PDB code in the search field, this protein will be displayed in the screen above the search field. Select the protein by clicking on it.
   c. Select the toggle boxes beside ‘Copy Proteins to MOE’ and ‘Overwrite Existing Files’ at the bottom of the RCSB Browse window.
   d. Click on the 'Download' button at the bottom of the window.
   e. An RCSB Download window will appear. Click ‘Start.’
   f. When the protein is downloaded into the Sequence Editor, the buttons at the bottom of the RCSB Download window will change to a ‘Close’ button. Close the RCSB window.

7. Look back at the Sequence Editor. Sequences for the Human AChE, along with sequences for the molecules it is complexed with appear as Chains 2-6. The only chains you should leave in the Sequence Editor are Chain 1 (the *D. viviparus* AChE) and Chain 2 (1B41.A Hydrolase/Toxin). To delete Chains 3-6, first select Chains 3-6. Hold down the Shift key while clicking on the boxes with the numbers 3-6 individually. Right-click over the selected boxes, and select ‘Delete Selected Chains.’ Click ‘Ok’ when the confirmation box appears.

8. In order to build a homology model, you must first align the sequence with unknown structure (*D. viviparus* AChE) to the template sequence.
   a. Select Chain 1, then click on ‘Invert Chains’ in the ‘Selection’ drop down menu of the Sequence Editor.
   b. Open the MOE-Align window by selecting ‘Align’ in the ‘Homology’ drop down menu of the Sequence Editor. In the MOE-Align window, select ‘freeze’ in the Chain Selection option. Click ‘Ok’ to close the window.

9. Now you will build the homology model.
   a. Select ‘Homology Model’ in the ‘Homology’ drop down menu of the Sequence Editor.
b. A Homology Model window will appear. Make sure that under the 'Models & Templates' heading, the sequence is 'Chain #1 <no name>' and the template is 'Chain #2 1B41.A.'

c. At the bottom of the same heading, change the number of models to 25.

d. Under the 'Model Refinement' heading at the bottom of the window, change the RMS gradient to 10 for both the intermediates and the final model.

e. Under the same heading, click on the 'Potential Setup' button. A window named Potential Setup – MMFF94x will appear. Click on the downward-pointing arrow next to ‘Load…’ in the upper left portion of the window. Choose CHARMM27. Click ‘Close’ to return to the Homology Model window.

f. Click ‘Ok,’ and MOE will begin to generate a homology model of *D. viviparus*. If a window pops up that asks if you would like to overwrite the file, click ‘Yes.’ Generating the model will take at least an hour. You can watch MOE’s progress by opening the Database Viewer. Open the Database Viewer by selecting ‘Open…’ from the ‘File’ drop down menu on the MOE main screen. Select promodel.mdb and then click ‘Ok.’ You can also see MOE’s progress in generating the model at the top, left corner of the MOE main screen.

10. When the homology model is complete, a composite model of the 25 models that were created will be displayed in the main window of MOE. In the ‘Render’ drop-down menu, select ‘Hide’ and then ‘All.’ Then, again in the ‘Render’ menu, select ‘Backbone’ and ‘Cartoon.’ You can now see the backbone of the protein and its secondary structure.

11. Now that you have created a model of *D. viviparus* AChE, you need to examine its restraints. Although models built from target sequences (those with unknown structure) that are very similar to their template sequence are usually accurate, sometimes homology models have outliers in the proposed sequence. Outliers are portions of the model that have abnormal or unreasonable stereochemical features.

   a. To view a list of the amino acids in the model of *D. viviparus* AChE that are outliers, select ‘Protein Geometry’ from the ‘Measure’ drop down menu in the Sequence Editor window.

   b. When the Protein Geometry window appears, choose ‘Data’ as the display option in the upper, right corner.

   c. To highlight the outliers in the Data list, hold down the Shift button as you first click on the first and then the last outlier in the list.

   d. Press the ‘Select Atoms’ button at the bottom of the Protein Geometry window. This selects the residues that are outliers on the model in the main MOE window. Selected residues appear as pink. Do not click anywhere on the model or on the black screen on the MOE main window; doing so will unselect the residues.
Using SWISS-MODEL to Create a Homology Model of *D. viviparus* AChE

1. Google ‘SWISS-MODEL,’ and click on the first link in the results.

2. In the left menu window, click ‘First Approach Mode.’

3. In the next page, enter your email address and a unique title for your project in the appropriate spaces. In the text field below ‘Provide a protein sequence or a UniProt AC Code,’ paste the amino acid sequence of *D. viviparus* AChE. Find the sequence, as before, using GenBank, and paste the mature sequence (with the first 16 amino acids cleaved) in the text field.

4. Click ‘Submit Modelling Request.’ Do not close the SWISS-MODEL window; the homology model results will be displayed in the window that you currently have opened.

5. When the model is finished running, click on ‘pdb’ beside ‘download model: as.’ A File Download – Security Warning window will appear. Click ‘Save.’ Save the model to the desktop.

6. You can view the model by opening Swiss-pdbviewer. Choose ‘Open PDB File…’ from the File drop-down menu and select the pdb file from the list that you just saved to the desktop.

Using PROCHECK to Determine the Accuracy of Homology Models

PROCHECK checks the stereochemical features of a model. By running a PROCHECK analysis on the models you have created, you will be able to tell which model is more accurate. You will run one analysis on each model you created, using the following protocol:

1. Open the website [http://www.ebi.ac.uk/thornton-srv/databases/pdbsum/upload.html](http://www.ebi.ac.uk/thornton-srv/databases/pdbsum/upload.html).

2. Fill in the fields with the appropriate information on this page, and click ‘SUBMIT.’

3. Soon an email will be sent to you with a link to the PDBsum pages for the structure you submitted. Click on the link.

4. When the custom-generated PDBsum page opens, click on ‘Run PROCHECK’ under the ‘Tools’ heading in the bottom, left portion of the page.

5. When the next window appears, click ‘here’ to see the results.
6. Click on the icon that looks like a stack of papers under the ‘Text file’ heading.

7. Another window will pop up. Scroll to the bottom of that window and look at the overall average of the G-factors. As stated at the bottom of the window, “Ideally, scores should be above -0.5. Values below -1.0 may need investigation.”

What is the overall average of G-factors for the model you created using MOE?

What is the overall average of G-factors for the model you created using SWISS-MODEL?

Which model is a more accurate model of *D. viviparus* AChE?

Now that you have determined which homology model is a better representation of *D. viviparus* AChE, you will use MOE to dock the inhibitor BW284c51.

1. SMILES (Simplified Molecular Input Line Entry Specification) is notation that software can read and interpret into a two or three dimensional model for a molecule. SMILES will help you, in this case, because you will not have to manually build the structure of BW284c51 in MOE. Google ‘pubchem.’ Click on the first link that appears. In the search page, enter ‘BW284c51.’ Click on the inhibitor’s name on the next page. Under the Descriptors Computed from Structure heading, highlight and copy the Canonical SMILES notation beginning with the first C and ending with the C just before the period (exclude the bromines).

2. In MOE, click on Builder on the right control panel. In the text field in the middle of the Molecule Builder window, press Ctrl+V to paste the SMILES notation for the inhibitor in the field. Click on the smiling face beside the text field, and close the Molecule Builder window. The three dimensional structure of BW284c51 should appear in the main window of MOE.

3. To also add the model of *D. viviparus* AChE to the window for docking, open the file for the preferred model. If it is saved on the desktop, the Path: field in the Open window should read c:/Documents and Settings/06dx020/Desktop (the numbers in 06dx020 will vary, depending on which computer in the lab you are using). When both structures are loaded, you will need to zoom out to see both of them in the field at the same time.

4. Open the Sequence Editor in the Window drop-down menu. To align the sequence with the number line so that the numbers correspond with the numbers in your ClustalW lab exercise, shift the sequence so that the first amino acid
(Leucine) is number 12 on the number line. To do so, press Ctrl and the middle button on the mouse and drag the chain to the desired position.

5. Select the amino acids that correspond to the aromatic amino acids listed in the chart above (except for the T-peptide residues and also those below for *T. californica*). You will need to refer to the sequence alignment you created using ClustalW. Select each by right-clicking on the name of the residue and choosing “Atom” and “Select.”

<table>
<thead>
<tr>
<th><em>T. californica AChE</em></th>
<th><em>D. viviparus AChE</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Gln74</td>
<td></td>
</tr>
<tr>
<td>His440</td>
<td></td>
</tr>
<tr>
<td>Glu199</td>
<td></td>
</tr>
<tr>
<td>Gly117</td>
<td></td>
</tr>
<tr>
<td>Asp72</td>
<td></td>
</tr>
</tbody>
</table>

6. Close the Sequence Editor window, and without clicking anywhere else in the MOE main window, select “Ball and Stick” from the “Render” drop-down menu. Also select “Label” in the right-hand panel of the screen and select “Residue.” The residues you selected should now be labeled, so check to make sure you selected the correct ones.

7. Select BW284c51 by dragging a box around it in the main window. In the “Selection” drop-down menu, select “Ligand.” This will label the inhibitor as the ligand of the docking simulation. Deselect BW284c51 by clicking somewhere in the black part of the MOE window.

8. Now, reselect the residues you originally selected on the AChE. To do this, hold down Ctrl and Shift while clicking on one atom from each of the residues. You may have to zoom in to see the AChE better and determine which residues are displayed as ball and stick.

9. In the “Compute” drop-down menu, choose “Simulations” and then “Dock…” The Dock window will appear.

10. Next to **Receptor:** click the question mark; the whole AChE molecule should briefly flash. Next to **Site:** select “Selected Atoms.” Clicking the question mark next to **Ligand:** should cause BW284c51 to flash.
11. Be sure that “Alpha Triangle” is selected next to **Placement:** and that “Affinity dG” is selected next to **Rescoring 1:**. Select “Forcefield” next to **Refinement:** and “Affinity dG” next to **Rescoring 2:**. Click “OK.”

12. Take note of the column labeled S in the Database Viewer window. This is the energy value for the conformations of the ligand inside the AChE pocket. The conformations are listed in order of the best to worst energy values. Record the number in the top row of the column.

13. Minimize the Database Viewer window. With the active site residues still selected, hide the rest of AChE and BW284c51 by selecting “Hide” and then “Unselected” on the right-hand panel of the main MOE window.

14. Unselect all atoms by clicking somewhere in the black part of the screen.

15. In the Database Viewer window, select ‘Browser…’ from the File drop-down menu. A window will appear that will allow you to scroll through the different conformations of the active site/ligand interaction.

16. Repeat the docking with ethopropazine, iso-OMPA, and physostigmine.