**Assignment 3**

Your name:    
Your email address:

Once you are done with the exercise, email your answers as an attachment to [gogarten@uconn.edu](mailto:gogarten@uconn.edu) and [daniel.s.phillips@uconn.edu](mailto:daniel.s.phillips@uconn.edu)

We will start today’s class with another **Team Based Learning** exercise.

We will use chimera for today’s exercise. Given the problems some students had with chimera under windows, please avail yourself to the laptops provided in class.

### Objectives:

* Know how to identify domains in multi domain proteins in chimera;
* create a multiple sequence alignment based on aligned structures in chimera
* align structures of very divergent proteins;
* inspect protein DNA interactions;
* identify the major and minor groove in a DNA molecule;

**Comparing divergent proteins with similar structures  
A) Inteins and mini inteins**

Most inteins are composed of two domains: one is responsible for protein splicing, and the other has endonuclease activity. A few inteins have lost the endonuclease domain completely and retain only the self-splicing domain and activity. The latter inteins are called **mini-inteins**.

The structures of several inteins have been determined through X-ray crystallography.  Today we will use the following:

* Saccharomyces cerevisiae intein (PMID: 9160747, [1VDE](https://j.p.gogarten.uconn.edu/mcb3421_2021/labs/1VDE.pdb)),
* the Mycobacterium xenopi mini intein (PMID: 9437427, [1AM2](https://j.p.gogarten.uconn.edu/mcb3421_2021/labs/1AM2.pdb))

**To do:**

1. Open 1VDE in chimera. This structure has two chains. Select **chainA**and save the **selected residues only** into their own pdb file (file > save pdb > fill out the form, check save selected residues only, save). Close the session.
2. Reopen the saved chain A in chimera. Open Mycobacterium mini intein 1AM2. Depict the structures as ribbons and color them according to the secondary structure (if you press the shift key while executing an action, it should act on both chains; but there also is a check mark in the "Tools > Depiction > Color Secondary Structure".  However, it also might look nice to use slightly different colors. for the two chains.   Rotate the two structures until you can see the similarities between mini intein and the large intein. (Use the "favorites > model panel" to rotate one or the other structure). Which part of the structures appears to be similar?   
   **Your answer** --->
3. Align two structures using Tools > structure comparison > matchmaker. Does the alignment correspond to your expectation?   
   **Your answer** --->
4. You can find out which part of *Saccharomyces cerevisiae* intein corresponds to the endonuclease domain by comparison of the two structures. The splicing domain matches between the two structures. The part that does not match is a candidate for the homing endonuclease domain.   
   Using "tools > compare structure > match -align" create a pairwise alignment of the two inteins based on the match between the two structures (you need to have had run matchmaker first!).   
   Color the putative self-splicing (i.e. the part that is present in 1VDE and in 1AM2) and endonuclease domains of 1VDE (no corresponding part in 1AM2) in two different colors   
   (selecting consecutive residues works easily via the alignment window; if you press shift, you can add to the selection.)  Do you get the same result as above?  
   **Your answer** --->  
   In Tools > Structure Comparison > Match align you obtain a sequence alignment.  
   In that alignment window, click on info, and then percent identity. This returns the % identity in the matched part of the sequences (at the bottom of the window displaying the structure)  
   What is the percent identity between the aligned parts of the splicing domain?   
   **Your answer** --->

**Save your project.**

1. Find and select the N and C terminals (first a.a. and the last a.a.) in both structures. If you hover over the beginning or end, the name of the residues pops up in a little window. CTRL click selects the amino acid or atom, and shift control click adds to the selection. Under actions>atoms/bonds>show side-chains make the side chains of the first and last amino acid visible.   
   Optional: hide the rest of the structure, if it distracts you.   
   Rotate the structure of the sidechains of the first and last aa and decide which atoms are closest. Select these atoms (ctrl click and shift ctrl click), then go to tools>Structure analyses> distances and click on the **create** button in the window that pops up. Repeat this for a few atoms from the first and the last aa. How close are beginning and end (in Ångström and in nanometers)?  
   **Your answer** --->   
   in 1VDE:   
   in 1AM2:

A close-up of several colorful arrows

Description automatically generated

Alignment between 1VDE chainA and 1AM2. The non-matching part is in green, the end of the intein in blue and the beginning in purple.

The vma-1 intein bound to DNA

1. Open Saccharomyces cerevisiae intein that is bound to its target DNA sequence.  
   Does the DNA - Protein interaction in 1LWS agree with your previous assignment of the self-splicing domain?   
   **(see the saved structure from the previous exercise)**

**Your answer** --->

1. Try to find a way to display the interactions between the amino acid side chains and the DNA helix.  
   One way to do it is to select two DNA chains and select aa in the neighboring zone. To do this you could first select chain A, then invert the selection (you now should have selected Chain B and C). Then select zone 4 or 5 Angstrom. Make the side chains visible, and display either the side chains or the DNA as spheres. One way to look at individual interactions is to turn the molecule so that one looks down the DNA helix, and then to use the viewing controls to only look at a cross section (or Slab) of the structure. If the bases of the DNA are displayed too cartoonish, you can change the display in Actions > Atoms/Bonds > nucleotide objects (select different option and click apply).  
   Most of the interactions of aa side chains are with the major groove of the DNA. Do you find residues that interact with the minor groove? If yes, which aa are involved:

**Your answer** --->

1. The Lys 340 and Glu 366 are residues that are important for interaction with DNA. Select those residues (Tools> sequence>sequence allows to show the primary sequence, which is a good way to select a particular aa). Which base pairs interact with these amino acids? (if you hover over an atom, a pop-up window gives the base, the number of the base, and the chain (e.g., G 23.B is the 23 rd base in chain B, which is a Guanin). It might look nice to look at a cross section of the molecule perpendicular to the DNA.

**Your answer** --->

**Comparing other divergent proteins with similar structures (optional)  
 the GRASP nucleotide binding site in enzymes that use ATP to synthesize something**

Glutathione synthetase and D-Alanine D-Alanine Ligase were long ago (1990) recognized by Jim Knox (MCB faculty) to be so similar in structure that there could be no doubt about their common evolutionary origin. Later these and other enzymes were discovered to have a novel ATP binging site (the GRASP domain). These domains were identified using profile aligments (will be covered later in this course). A description of the GRASP domain family is at <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3249071/>.

Save the following files to your computer and load them into chimera (new session).

[2DLN.pdb](https://j.p.gogarten.uconn.edu/bioinf/2DLN.pdb) (D-Alanine D-Alanine Ligase - D-ala is an important part of the bacterial cell wall, more [here](http://en.wikipedia.org/wiki/D-alanine-D-alanine_ligase),

[1GSA.pdb](https://j.p.gogarten.uconn.edu/bioinf/1GSA.pdb)(glutathione synthetase from E. coli, [glutathione](http://en.wikipedia.org/wiki/Glutathione) is the biological equivalent of mercaptoethanol),

[cpsBfrag.pdb](https://j.p.gogarten.uconn.edu/bioinf/cpsBFRAG.pdb), load [cpsFfrag.pdb](https://j.p.gogarten.uconn.edu/bioinf/cpsFFRAG.pdb) ([Carbamoyl phosphate synthetase](http://en.wikipedia.org/wiki/Carbamoyl_phosphate_synthetase) is an enzyme consisting of several domains. These are the front and the back of (1BXR download the fragments from the link, not the whole pdb file).

Based on your first impression, are these structures similar? homologous?

**Your answer** --->

Can you use Tools > Structure comparison > Matchmaker to align these structures?

**Your answer** --->

See [here](https://j.p.gogarten.uconn.edu/bioinf/graspdomainsbmp.htm) for an illustration of the structures in similar orientation.

**Comparing other divergent proteins with similar structures  
B) Inteins and mini inteins**

Most inteins are composed of two domains: one is responsible for protein splicing, and the other has endonuclease activity. A few inteins have lost the endonuclease domain completely and retain only the self-splicing domain and activity. The latter inteins are called **mini-inteins**.

The structures of several inteins have been determined through X-ray crystallography.  Today we will use the following:

* Saccharomyces cerevisiae intein (PMID: 9160747, [1VDE](https://j.p.gogarten.uconn.edu/mcb3421_2021/labs/1VDE.pdb)),
* the Mycobacterium xenopi mini intein (PMID: 9437427, [1AM2](https://j.p.gogarten.uconn.edu/mcb3421_2021/labs/1AM2.pdb))

**To do:**

1. Open 1VDE in chimera. This structure has two chains. Select **chainA**and save the **selected residues only** into their own pdb file (file > save pdb > fill out the form, check save selected residues only, save). Close the session.
2. Reopen the saved chain A in chimera. Open Mycobacterium mini intein 1AM2. Depict the structures as ribbons and color them according to the secondary structure (if you press the shift key while executing an action, it should act on both chains; but there also is a check mark in the "Tools > Depiction > Color Secondary Structure".  However, it also might look nice to use slightly different colors. for the two chains.   Rotate the two structures until you can see the similarities between mini intein and the large intein. (Use the "favorites > model panel" to rotate one or the other structure). Which part of the structures appears to be similar?   
   **Your answer** --->
3. Align two structures using Tools > structure comparison > matchmaker. Does the alignment correspond to your expectation?   
   **Your answer** --->
4. You can find out which part of *Saccharomyces cerevisiae* intein corresponds to the endonuclease domain by comparison of the two structures. The splicing domain matches between the two structures. The part that does not match is a candidate for the homing endonuclease domain.   
   Using "tools > compare structure > match -align" create a pairwise alignment of the two inteins based on the match between the two structures (you need to have had run matchmaker first!).   
   Color the putative self-splicing (i.e. the part that is present in 1VDE and in 1AM2) and endonuclease domains of 1VDE (no corresponding part in 1AM2) in two different colors   
   (selecting consecutive residues works easily via the alignment window; if you press shift, you can add to the selection.)  Do you get the same result as above?  
   **Your answer** --->  
   In Tools > Structure Comparison > Match align you obtain a sequence alignment.  
   In that alignment window, click on info, and then percent identity. This returns the % identity in the matched part of the sequences (at the bottom of the window displaying the structure)  
   What is the percent identity between the aligned parts of the splicing domain?   
   **Your answer** --->

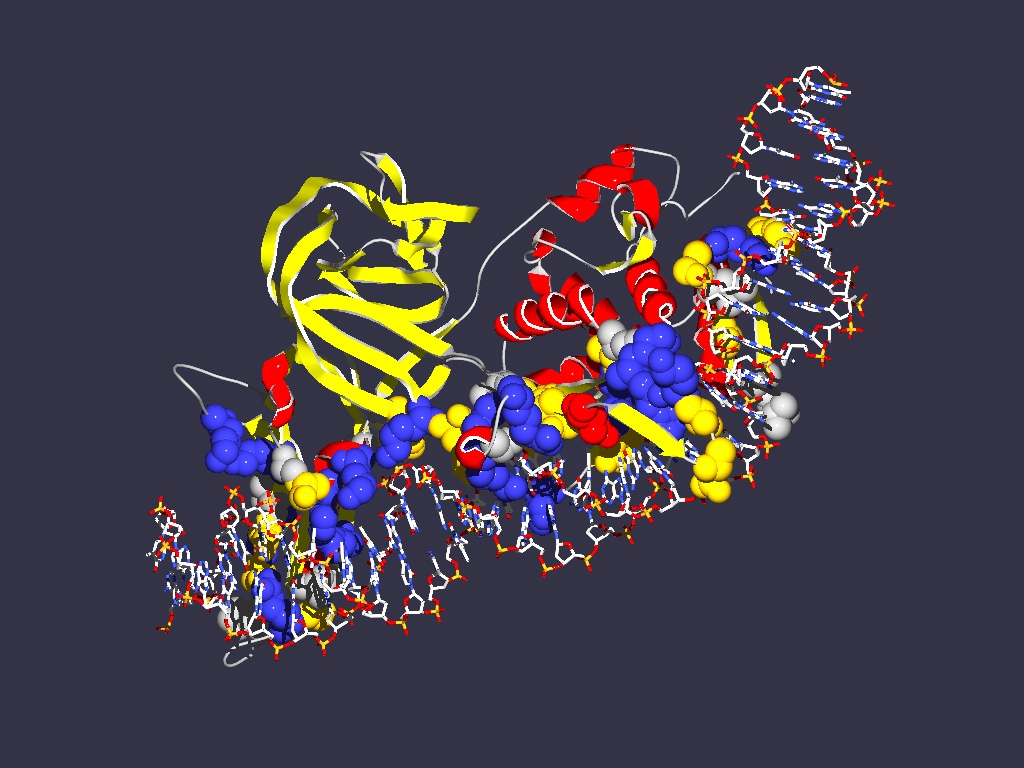
**Save your project.**

1. Find and select the N and C terminals (first a.a. and the last a.a.) in both structures. If you hover over the beginning or end, the name of the residues pops up in a little window. CTRL click selects the amino acid or atom, and shift control click adds to the selection. Under actions>atoms/bonds>show side-chains make the side chains of the first and last amino acid visible.   
   Optional: hide the rest of the structure, if it distracts you.   
   Rotate the structure of the sidechains of the first and last aa and decide which atoms are closest. Select these atoms (ctrl click and shift ctrl click), then go to tools>Structure analyses> distances and click on the **create** button in the window that pops up. Repeat this for a few atoms from the first and the last aa. How close are beginning and end (in Ångström and in nanometers)?  
   **Your answer** --->   
   in 1VDE:   
   in 1AM2:

A close-up of several colorful arrows

Description automatically generated

Alignment between 1VDE chainA and 1AM2. The non-matching part is in green, the end of the intein in blue and the beginning in purple.



1LWS.pdb

If you have trouble to wrap your head around the major and minor groove, build your own DNA molecule in chimera.   
Favorites > Build structure   
Select DNA and B-form enter a sequence (As, Gs, Ts and Cs)

A screenshot of a computer

Description automatically generated

Click apply.

A structure of a molecule

Description automatically generated  
dsDNA with hydrogen bonds

A close-up of a molecule

Description automatically generated  
Same but showing the surface.

Note: DNA is NOT a symmetric spiral as depicted in numerous logos and artwork:

A graph showing the number of thermometers

Description automatically generated A diagram of a dna molecule

Description automatically generated  
A dna double helix with text

Description automatically generated

Google DNA, then select images for many more horrifying examples/

For more things to try out with divergent protein sequences go to the bottom of the assignment 3 page at <https://j.p.gogarten.uconn.edu/mcb3421_2021/labs/assign3.html>