Assignment 3_2024

Your name:
Your email address:

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

In case of problems with ChimeraX, 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;
- Predict the structure of an intein using alphafold

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:

1. Saccharomyces cerevisiae intein (PMID: 9160747, 1VDE),
2. the Mycobacterium xenopi mini intein (PMID: 9437427, 1AM2)

To do:

1. Open 1VDE in chimeraX. The unit cell of the crystal has two nearly identical 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. It might look nice to use slightly different colors for the two chains (different shades of red and yellow). Rotate the two structures until you can see the similarities between mini intein and part of the large intein. (Use the right mouse button menu to rotate one or the other structure). Which part of the structures appears to be similar?
   Your answer --->

   Align two structures using Tools > structure comparison > matchmaker. Does the alignment correspond to your expectation?
   Your answer --->

   Save your project.

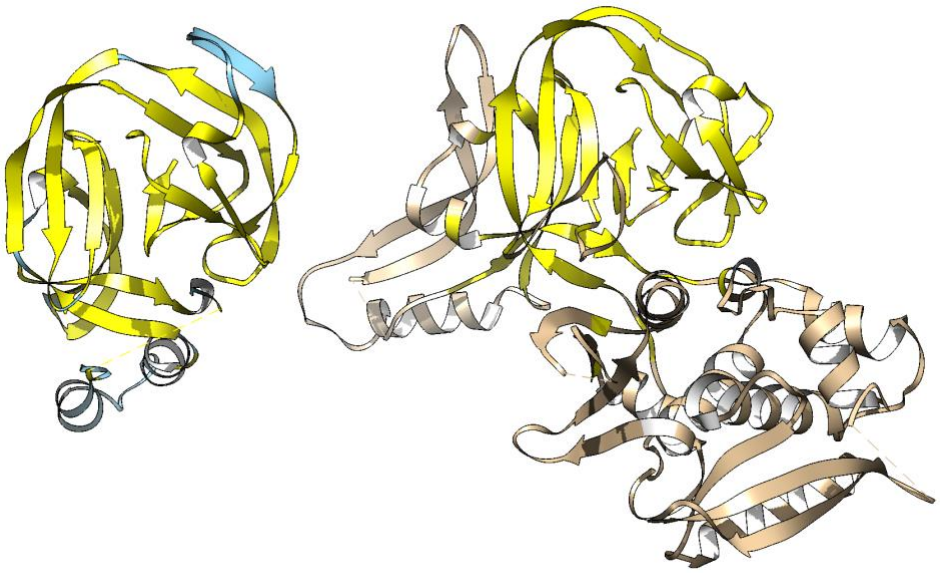
This page is for your information, we will not do this in class today

Note: chimera, but not chimeraX provides the possibility to align sequences based on their structure:

In chimera after you loaded chainA_1VDE and 1AM2, run Tools>Structure comparison> Matchmaker, then Tools>Structure comparison>Match-Align.

This results in an alignment of the primary sequences. You can select the matching residues:

You then can use action>color to color the aligned regions:



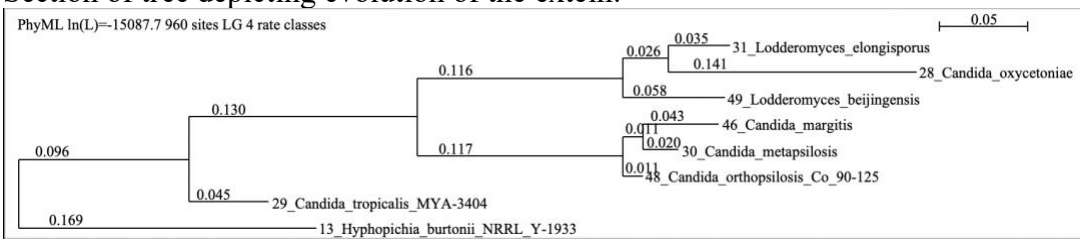
3. Back to chimeraX: Find and select the N and C terminals (first a.a. and the last a.a.) in both structures. The easiest is to select them via Tools>Sequence>SequenceViewer.) Click selects the amino acid or atom, and shift click adds to the selection.

   Use actions>atoms/bonds>show>side-chains to make the side chains of the first and last amino acids visible.
   Hide the rest of the structure (invert selection, actions>cartoon>hide).
   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:

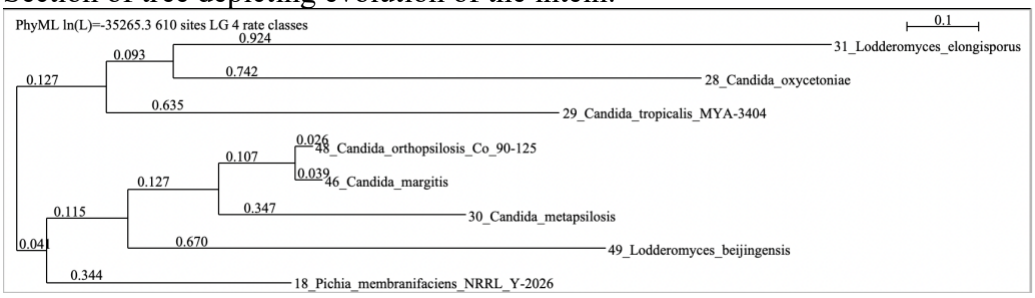
Using alphafold from within chimera to predict the structure of intein:

The intein found in the *Saccharomyces cerevisiae* gene encoding the catalytic subunit of the vacuolar ATPase (gene name *vma*1) is also found in many other yeasts (different species and genera). Not surprisingly, the intein is much less conserved than the host gene (the catalytic subunit). We want to figure out, if along with changes in the sequence, the structure changed as well.

Section of tree depicting evolution of the extein:

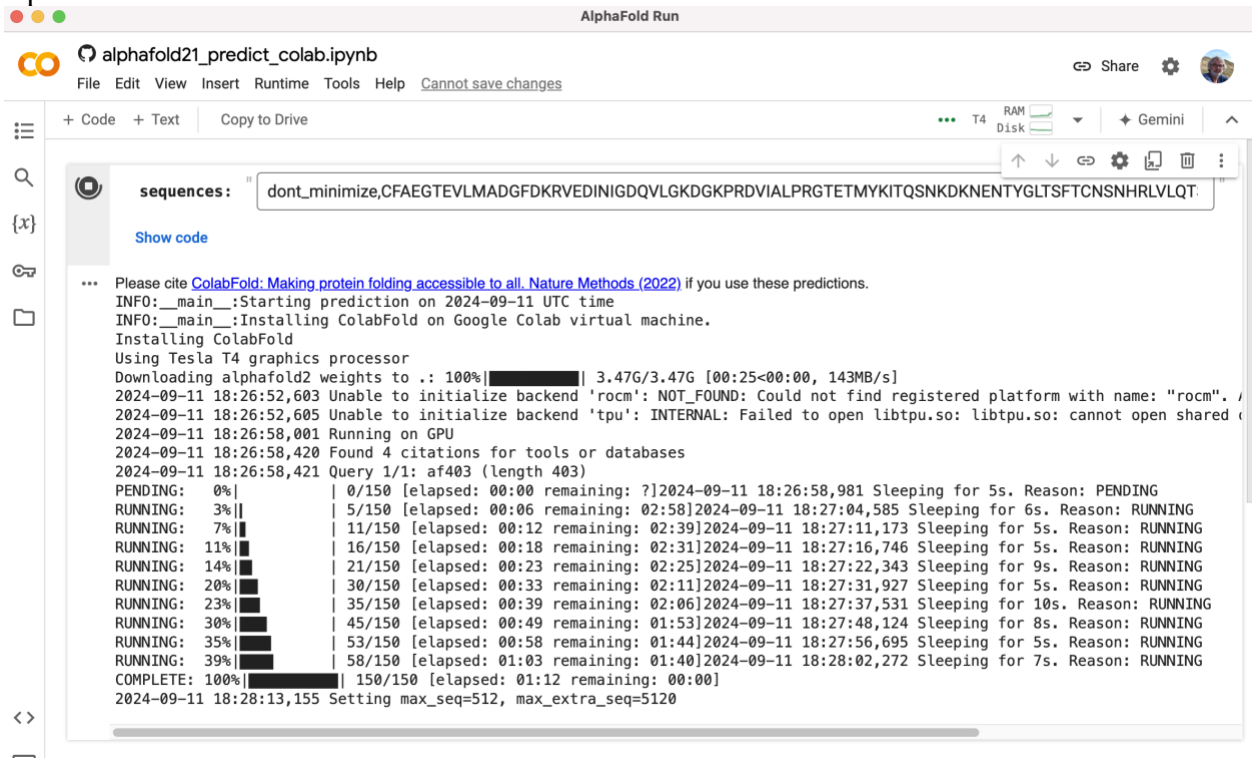


Section of tree depicting evolution of the intein:

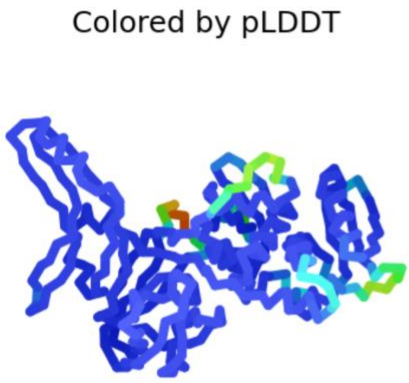


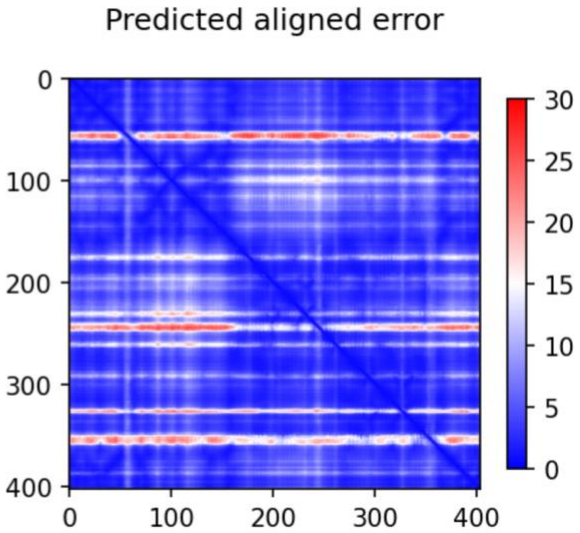
Numbers at the branch give the average number amino acid substitution per site along the branch.

1. The class notebook in the "_collaboration space" has a folder with inteins from yeasts, and the first page has a listing of several inteins from different yeasts. The sequences are in the so-called FASTA format (more on this later). The annotation lines start with a > symbol. Each student should work on an intein from a different species. The instructor will assign each student a number (the first entry following the >).
   You need to take notes of the species, annotation line, and sequence of the intein assigned to you. Chimera's interface to alphafold does not keep track of anything, and **it only wants and remembers the sequence, not including the annotation line**.
   - In the shared document, write your name at the end of the annotation line!
   - Open a **new session** in chimeraX.
   - Under Tools > Structure Prediction select alphafold.
   - Somewhere along the line you will be asked to login with your gmail username and password.
   - Be careful when clicking on things. Per gmail account you have only a limited number of predictions that you can calculate.
   - Paste the sequence only – without annotation line – into the window. (Make sure you have the complete sequence – all the vma1 inteins end on HN.)
   - Click predict.
   - Click on run anyway. (danger: if you click somewhere else, the window disappears)
   - Somewhere on your screen should be a window that reports progress on the alpha-fold run:



2. Be patient! After about 6-8 minutes you should see the first prediction. Something like this:





This will be updated periodically.

3. When the prediction is done (about 20-30 minutes after you started), the predicted structure will be opened in the chimera main window, the ribbon display will be colored accoring to reliability: blue = good, yellow = poor.
4. Save this structure as a chimeraX project under the species name and the accession number. E.g.: 22_Ascoidea_rubescens_DSM_1968 _XP_020046594.1.cxs
5. In chimeraX align the predicted structure to 1AM2 and 1VDE chain A.

   Do you observe any differences between 1VDE and the predicted structure?
   Your answer --->
   Are these differences in reliably predicted parts of the predicted intein structure?
   Your answer --->
   Are these differences in the self-splicing or the homing endonuclease domain?
   Your answer --->

The vma-1 intein bound to DNA

1. Close the current Chimera X session (make sure you documented your findings first).
   Open *Saccharomyces cerevisiae* intein that is bound to its target DNA sequence (Fetch by ID 1LWS).
   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)

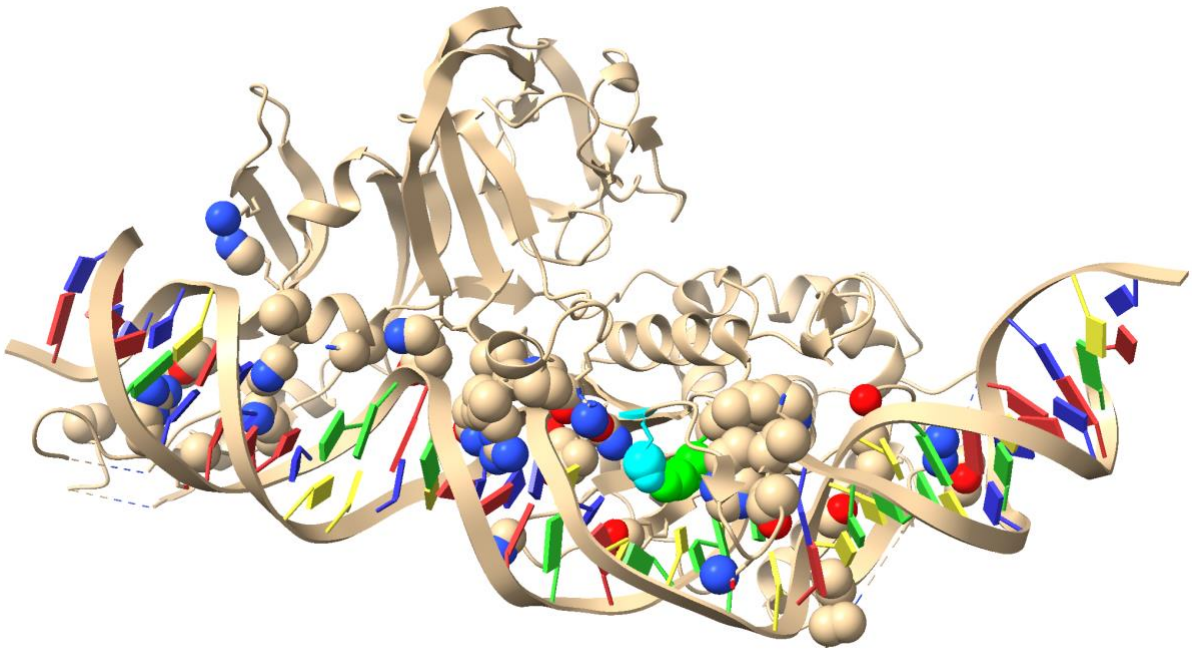
   Your answer --->

2. 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); or you could change the selection mode and select chain B and C successively. 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: open Tools > General > Side View. The yellow lines in the side view window are the lines where become invisible (clipped). Rotate the structure and move the Yellow line, so that you only see one turn on the DNA helix. 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 --->

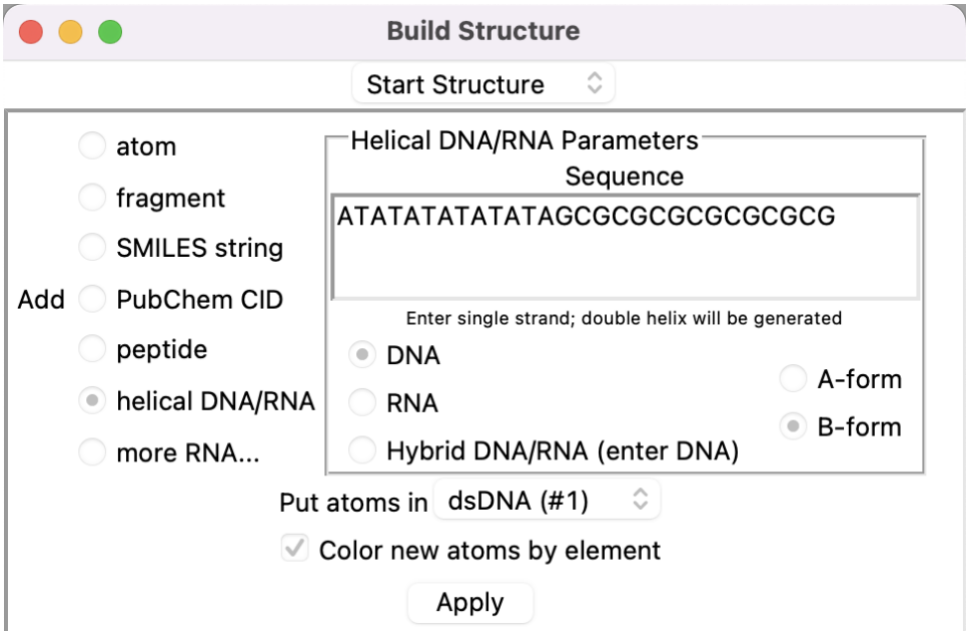
3. The Lys 340 (K) and Glu 366 (E) are residues that are important for interaction with DNA. Select those residues (Tools> sequence>sequence viewer allows to show the primary sequence, which is a good way to select a particular aa).
   Which base pairs interact with these amino acids? (if you hover over 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 --->

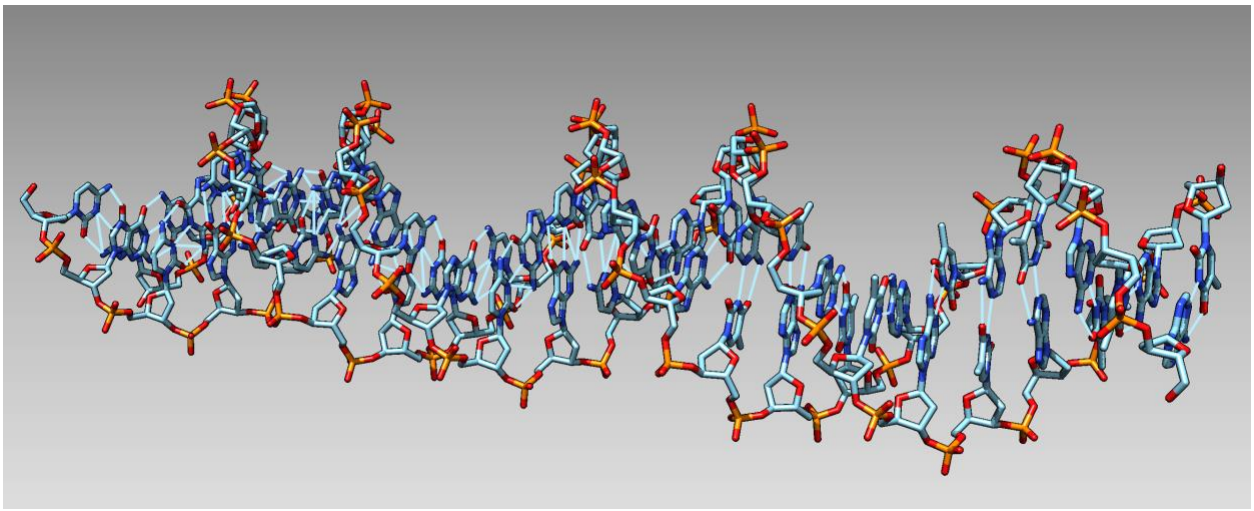


1LWS.pdb

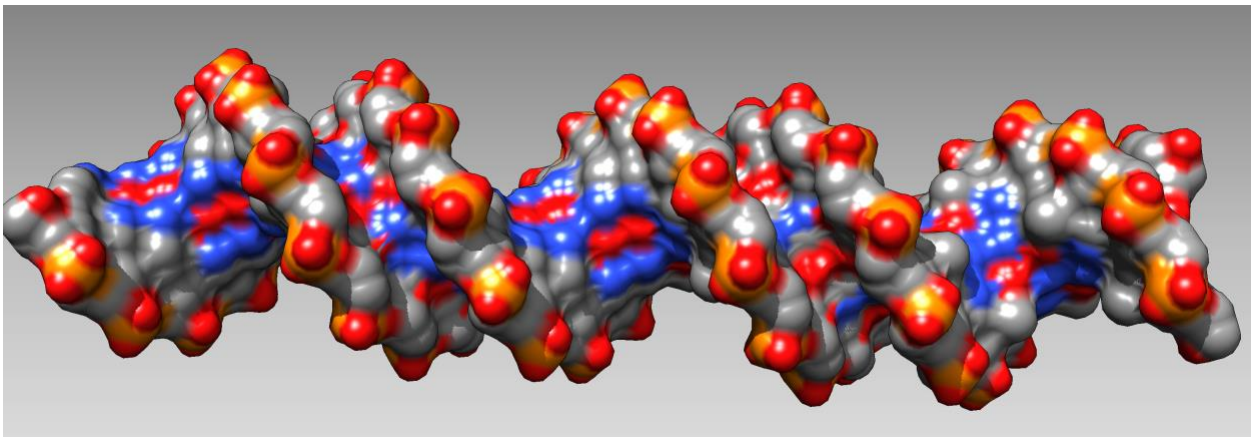
If you have trouble to wrap your head around the major and minor groove, build your own DNA molecule in chimera.
Tools > Structure Editing > Build structure
Select nucleic double helix and B-form enter a sequence (As, Gs, Ts and Cs)



Click apply.

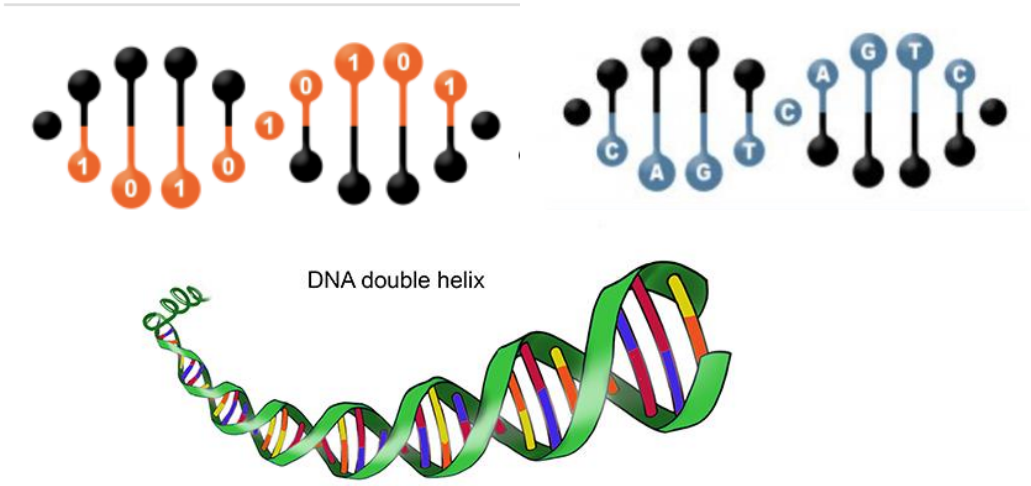


dsDNA with hydrogen bonds



Same but showing the surface.

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



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