

Lab VII: Advanced Building – Linux & OS X

In this lab you will:

- 1) Build a protein structure that consists of three domains, d1, d2 and d3 joined by disordered flexible linkers. There are several missing amino acids at the beginning of d1 and at the end of d3, as well as between d1 and d2 and between d2 and d3.
- 2) Use PDB Scan, PSFGEN and NAMD to add the missing amino acids and build the full structure.

Working directory: \$HOME/Desktop/aps_2016/exercises/lab_VII

NOTE: in this lab, the “>” symbol means that you are in an active terminal window and what follows this symbol is what you type. Make sure you know what directory you are in (> pwd) and what files are in your current directory (> ls). The symbol “vmd>” means you type what follows inside the VMD console (not a Linux terminal).

Directories:

- d1_d2_files/
- d1_d2_d3_files/
- common/

Files:

- d1_d2_files/
 - min_dyn0
- d1_d2_files/output_building/
 - d1.pdb
 - d2.pdb
 - d1_d2_full_sequence.txt
 - rn_d1.pdb
 - rn_d2.pdb
- d1_d2_d3_files/output_building/
 - d3.pdb
 - d1_d2_d3_full_sequence.txt
 - rn_d3.pdb
- common/
 - par_all27_prot_na.inp
 - top_all27_prot_na.inp
 - amino_acid_codes.pdf

Note that in this problem we are no longer giving explicit commands to run PSFGEN or NAMD, etc. If you have trouble, refer to previous labs (II & III) or ASK a question!

1. Build structure d1 + d2 and minimize.

You will first build a structure, d1+d2, by adding in missing amino acid names at the beginning of d1, between d1 and d2 and at the end of d2 (which will also serve as the linker between d2 and d3). The structures will be used in your PSFGEN build script to generate a new d1+d2 starting structure (new PDB and PSF files). You will then create a fixed atom file to restrain the known structural components while allowing for the newly added atoms to move. With these files you will minimize the structures and perform a short (20 ps) dynamics run using NAMD. If this is successful, the file will be ready to add the third domain, d3.

Procedure:

- a) Migrate to the d1_d2_files/ directory and create a new directory named output.
 - i. `> mkdir output`
- b) Migrate to the output_building/ directory and inspect the d1_d2_full_sequence.txt file and compare it to the amino acids listed in the d1.pdb and d2.pdb files. Note the missing amino acids (resid and resname).

QSKELVFN	DG	DPAKTPVTPD	ASRPATFVDA	NGQPVTGNTV	PAMSNGQQVG	50	
TYELDPNIGQ		VTFKPNKTFV	GTPDPVAVQV	SDTNGVPHRA	RYQPTVTK	100	
PTGTGATSTG		PQGVPQTGTF	IF	QGGDPLVP	IDETVEPTFE	DGSKEKTIPG	150
QGTYTIAPDG		TVTFTPKQF	VGNPDPVTVK	RVDKNGTPVT	ATYSPEFTK	V	200
TPTSTDATSN		GIQGQPQKGT	PTF				

Figure 1. d1+d2 full sequence. Missing residues are highlighted in green.

In order to properly join the d1 and d2 structures, the amino acids in the files have to be renumbered to account for those that are missing. For convenience, the renumbered files are provided as rn_d1.pdb and rn_d2.pdb. Note that rn_d1.pdb starts at resid 9 and ends at resid 98 whereas rn_d2.pdb starts at resid 123 and ends at resid 199. You will add the amino acids with resid 1-9, 99-122 and 200-223.

- c) Open *VMD* and change to the working directory (\$HOME/Desktop/aps_2016/exercises/lab_VII/d1_d2_files/output_building/).
- d) Load the rn_d1.pdb and rn_d2.pdb files. Make them two different colors. Note that they are similar and lie on top of one another. Thus, one or both will need to be moved in order to make the linker that joins the two structures. Create representations that highlights amino acids 98 and 123 so you can see where the C terminus of d1 is relative to the N terminus of d2.
 - i. In the Graphical Representations window, select rn_d1.pdb.
 - ii. Click Create Rep
 - iii. Edit Selected Atoms by typing "resid 98" and hit RETURN
 - iv. Select a Coloring Method and Drawing Method (Color ID and Bonds are recommended).

- v. Do the same thing for resid 123 of rn_d2.pdb.

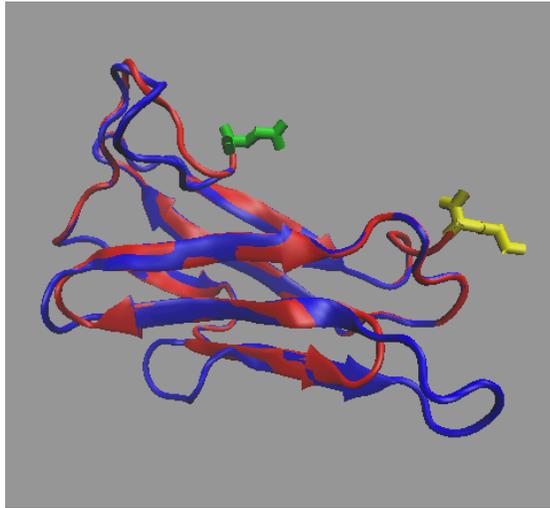


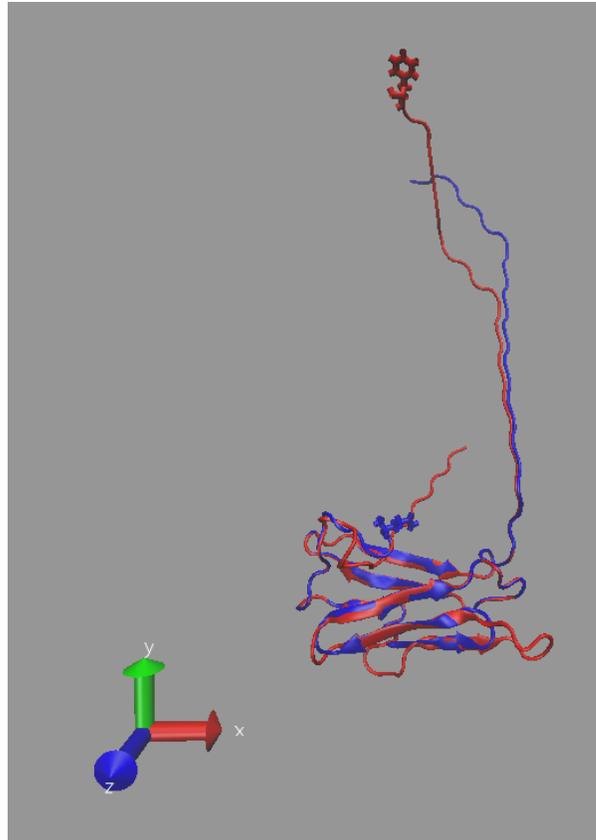
Figure 1. rn_d1.pdb(red)with resid 98 highlighted (yellow) and rn_d2.pdb (blue) with resid 123 highlighted (green).

- e) Add the missing amino acids to rn_d1.pdb to see how the file(s) need to be moved in order to join the two domains.
- f) Using the following command, create two new PDB files that you are going to edit to create PDB files with the added amino acids.
- `> grep "CA" rn_d1.pdb > rn_d1_ca.pdb`
 - `> grep "CA" rn_d2.pdb > rn_d2_ca.pdb`
 - NOTE: The "grep" program "grabs" only those lines in the file that have the characters "CA" and then we redirect that output using ">" to a file called rn_d1_ca.pdb.
 - NOTE2: In other words, the command above makes a new PDB file with just the CA atoms.
- g) Edit your newly created rn_d1_ca.pdb file to add in the missing CA atoms at the beginning and end of domain. Remember that **the coordinates do not matter** in this file. See Figure 2 below for an example of adding amino acids to the system. The amino_acid_codes.pdf document in the common/ directory may be helpful to translate from the one-letter codes in the sequence to the three-letter codes in the pdb file. Name your edited file rn_d1_new_ca.pdb.

ATOM	2	CA	GLN	A	1	-2.834	6.576	1.420	1.00	0.66	A	C
ATOM	2	CA	SER	A	2	-2.834	6.576	1.420	1.00	0.66	A	C
ATOM	2	CA	LYS	A	3	-2.834	6.576	1.420	1.00	0.66	A	C
ATOM	2	CA	GLU	A	4	-2.834	6.576	1.420	1.00	0.66	A	C
ATOM	2	CA	LEU	A	5	-2.834	6.576	1.420	1.00	0.66	A	C
ATOM	2	CA	VAL	A	6	-2.834	6.576	1.420	1.00	0.66	A	C
ATOM	2	CA	PHE	A	7	-2.834	6.576	1.420	1.00	0.66	A	C
ATOM	2	CA	ASN	A	8	-2.834	6.576	1.420	1.00	0.66	A	C
ATOM	2	CA	ASP	A	9	-2.834	6.576	1.420	1.00	0.66	A	C
ATOM	10	CA	GLY	A	10	-5.541	4.440	3.062	1.00	0.71	A	C
ATOM	14	CA	ASP	A	11	-6.080	5.872	6.544	1.00	0.65	A	C

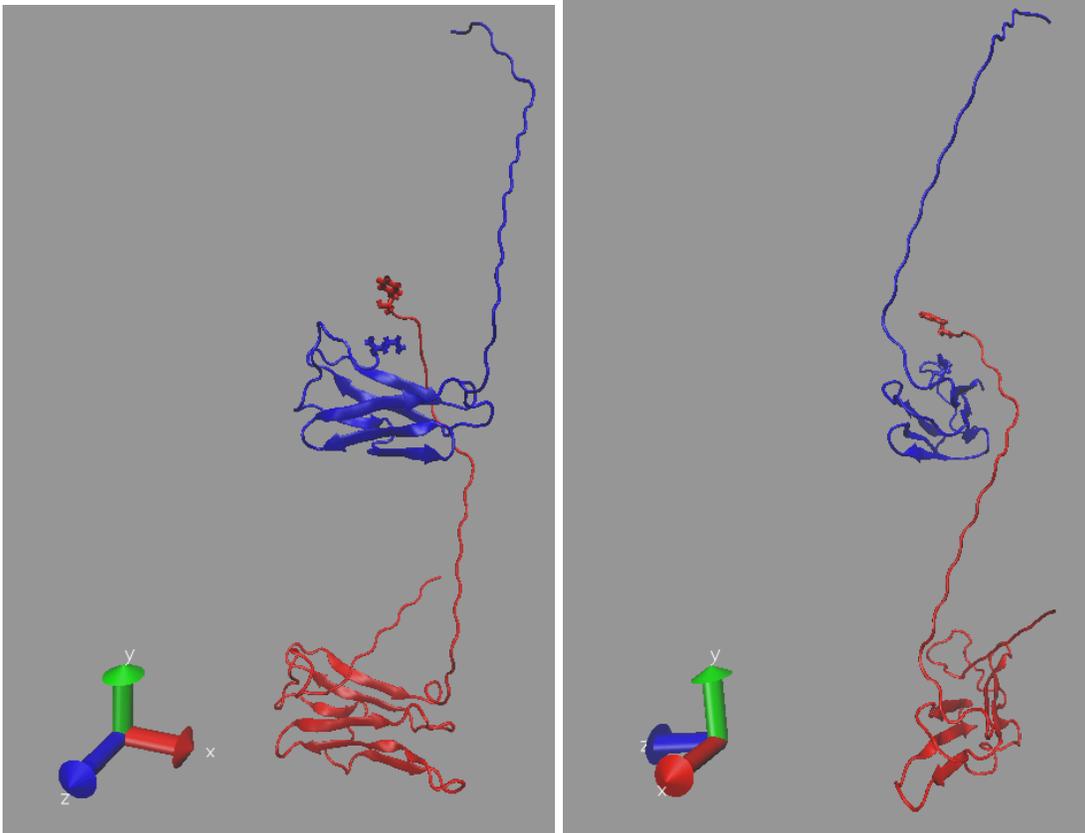
Figure 2. Example of adding missing residues to the sequence file.

- h) Migrate to the `d1_d2_files/` directory. Write a PSFGEN input file to read your edited `rn_d1_new_ca.pdb` file along with the `rn_d1.pdb` file that contains the known coordinates and create a new pdb/psf pair, `rn_d1_new.pdb` (and `psf`). Name this file `build_d1.psfgen`. The only patches needed in the segment definition are `NTER` and `NONE`, since we will be adding `d2` later. (You could use `CTER` here, since PSFGEN will figure out that this isn't the last amino acid residue when we connect it to `d2` later. You can try it both ways and examine the resultant file to see the difference in resid 122.) You will also need the following aliases in the `build_d1.psfgen` input file:
- ```
alias residue HIS HSE
alias atom ILE CD1 CD
alias atom SER HG HG1
```
- Name your new psf/pdb files: `rn_d1_new.psf` and `rn_d1_new.pdb`. Run your script to create the new files in the `output/` directory.
- i) Repeat step 1g for `rn_d2_ca.pdb` to add the missing amino acids at the end of domain `d2`. Name your new file `rn_d2_new_ca.pdb`.
- j) Repeat step 1h to write a PSFGEN input file to create a new pdb/psf pair, `rn_d2_new.pdb` (and `psf`). Name this file `build_d2.psfgen`. The same patches, `NTER` (you could use `NONE` here since `d2` will be between `d1` and `d3` later) and `NONE` (since we will be adding `d3` later), as well as the same aliases can be used.
- k) Open a new VMD session and change to the `/output` directory. Load the new psf/pdb pairs into VMD. Create a new rep to highlight resid 122 from `d1` and resid 123 from `d2`. These are the two residues that need to be closer to each other to make the `d1 + d2` structure. Your structures should look something like the one in Figure 3 below.



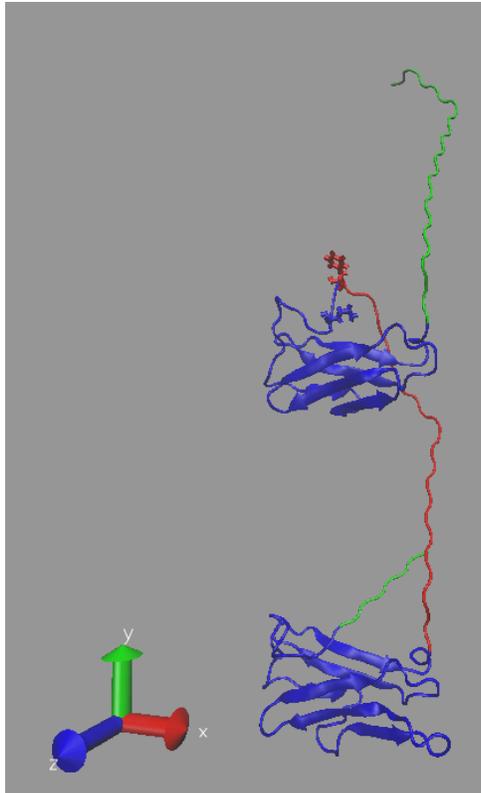
*Figure 3. d1(red) and d2 (blue) with missing residues added.*

- l) Copy `rn_d1_new.pdb` and `rn_d2_new.pdb` to the `output_building/` directory and migrate to that directory in VMD.
- m) **Make sure d2 is the top molecule.** Using the VMD TkConsole, translate d2 along the y axis to get it closer to the new d1. Rotate the molecules (with the mouse) to make sure you are happy with the new location of d2. When you are satisfied, save the new d2 coordinates as `moved_rn_d2.pdb`.
  - i. `vmd> pwd` (If you aren't in the `output_building/` directory, migrate there.)
  - ii. `vmd> set sel [atomselect top all]`
  - iii. `vmd> $sel moveby {0 60 0}`
  - iv. `vmd> $sel writepdb moved_rn_d2_new.pdb`
 Your structure should look something like the one in Figure 4 below.
- n) Use a text editor to combine the two pdb files with the known coordinates, i.e., `rn_d1_new.pdb` and `moved_rn_d2_new.pdb`. Name the file `d1_d2_new_moved.pdb`. Examine this new file. The resid numbers should be sequential from 1 to 223. However, note that the atom numbers start at 1 again at resid 123. This will be fixed in the next step below.



**Figure 4.** *d1 with missing residues (red) and moved d2 with missing residues (blue).*

- o) Migrate to the `d1_d2_files/` directory. Write a new PSFGEN input file to make a new psf/pdb pair of files for the complete d1+d2 structure. Name this file `build_d1_d2.psfgen`. The only patches needed in the segment definition are NTER and NONE (since we are adding d3 later). The segment definition and known coordinate files are now the same, i.e., `d1_d2_new_moved.pdb`. Add the same aliases you used before. Name your new psf/pdb files `d1_d2.psf` and `d1_d2.pdb`. Run your script to create the new files in the `output/` directory. Examine the `d1_d2.pdb` file and note that the atom as well as resid numbers are sequential from beginning to end. Also note that resid 123 no longer has the NTER patch. Compare it to resid 123 in `rn_d2_new.pdb` to see this more clearly.
- p) Open a new VMD session and change to this directory. Load the new psf/pdb files for the d1+d2 structure and view the structure. Identify the amino acids that were added using Create Rep and make them different colors. Create reps to highlight resid 122 and resid 123 where the two structures were joined. See the Figure 5 below for an idea of what the structure should now look like.
- q) The next step is to relax the parts of the structure that have been added, especially the bond between resid 122 and resid 123. This will be done by minimizing this new structure and performing a 20 ps dynamics run in vacuum with fixed atoms.

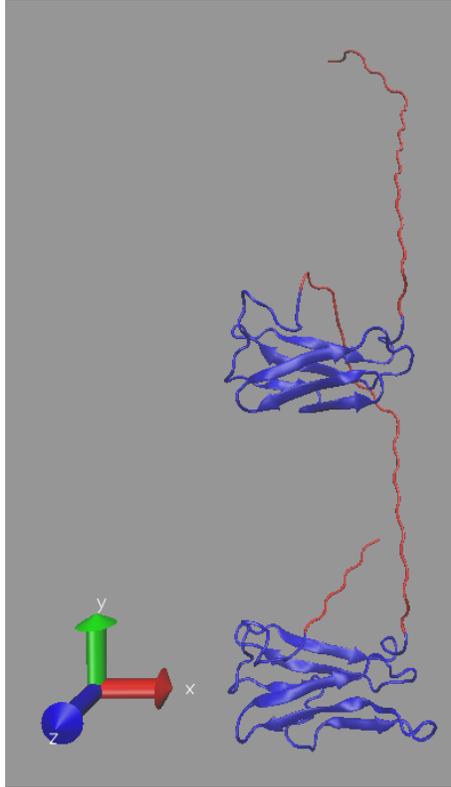


**Figure 5.** *d1 + d2* with added amino acids highlighted (green and red). The Bond Drawing Method is used for resid 122 (red) and resid 123(blue).

- r) Copy the `d1_d2.pdb` and `d1_d2.psf` files to the `output_building/` directory and migrate to that directory in VMD.
- s) Type the commands in the TkConsole to create a fixed at PDB file that restrains the coordinates for the residues that we just added. In other words, all amino acids with resid values corresponding to coordinates for the added amino acids should have BETA values set to zero (0.00). Otherwise they should be set to one (1.00). Name the fixed atom PDB file `d1_d2_fixed_atoms.pdb`.

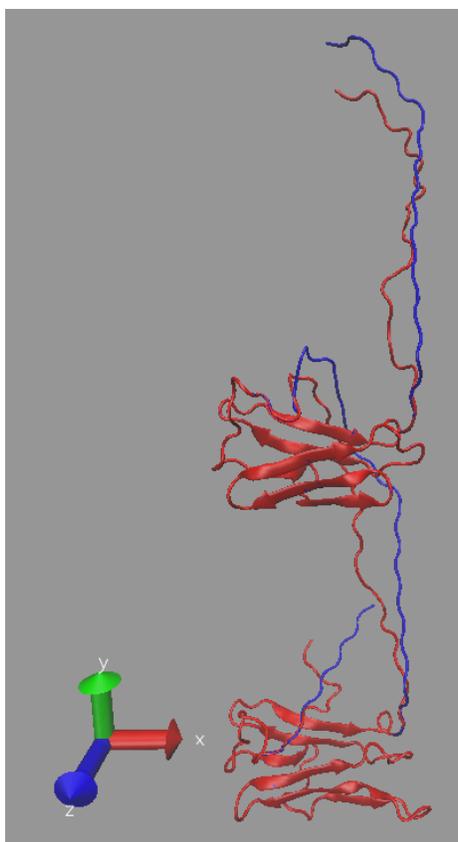
- i. **vmd> pwd** (If you aren't in the `output_building/` directory, migrate there.)
- ii. **vmd> set sel [atomselect top all]**
- iii. **vmd> \$sel set beta 1.00**
- iv. **vmd> set sel1 [atomselect top "(resid < 9) or (resid > 98 and resid < 123) or (resid > 199 and resid < 224)"]**
- v. **vmd> \$sel1 set beta 0.00**
- vi. **vmd> \$sel writepdb d1\_d2\_fixed\_atoms.pdb**

Examine this file to make sure that the proper values are in the BETA column. You can either use a text editor to check the values or load this file into VMD and select the Beta Coloring Method.



**Figure 6.** *d1 + d2 with fixed atoms designated in Beta column highlighted (red).*

- t) We will now minimize this new structure (10000 steps) and perform a 20 ps dynamics run in vacuum with fixed atoms. Migrate to the `d1_d1_files/` directory. Examine the `min_dyn0` input file and make sure that you understand the inputs and outputs. Note that the timestep is 2.0, so 10000 steps is equivalent to 20 ps. Once you understand the inputs, run the script using NAMD. Your final structure will be called `min_dyn0.coor`. Copy this file to the `output_building/` directory and rename it `d1_d2_fixed_min_dyn.pdb`.
- u) Load the final structure into VMD and compare it to the structure before minimization and dynamics. Close VMD when finished.



**Figure 7.**  $d1 + d2$  before (blue) and after (red) fixed-atom minimization and dynamics.

## 2. Build structure d1 + d2 + d3 and minimize.

You will build the full structure d1+d2+d3 by adding in missing amino acid names at the end of d3 and then using PSFGEN to combine the complete d3 structure your newly-built d1+d2 structure. You will minimize the entire structure and perform a 20 ps dynamics run with fixed atoms using NAMD. This will be followed by an unrestrained minimization and 20 ps dynamics run. The structure is then ready for simulation using SASSIE.

Procedure:

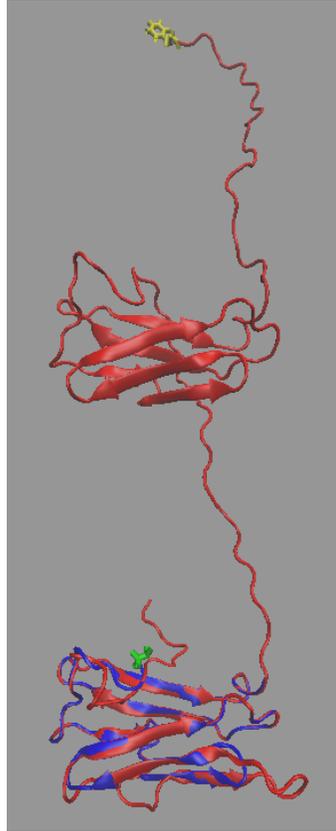
- a) Migrate to the d1\_d2\_d3\_files/ directory and create a new directory named output.
  - i. > **mkdir output**
- b) Migrate to the output\_building/ directory. Inspect the d1\_d2\_d3\_full\_sequence.txt file and compare it to the amino acids listed in the d3.pdb and your recently-created d1\_d2\_fixed\_min\_dyn.pdb files. (Copy the latter file to the current directory.) Note the missing amino acids (resid and rename).

|                   |                   |             |            |            |     |
|-------------------|-------------------|-------------|------------|------------|-----|
| QSKELVFNDG        | DKAKTPVTPD        | ASRPATFVDA  | NGQPVTGNTV | PAMSNQQQVG | 50  |
| TYELDPNTGQ        | VTFKPNKTFV        | GTPDPVAVQV  | SDTNGVPHRA | RYQPTVTKVT | 100 |
| PTGTGATSTG        | PQGVPQTGTP        | TFQGGDPLVP  | IDETVEPTFE | DGSKEKTIPG | 150 |
| QGTYYIAPDG        | TVTFTPDKQF        | VGNPDPVTVK  | RVDKNGTPVT | ATYSPEFTKV | 200 |
| TPTSTDATSN        | GIQGQPQKGT        | PTFTEGNPLV  | PIDDTKPMTF | EDGQSTKTVP | 250 |
| GVGEYSINPD        | GSITFTPDKK        | YVGTPNPVTV  | KRVDKNGTEV | TATYTPTVTK | 300 |
| <b>VTPTSTDATS</b> | <b>NGIQGQPQKG</b> | <b>TPTF</b> |            |            |     |

*Figure 8. d1+d2+d3 full sequence. Missing residues are highlighted in green.*

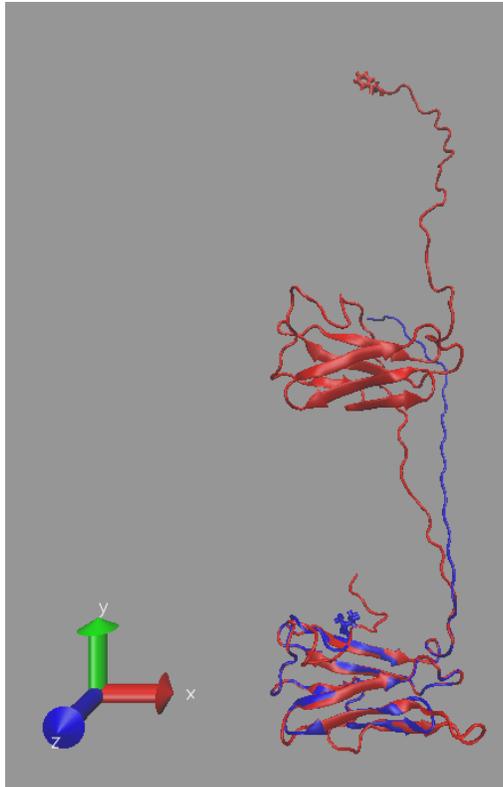
Once again, a renumbered file, m\_d3.pdb is provided for your convenience.

- c) Open VMD and change to the working directory (\$HOME/Desktop/aps\_2016/exercises/lab\_VII/d1\_d2\_d3\_files/output\_building/).
- d) Load the d1\_d2\_fixed\_min\_dyn.pdb and m\_d3.pdb files to determine where the d3 domain is relative to the d1+d2 domains.



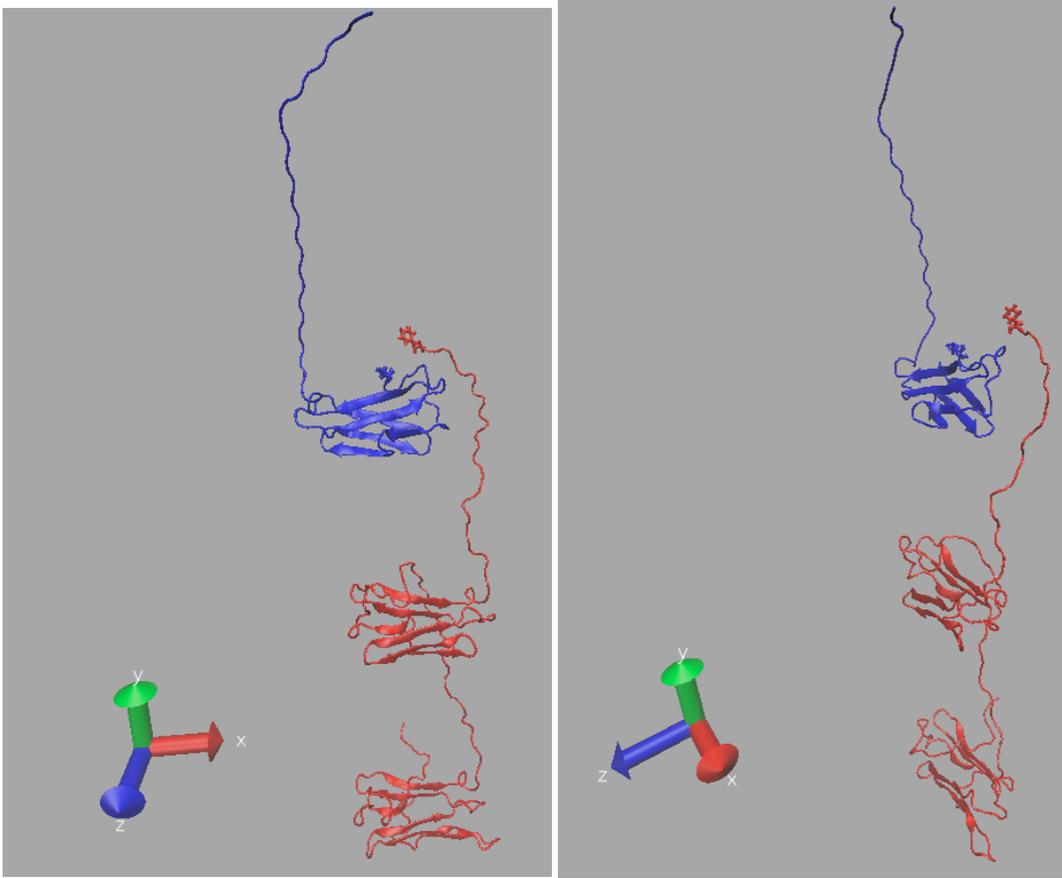
**Figure 9.** *d1\_d2\_fixed\_min\_dyn.pdb* (red) with resid 223 highlighted (yellow) and *rn\_d3.pdb* (blue) with resid 224 highlighted (green).

- e) Repeat steps 1e) through 1h) above to add the missing amino acids to *rn\_d3.pdb*. This time, the patches needed in the segment definition are NTER (NONE can be used here as well) and CTER, since d3 is the last part of the full structure. Use the same aliases as before. Run your PSFGEN script to create the new files in the output/ directory and name them *rn\_d3\_new.psf* and *rn\_d3\_new.pdb*.
- f) Examine the new psf/pdb pair of files along with the recently-created *d1\_d2\_fixed\_min\_dyn.pdb* file in VMD to see how the structures need to be moved in order to join the two domains. Creating a new rep to highlight the C-terminal resid in *d1\_d2\_fixed\_min\_dyn.pdb* and the N-terminal resid in *rn\_d3\_new.pdb* may be helpful.



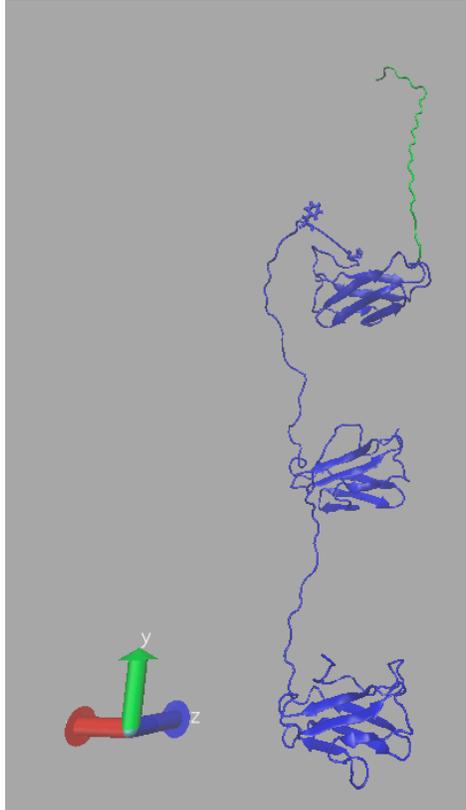
**Figure 10.** *d1 + d2 (red) and d3 (blue) with missing residues added.*

- g) Move `rn_d3_new.pdb` to move these two amino acids closer to each other. Name the new file `rn_d3_new_moved.pdb`. Note that you may have to **rotate** and **translate** d3 in order to avoid overlap.



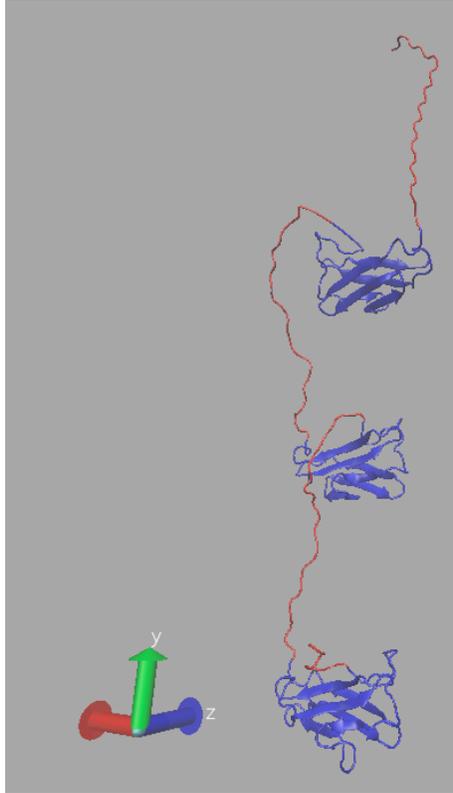
**Figure 11.** *d1+ d2 (red) and moved d3 with missing residues (blue).*

- h) Use a text editor to combine the two pdb files with the known coordinates, i.e., `d1_d2_fixed_min_dyn.pdb` and `rn_d3_new_moved.pdb`. Name the file `d1_d2_d3_new_moved.pdb`.
- i) Write a new PSFGEN input file to make a new psf/pdb pair of files for the complete `d1+d2+d3` structure. Name this file `build_d1_d2_d3.psfgen`. Name the new files `d1_d2_d3.pdb` and `d1_d2_d3.psf`.
- j) Copy the `d1_d2_d3.pdb` and `d1_d2_d3.psf` files to the `output_building/` directory and migrate to that directory in VMD.



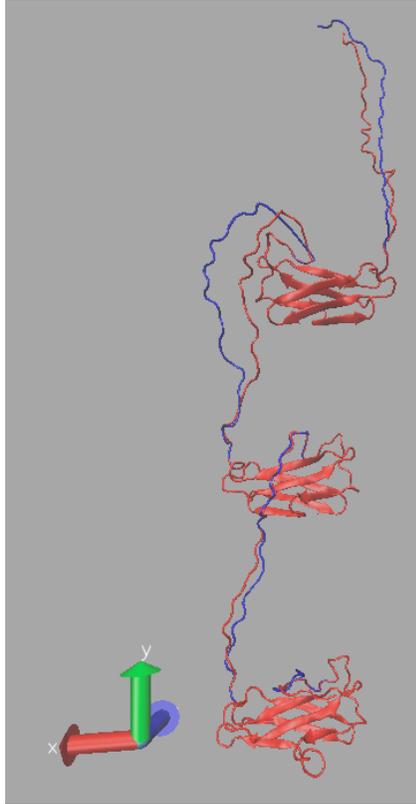
**Figure 12.** *d1 + d2 + d3 with added amino acids highlighted (green). The Bond Drawing Method is used for resid 223 and resid 224.*

- k) Create a fixed atom PDB file from `d1_d2_d3.pdb` using VMD in which the BETA values for **all** of the added amino acids (in this section as well as section 1 above) are equal to 0.00. Follow the example commands in 1s above. You can create several different selections in VMD and then write the final PDB file after you have set BETA to 0.00 for each of them. Name the new PDB file `d1_d2_d3_fixed_atoms.pdb`.



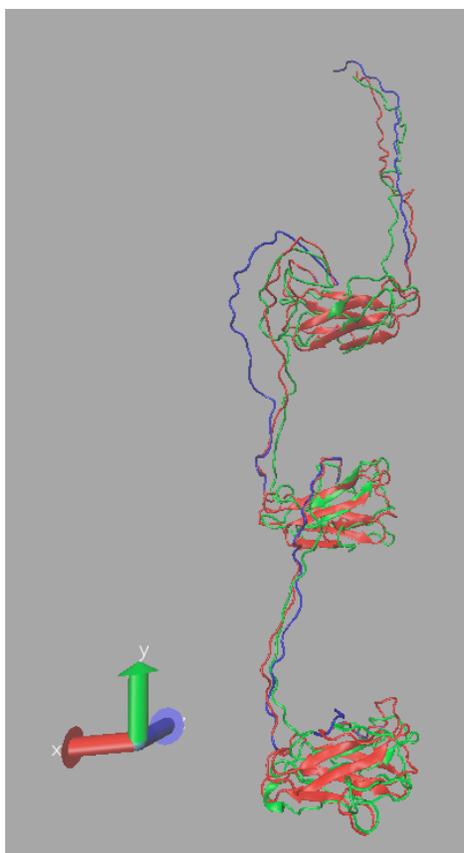
**Figure 13.**  $d1 + d2 + 3$  with fixed atoms designated in Beta column highlighted (red).

- l) Minimize the fixed atom structure and perform a 20 ps dynamics run in vacuum with fixed atoms. You can use min\_dyn0 from section 1t above as a template for your NAMD script. If your script runs successfully, then you are ready to remove the restraints.
- m) Copy min\_dyn0.coor to d1\_d2\_d3\_fixed\_min\_dyn.pdb and put it in the output\_building/ directory. Examine the d1\_d2\_d3\_fixed\_min\_dyn.pdb structure with VMD and compare it to d1\_d2\_d3\_fixed\_atoms.pdb.



**Figure 14.** *d1 + d2 + d3 before (blue) and after (red) fixed-atom minimization and dynamics.*

- n) Copy the min\_dyn0 NAMD script to min\_dyn1. Edit min\_dyn1 to remove or comment out the fixed atom commands. Run the simulation again starting with d1\_d2\_d3\_fixed\_min\_dyn.pdb to minimize the structure for 5000 steps and then perform a 20 ps dynamics run with no restraints.
- o) Copy the resultant min\_dyn1.coor file to the output\_building/ directory and name it d1\_d2\_d3\_min\_dyn.pdb. Examine the new structure with VMD and compare it to d1\_d2\_d3\_fixed\_min\_dyn.pdb and d1\_d2\_d3\_fixed\_atom.pdb.



**Figure 15.** *d1 + d2 + d3 before (blue) and after (red) fixed-atom minimization and dynamics and after unrestrained minimization and dynamics (green).*

- p) Your structure is now ready for further simulation using, for example, the SASSIE Monomer Monte Carlo module.