M01: Virtual Lab-Introduction of PDB and PDBx/mmCIF files

Problems

Navigation to the following PDB entry: **https://www.rcsb.org/structure/7PFL**.

Using the knowledge gained from the virtual lab, compile a document describing the PDBx/mmCIF files associated with this PDB entry:

1. **What features do you see in the file? What features are difficult to decipher in the file?**
2. **Are chains and models included? If so, how are they indicated?**
3. **Does the structure display multiple conformations? How would you be able to tell?**
4. **Are there features, such as asymmetric units or biological assemblies that are interesting? How are they indicated in the file?**

Compile all code, screenshots, and outputs from the virtual lab into a single document (you may use a Word document, PDF, Google tools, or create your own **Jupyter notebook** and submit it here by typing the URL or attaching the file.

Q&A

What features do you see in the file? What features are difficult to decipher in the file?

Features presented in 7PFL file.

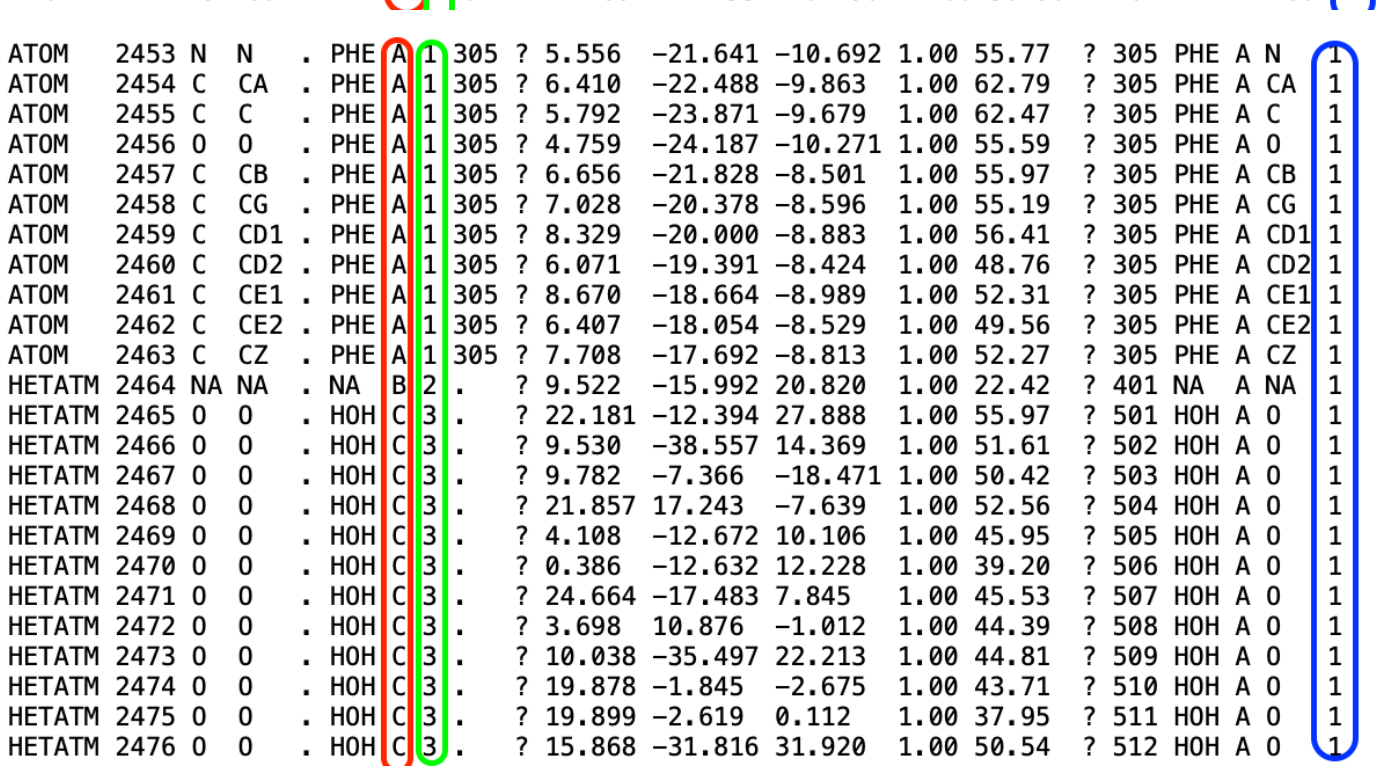
- Basic database information of the protein
- Author info
- Citation
- Asymmetric unit, Unit cell parameters
- Entities
- Reference structure
- Chemical description of ligands
- Experimental info (method)
- Refinement info
- Atomic coordinates
- Structure assembly
- Validation

It is really difficult to decipher atom-level data to protein 3-D structure.

Are chains and models included? If so, how are they indicated?

Chains and model information was included in the cif file. **_atom_site.label_asym_id** and **_atom_site.label_entity_id** indicated the transition of reading from chain A to chain B and then to chain C, but there is only one unique chain(chain A) in this protein. And only one model was recorded in the file, which was indicated by **_atom_site.pdbx_PDB_model_num**.

```
#
loop_
_atom_site.group_PDB
_atom_site.id
_atom_site.type_symbol
_atom_site.label_atom_id
_atom_site.label_alt_id
_atom_site.label_comp_id
_atom_site.label_asym_id
_atom_site.label_entity_id
_atom_site.label_seq_id
_atom_site.pdbx_PDB_ins_code
_atom_site.Cartn_x
_atom_site.Cartn_y
_atom_site.Cartn_z
_atom_site.occupancy
_atom_site.B_iso_or_equiv
_atom_site.pdbx_formal_charge
_atom_site.auth_seq_id
_atom_site.auth_comp_id
_atom_site.auth_asym_id
_atom_site.auth_atom_id
_atom_site.pdbx_PDB_model_num
ATOM   1    N  N   . SER A 1 1   ? -3.231 -7.482  -17.009 1.00 40.61  ? 1   SER A N   1
ATOM   2    C  CA  . SER A 1 1   ? -2.514 -6.334  -16.473 1.00 32.76  ? 1   SER A CA  1
ATOM   3    C  C   . SER A 1 1   ? -2.760 -6.189  -14.976 1.00 38.09  ? 1   SER A C   1
ATOM   4    O  O   . SER A 1 1   ? -3.347 -7.070  -14.346 1.00 37.15  ? 1   SER A O   1
ATOM   5    C  CB  . SER A 1 1   ? -2.928 -5.054  -17.204 1.00 44.33  ? 1   SER A CB  1
ATOM   6    O  OG  . SER A 1 1   ? -2.213 -3.932  -16.717 1.00 53.44  ? 1   SER A OG  1
ATOM   7    N  N   . GLY A 1 2   ? -2.318 -5.075  -14.415 1.00 36.15  ? 2   GLY A N   1
ATOM   8    C  CA  . GLY A 1 2   ? -2.421 -4.863  -12.980 1.00 32.34  ? 2   GLY A CA  1
ATOM   9    C  C   . GLY A 1 2   ? -1.173 -5.313  -12.250 1.00 32.68  ? 2   GLY A C   1
ATOM   10   O  O   . GLY A 1 2   ? -0.466 -6.231  -12.666 1.00 34.95  ? 2   GLY A O   1
ATOM   11   N  N   . PHE A 1 3   ? -0.900 -4.646  -11.130 1.00 30.63  ? 3   PHE A N   1
ATOM   12   C  CA  . PHE A 1 3   ? 0.304  -4.910  -10.352 1.00 30.17  ? 3   PHE A CA  1
ATOM   13   C  C   . PHE A 1 3   ? -0.053 -4.973  -8.875  1.00 27.89  ? 3   PHE A C   1
ATOM   14   O  O   . PHE A 1 3   ? -0.659 -4.041  -8.339  1.00 34.53  ? 3   PHE A O   1
ATOM   15   C  CB  . PHE A 1 3   ? 1.370  -3.835  -10.608 1.00 29.29  ? 3   PHE A CB  1
ATOM   16   C  CG  . PHE A 1 3   ? 2.769  -4.294  -10.319 1.00 34.45  ? 3   PHE A CG  1
ATOM   17   C  CD1 . PHE A 1 3   ? 3.437  -5.124  -11.209 1.00 34.19  ? 3   PHE A CD1 1
ATOM   18   C  CD2 . PHE A 1 3   ? 3.416  -3.902  -9.158  1.00 34.86  ? 3   PHE A CD2 1
ATOM   19   C  CE1 . PHE A 1 3   ? 4.723  -5.555  -10.946 1.00 28.71  ? 3   PHE A CE1 1
ATOM   20   C  CE2 . PHE A 1 3   ? 4.705  -4.328  -8.889  1.00 31.54  ? 3   PHE A CE2 1
ATOM   21   C  CZ  . PHE A 1 3   ? 5.359  -5.156  -9.782  1.00 30.72  ? 3   PHE A CZ  1
ATOM   22   N  N   . ARG A 1 4   ? 0.330  -6.069  -8.220  1.00 29.26  ? 4   ARG A N   1
ATOM   23   C  CA  . ARG A 1 4   ? -0.016 -6.322  -6.828  1.00 22.84  ? 4   ARG A CA  1
ATOM   24   C  C   . ARG A 1 4   ? 1.209  -6.796  -6.061  1.00 18.80  ? 4   ARG A C   1
ATOM   25   O  O   . ARG A 1 4   ? 2.084  -7.467  -6.613  1.00 23.84  ? 4   ARG A O   1
ATOM   26   C  CB  . ARG A 1 4   ? -1.123 -7.384  -6.712  1.00 22.02  ? 4   ARG A CB  1
ATOM   27   C  CG  . ARG A 1 4   ? -2.527 -6.864  -6.935  1.00 25.79  ? 4   ARG A CG  1
ATOM   28   C  CD  . ARG A 1 4   ? -3.068 -6.208  -5.671  1.00 31.87  ? 4   ARG A CD  1
ATOM   29   N  NE  . ARG A 1 4   ? -4.504 -5.956  -5.769  1.00 39.42  ? 4   ARG A NE  1
ATOM   30   C  CZ  . ARG A 1 4   ? -5.262 -5.545  -4.758  1.00 42.58  ? 4   ARG A CZ  1
ATOM   31   N  NH1 . ARG A 1 4   ? -4.724 -5.344  -3.562  1.00 33.57  ? 4   ARG A NH1 1
ATOM   32   N  NH2 . ARG A 1 4   ? -6.560 -5.336  -4.943  1.00 35.52  ? 4   ARG A NH2 1
ATOM   33   N  N   . LYS A 1 5   ? 1.261  -6.448  -4.777  1.00 24.01  ? 5   LYS A N   1
ATOM   34   C  CA  . LYS A 1 5   ? 2.252  -7.041  -3.889  1.00 21.41  ? 5   LYS A CA  1
ATOM   35   C  C   . LYS A 1 5   ? 1.854  -8.490  -3.657  1.00 27.88  ? 5   LYS A C   1
ATOM   36   O  O   . LYS A 1 5   ? 0.847  -8.773  -3.002  1.00 33.74  ? 5   LYS A O   1
ATOM   37   C  CB  . LYS A 1 5   ? 2.346  -6.268  -2.576  1.00 28.19  ? 5   LYS A CB  1
ATOM   38   C  CG  . LYS A 1 5   ? 3.637  -6.536  -1.812  1.00 31.82  ? 5   LYS A CG  1
ATOM   39   C  CD  . LYS A 1 5   ? 3.660  -5.813  -0.477  1.00 35.58  ? 5   LYS A CD  1
ATOM   40   C  CE  . LYS A 1 5   ? 3.707  -4.306  -0.650  1.00 37.18  ? 5   LYS A CE  1
ATOM   41   N  NZ  . LYS A 1 5   ? 5.010  -3.856  -1.201  1.00 38.92  ? 5   LYS A NZ  1
ATOM   42   N  N   . MET A 1 6   ? 2.630  -9.406  -4.219  1.00 23.92  ? 6   MET A N   1
ATOM   43   C  CA  . MET A 1 6   ? 2.230  -10.794 -4.381  1.00 24.25  ? 6   MET A CA  1
ATOM   44   C  C   . MET A 1 6   ? 3.230  -11.691 -3.671  1.00 20.02  ? 6   MET A C   1
ATOM   45   O  O   . MET A 1 6   ? 4.441  -11.547 -3.864  1.00 22.12  ? 6   MET A O   1
ATOM   46   C  CB  . MET A 1 6   ? 2.154  -11.132 -5.870  1.00 25.19  ? 6   MET A CB  1
ATOM   47   C  CG  . MET A 1 6   ? 1.789  -12.537 -6.196  1.00 30.96  ? 6   MET A CG  1
```



```
ATOM   2453 N  N   . PHE A 1 305 ? 5.556  -21.641 -10.692 1.00 55.77  ? 305 PHE A N   1
ATOM   2454 C  CA  . PHE A 1 305 ? 6.410  -22.488 -9.863  1.00 62.79  ? 305 PHE A CA  1
ATOM   2455 C  C   . PHE A 1 305 ? 5.792  -23.871 -9.679  1.00 62.47  ? 305 PHE A C   1
ATOM   2456 O  O   . PHE A 1 305 ? 4.759  -24.187 -10.271 1.00 55.59  ? 305 PHE A O   1
ATOM   2457 C  CB  . PHE A 1 305 ? 6.656  -21.828 -8.501  1.00 55.97  ? 305 PHE A CB  1
ATOM   2458 C  CG  . PHE A 1 305 ? 7.028  -20.378 -8.596  1.00 55.19  ? 305 PHE A CG  1
ATOM   2459 C  CD1 . PHE A 1 305 ? 8.329  -20.000 -8.883  1.00 56.41  ? 305 PHE A CD1 1
ATOM   2460 C  CD2 . PHE A 1 305 ? 6.071  -19.391 -8.424  1.00 48.76  ? 305 PHE A CD2 1
ATOM   2461 C  CE1 . PHE A 1 305 ? 8.670  -18.664 -8.989  1.00 52.31  ? 305 PHE A CE1 1
ATOM   2462 C  CE2 . PHE A 1 305 ? 6.407  -18.054 -8.529  1.00 49.56  ? 305 PHE A CE2 1
ATOM   2463 C  CZ  . PHE A 1 305 ? 7.708  -17.692 -8.813  1.00 52.27  ? 305 PHE A CZ  1
HETATM 2464 NA NA  . NA  B 2 .   ? 9.522  -15.992 20.820  1.00 22.42  ? 401 NA  A NA  1
HETATM 2465 O  O   . HOH C 3 .   ? 22.181 -12.394 27.888  1.00 55.97  ? 501 HOH A O   1
HETATM 2466 O  O   . HOH C 3 .   ? 9.530  -38.557 14.369  1.00 51.61  ? 502 HOH A O   1
HETATM 2467 O  O   . HOH C 3 .   ? 9.782  -7.366  -18.471 1.00 50.42  ? 503 HOH A O   1
HETATM 2468 O  O   . HOH C 3 .   ? 21.857 17.243  -7.639  1.00 52.56  ? 504 HOH A O   1
HETATM 2469 O  O   . HOH C 3 .   ? 4.108  -12.672 10.106  1.00 45.95  ? 505 HOH A O   1
HETATM 2470 O  O   . HOH C 3 .   ? 0.386  -12.632 12.228  1.00 39.20  ? 506 HOH A O   1
HETATM 2471 O  O   . HOH C 3 .   ? 24.664 -17.483 7.845   1.00 45.53  ? 507 HOH A O   1
HETATM 2472 O  O   . HOH C 3 .   ? 3.698  10.876  -1.012  1.00 44.39  ? 508 HOH A O   1
HETATM 2473 O  O   . HOH C 3 .   ? 10.038 -35.497 22.213  1.00 44.81  ? 509 HOH A O   1
HETATM 2474 O  O   . HOH C 3 .   ? 19.878 -1.845  -2.675  1.00 43.71  ? 510 HOH A O   1
HETATM 2475 O  O   . HOH C 3 .   ? 19.899 -2.619  0.112   1.00 37.95  ? 511 HOH A O   1
HETATM 2476 O  O   . HOH C 3 .   ? 15.868 -31.816 31.920  1.00 50.54  ? 512 HOH A O   1
```

Does the structure display multiple conformations? How would you be able to tell?

Yes the structure shows multiple conformations, scientists used occupancy to estimate the amount of each conformation that observed in the crystal. In the cif **_atom_site.occupancy** display the amount of each conformation and **_atom_site.label_alt_id** shows the alternate conformations. For example, occupancy = 1.00 means that the atom is found in all of the molecules in the same place in the crystal. However, if there is another conformation, a metal ion only binds half of the molecules in the crystal, electron density map would only present weak image, and the occupancy would be 0.5.

```
ATOM   164  C  C   .  GLY A 1 23  ? 7.143   -25.787 28.831  1.00 49.16  ? 23  GLY A C   1
ATOM   165  O  O   .  GLY A 1 23  ? 6.203   -26.317 28.228  1.00 52.33  ? 23  GLY A O   1
ATOM   166  N  N   .  THR A 1 24  ? 7.101   -24.526 29.257  1.00 55.28  ? 24  THR A N   1
ATOM   167  C  CA  .  THR A 1 24  ? 5.920   -23.685 29.113  1.00 49.18  ? 24  THR A CA  1
ATOM   168  C  C   .  THR A 1 24  ? 5.943   -22.830 27.854  1.00 47.16  ? 24  THR A C   1
ATOM   169  O  O   .  THR A 1 24  ? 5.019   -22.037 27.644  1.00 49.97  ? 24  THR A O   1
ATOM   170  C  CB  .  THR A 1 24  ? 5.771   -22.774 30.337  1.00 63.18  ? 24  THR A CB  1
ATOM   171  O  OG1 .  THR A 1 24  ? 4.586   -21.979 30.205  1.00 76.37  ? 24  THR A OG1 1
ATOM   172  C  CG2 .  THR A 1 24  ? 6.982   -21.857 30.463  1.00 57.16  ? 24  THR A CG2 1
ATOM   173  N  N   .  THR A 1 25  ? 6.963   -22.966 27.017  1.00 43.37  ? 25  THR A N   1
ATOM   174  C  CA  .  THR A 1 25  ? 7.134   -22.128 25.839  1.00 42.67  ? 25  THR A CA  1
ATOM   175  C  C   .  THR A 1 25  ? 6.816   -22.930 24.586  1.00 30.34  ? 25  THR A C   1
ATOM   176  O  O   .  THR A 1 25  ? 7.324   -24.043 24.412  1.00 38.93  ? 25  THR A O   1
ATOM   177  C  CB  .  THR A 1 25  ? 8.561   -21.584 25.768  1.00 39.76  ? 25  THR A CB  1
ATOM   178  O  OG1 .  THR A 1 25  ? 8.860   -20.879 26.980  1.00 41.33  ? 25  THR A OG1 1
ATOM   179  C  CG2 .  THR A 1 25  ? 8.726   -20.653 24.572  1.00 34.53  ? 25  THR A CG2 1
ATOM   180  N  N   A  THR A 1 26  ? 5.988   -22.360 23.713  0.49 33.73  ? 26  THR A N   1
ATOM   181  N  N   B  THR A 1 26  ? 5.965   -22.374 23.731  0.51 33.71  ? 26  THR A N   1
ATOM   182  C  CA  A  THR A 1 26  ? 5.608   -22.992 22.455  0.49 32.76  ? 26  THR A CA  1
ATOM   183  C  CA  B  THR A 1 26  ? 5.637   -22.989 22.455  0.51 32.75  ? 26  THR A CA  1
ATOM   184  C  C   A  THR A 1 26  ? 5.793   -22.001 21.314  0.49 29.53  ? 26  THR A C   1
ATOM   185  C  C   B  THR A 1 26  ? 5.869   -21.978 21.342  0.51 29.53  ? 26  THR A C   1
ATOM   186  O  O   A  THR A 1 26  ? 5.373   -20.845 21.416  0.49 29.44  ? 26  THR A O   1
ATOM   187  O  O   B  THR A 1 26  ? 5.582   -20.786 21.496  0.51 29.36  ? 26  THR A O   1
ATOM   188  C  CB  A  THR A 1 26  ? 4.155   -23.487 22.502  0.49 34.08  ? 26  THR A CB  1
ATOM   189  C  CB  B  THR A 1 26  ? 4.184   -23.508 22.430  0.51 34.08  ? 26  THR A CB  1
ATOM   190  O  OG1 A  THR A 1 26  ? 4.055   -24.595 23.404  0.49 31.71  ? 26  THR A OG1 1
ATOM   191  O  OG1 B  THR A 1 26  ? 3.936   -24.196 21.196  0.51 36.79  ? 26  THR A OG1 1
ATOM   192  C  CG2 A  THR A 1 26  ? 3.685   -23.923 21.122  0.49 36.29  ? 26  THR A CG2 1
ATOM   193  C  CG2 B  THR A 1 26  ? 3.183   -22.367 22.587  0.51 32.38  ? 26  THR A CG2 1
ATOM   194  N  N   .  LEU A 1 27  ? 6.425   -22.454 20.233  1.00 26.19  ? 27  LEU A N   1
ATOM   195  C  CA  .  LEU A 1 27  ? 6.613   -21.633 19.045  1.00 27.94  ? 27  LEU A CA  1
ATOM   196  C  C   .  LEU A 1 27  ? 6.610   -22.556 17.831  1.00 24.65  ? 27  LEU A C   1
ATOM   197  O  O   .  LEU A 1 27  ? 6.294   -23.745 17.936  1.00 24.88  ? 27  LEU A O   1
ATOM   198  C  CB  .  LEU A 1 27  ? 7.891   -20.789 19.152  1.00 22.26  ? 27  LEU A CB  1
ATOM   199  C  CG  .  LEU A 1 27  ? 9.198   -21.381 19.681  1.00 26.76  ? 27  LEU A CG  1
```

Are there features, such as asymmetric units or biological assemblies that are interesting? How are they indicated in the file?

The asymmetric unit is the smallest portion of a crystal structure to which symmetry operations can be applied in order to generate the complete unit cell. The biological assembly is believed to be the functional form of the molecule. Based on the 3D viewer on PDB, it seems like the asymmetric unit is only a portion of the biological assembly. It looks like that the biological assembly was built by two asymmetric units. In mmCIF format files, details about the structural elements that form each biological assembly are found in the **pdbx_struct_assembly**, **pdbx_struct_assembly_gen** and **pdbx_struct_oper_list categories**.

```
#
_pdbx_struct_assembly.id                    1
_pdbx_struct_assembly.details               author_and_software_defined_assembly
_pdbx_struct_assembly.method_details        PISA
_pdbx_struct_assembly.oligomeric_details    dimeric
_pdbx_struct_assembly.oligomeric_count      2
#
_pdbx_struct_assembly_gen.assembly_id       1
_pdbx_struct_assembly_gen.oper_expression   1,2
_pdbx_struct_assembly_gen.asym_id_list      A,B,C
#
loop_
_pdbx_struct_assembly_prop.biol_id
_pdbx_struct_assembly_prop.type
_pdbx_struct_assembly_prop.value
_pdbx_struct_assembly_prop.details
1 'ABSA (A^2)' 3340  ?
1 MORE         -37   ?
1 'SSA (A^2)'  25190 ?
#
loop_
_pdbx_struct_oper_list.id
_pdbx_struct_oper_list.type
_pdbx_struct_oper_list.name
_pdbx_struct_oper_list.symmetry_operation
_pdbx_struct_oper_list.matrix[1][1]
_pdbx_struct_oper_list.matrix[1][2]
_pdbx_struct_oper_list.matrix[1][3]
_pdbx_struct_oper_list.vector[1]
_pdbx_struct_oper_list.matrix[2][1]
_pdbx_struct_oper_list.matrix[2][2]
_pdbx_struct_oper_list.matrix[2][3]
_pdbx_struct_oper_list.vector[2]
_pdbx_struct_oper_list.matrix[3][1]
_pdbx_struct_oper_list.matrix[3][2]
_pdbx_struct_oper_list.matrix[3][3]
_pdbx_struct_oper_list.vector[3]
1 'identity operation'         1_555 x,y,z   1.0000000000  0.0000000000 0.0000000000 0.0000000000 0.0000000000 1.0000000000
0.0000000000 0.0000000000 0.0000000000 0.0000000000 1.0000000000  0.0000000000
2 'crystal symmetry operation' 2_555 -x,y,-z -1.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000 1.0000000000
0.0000000000 0.0000000000 0.0000000000 0.0000000000 -1.0000000000 0.0000000000
#
```

In [ ]: