1

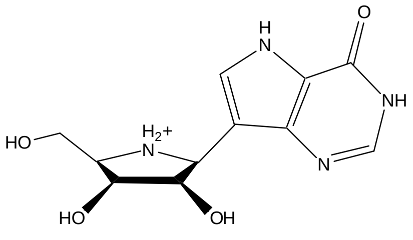
Importing and Cleaning Protein Crystal Structures

Overview

This exercise describes how to open Protein Data Bank (PDB) files, view protein structures and sequences and build proteins and peptides from their sequence. You'll learn how to:

- open and view PDB molecules
- inspect and correct hetero groups
- add hydrogen atoms
- balance charges

Background



immucillin-H

In this exercise, we will investigate the 1.75 Angstrom resolution crystal structure (`1G2O`) of *Mycobacterium tuberculosis* purine nucleoside phosphorylase (TB-PNP) complexed with the transition-state analogue Immucillin-H (ImmH) and phosphate.[1]

TB kills millions each year and it has been estimated that one-third of the world's population is infected with latent TB.

Analysis of the genome sequence of *Mycobacterium tuberculosis* (TB) predicted that it expresses purine nucleoside phosphorylase (PNP) which catalyzes the phosphorolysis of purine nucleotides to purine bases and deoxynucleosides to (deoxy)ribosyl 1-phosphate. PNP recycles purines, a crucial function for organisms that do not synthesize purines. It is thought that inhibition of this enzyme will cause physiological in TB that will cause the bacterium to enter a latent state thereby preventing the development of active TB in infected individuals[2].

One problem with this approach is that humans also express PNP. In humans it is known that inhibitors of human PNP have potential clinical use as immunosuppressants. Since there is a risk that a TB-PNP inhibitor would also inhibit human PNP, leading to unacceptable side-effects during treatment, there is a need to develop TB-PNP inhibitors that do not bind to human PNP. Although TB-PNP is evolutionarily related to human PNP, a recent crystal structure shows a significant difference in the hydrogen bonding to ImmH between the mammalian and TB PNPs. It is therefore likely that molecular modeling can be useful in designing selective inhibitors for TB PNP. The need to develop selective inhibitors is a common problem in discovery and development.

The goal of these exercises is to teach you how to use BioMedCAChe to design your own selective inhibitors by using peer reviewed studies from the current scientific literature applied to the important problem of obtaining new drugs for the treatment of TB.

[1] Shi, W., Basso, L. A., Santos, D. S., Tyler, P. C., Furneaux, R. H., Blanchard, J. S., Almo, S. C. and Schramm, V. L., "Structures of Purine Nucleoside Phosphorylase from *Mycobacterium tuberculosis* in Complexes with Immucillin-H and Its Pieces," *Biochemistry*, **2001**, *40*, 8204-8215.

[2] Ojha, A. K., Muckherjee, T. K., and Chatterji, D., "High Intracellular Level of Guanosine Tetraphosphate in *Mycobacterium smegmatis* Changes the Morphology of the Bacterium", *Infect. Immun.,* **2000**, *68*, 4084-4091.

Importing proteins

The Protein Data Bank maintained by the Research Collaboratory for Structural Bioinformatics (RCSB)[1] contains tens of thousands x-ray crystal structures of proteins and other biomolecules.

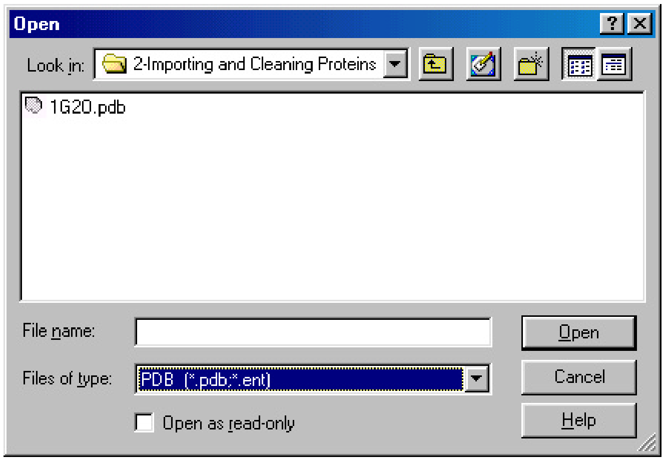
NOTE

The PDB file is `1g2o`. The last letter is ‘oh’, not zero.

The structures of proteins and other biomolecules can be down loaded from `http://www.rcsb.org/pdb/` as PDB files. PDB files have an extension of `.pdb` or `.ent`. To view the 3D structure of TB-PNP complexed with ImmH, you simply download and open `1G20.pdb` in the CAChe workspace.

To open the PDB file

1. Choose **File** | **Open** to display the Open dialog box.



Look in: 2-Importing and Cleaning Proteins

1G20.pdb

File name:

Files of type: PDB (*.pdb;*.ent)

□ Open as read-only

2. Select the arrow button in the **Look in** drop down box to display a drop-down list of:
   - folders in the directory structure above the currently open folder
   - available drives from which you can open the file.
3. Select the arrow button in the **Files of type** box and choose **PDB (*.pdb, *.ent)** from the drop-down list.
4. Choose the folder or drive where 1G20.pdb is located from the drop-down list.

   Do one of the following:

   - To move up a level of folders to locate the folder containing the file, select the dialog box button shown to the left.
   - To locate the file in a subfolder, click and drag the scroll bar in the scrolling list to display the folder where the file is located, and double-click on the folder to open it.

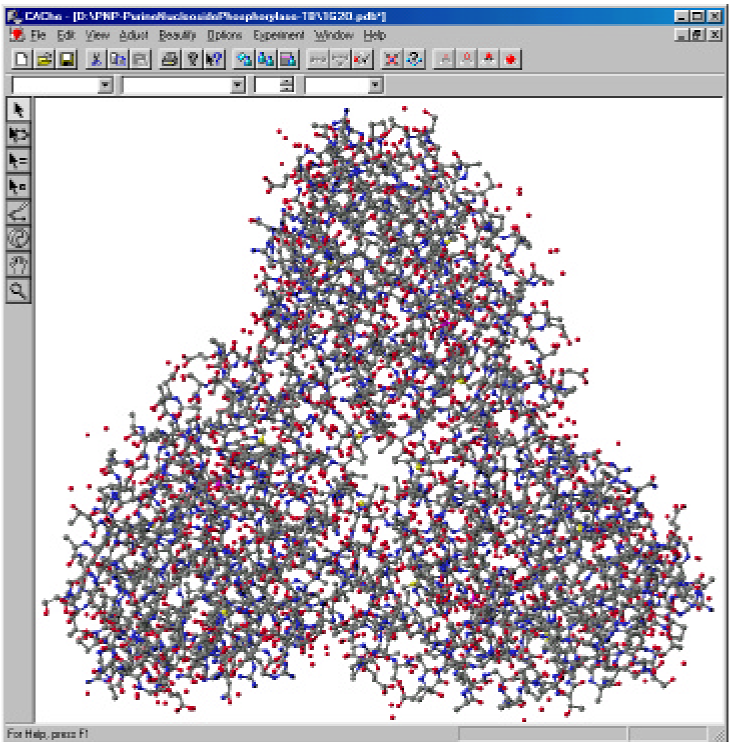
[1] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, “The Protein Data Bank”, *Nucl. Acits. Res.*, **2000**, *28*, 235-242.

5. Do one of the following:
   - Click and drag the scroll bar in the scrolling list to locate 1G20.pdb and select the file by clicking on it.
   - Click in the **File name** text box and type 1G20.pdb.
6. Select **Open** to open the file and to close the Open dialog box.

A new workspace opens and the obviously trimeric structure displays.

📖 NOTE

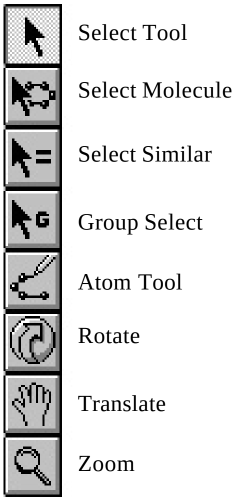
It may take a minute to load the PDB file, depending on the speed of your computer.

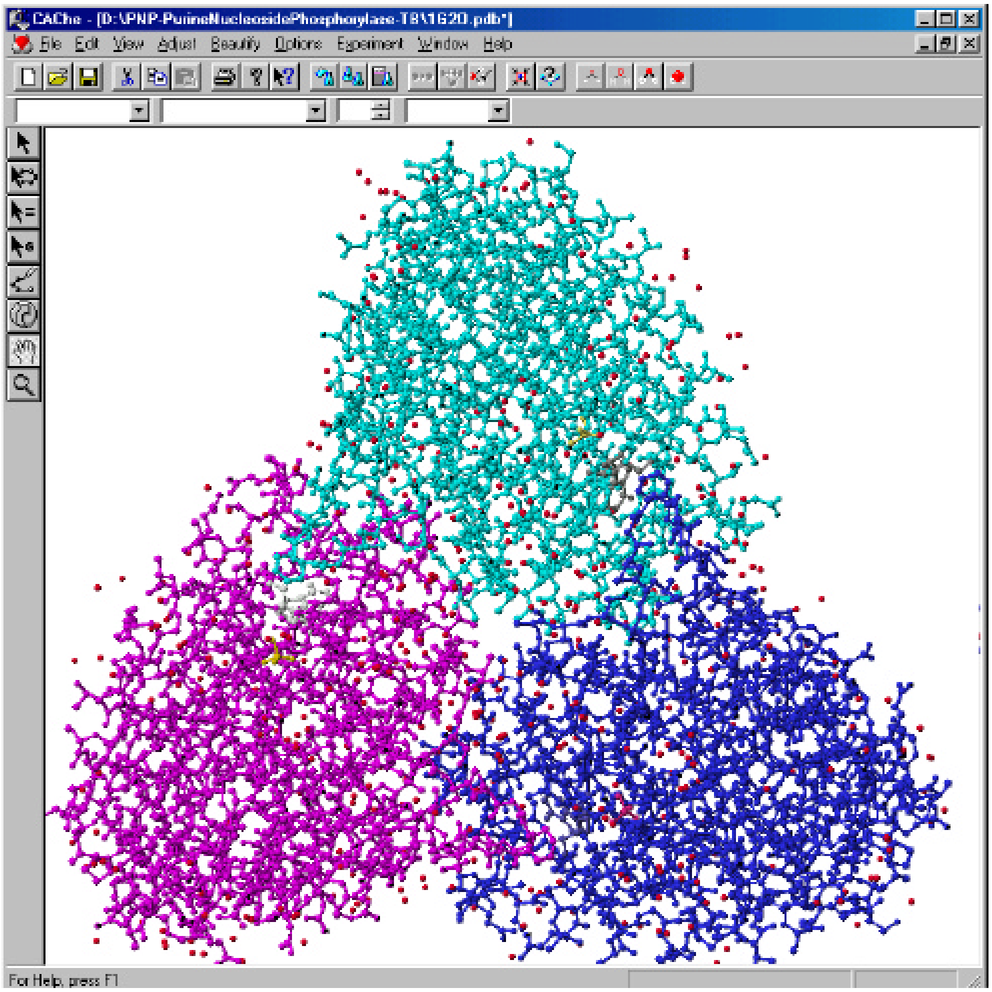


To see clearly the three chains and their ligands in the trimeric structure:

1. choose **View | Color by Molecule**.

   Each chain and ligand gets a different color. Water molecules all remain red. Choose the select by molecule tool and click on each colored component to examine it in more detail.




CAChe - [D:\PNP-PurineNucleosidePhosphorylase-T8\1G20.pdb*]
File Edit View Adjust Beautify Options Experiment Window Help
For Help, press F1

📖 TIP

Opening csf files is much faster than opening PDB files. 1G20.csf will open in a few seconds.

At this point you should save the file as a chemical sample with the name 1G2O.csf. From now on you will work with the CAChe chemical sample file (csf) rather than the PDB file.so that you can retain all of the views and information you create in modeling.

Saving a PDB molecule

To save a PDB molecule

1. Select **File | Save**.

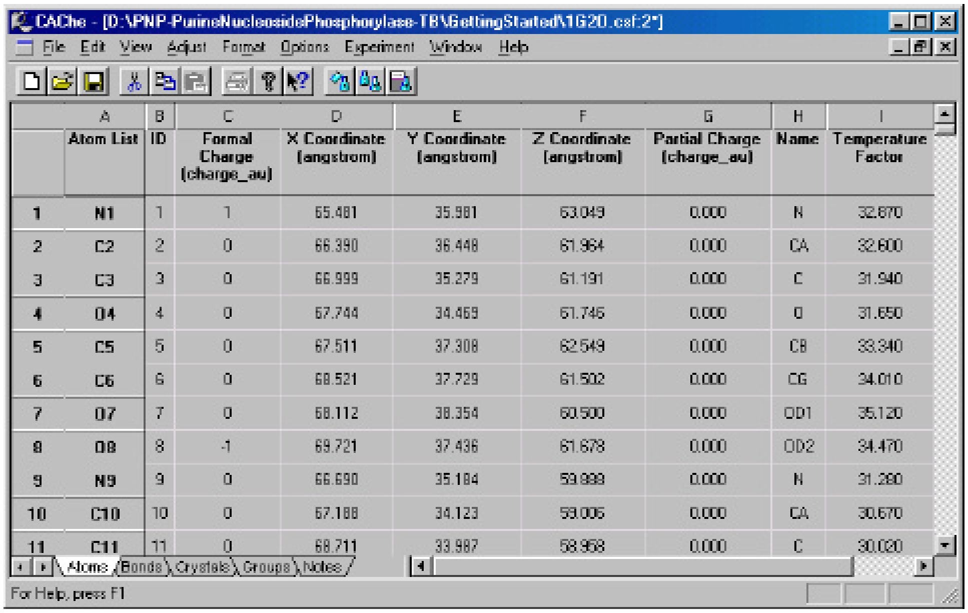
When you save this workspace, you can chose to save it as a chemical sample file (`*.csf`), a PDB file, or other file type. Save the workspace as a chemical sample file (`*.csf`) to preserve all of the information you have added. Otherwise, information such as rendering style and computed atom properties will be lost.

📖 NOTE

You will be informed that the charge on the protein is +3. Click **Save**. You will adjust the charge later.

All of the information in the PDB file has been stored in the CAChe chemical sample format. For example, residues have been stored as Groups and you can use the Select Group tool to select residues.

To view this information, choose **Analyze | Chemical Properties Spreadsheet.** The chemical sample properties window appears.


CAChe - [D:\PNP-PurineNucleosidePhosphorylase-TB\GettingStarted\1G20.csf:2*]
File Edit View Adjust Format Options Experiment Window Help
Atoms Bonds Crystals Groups Notes

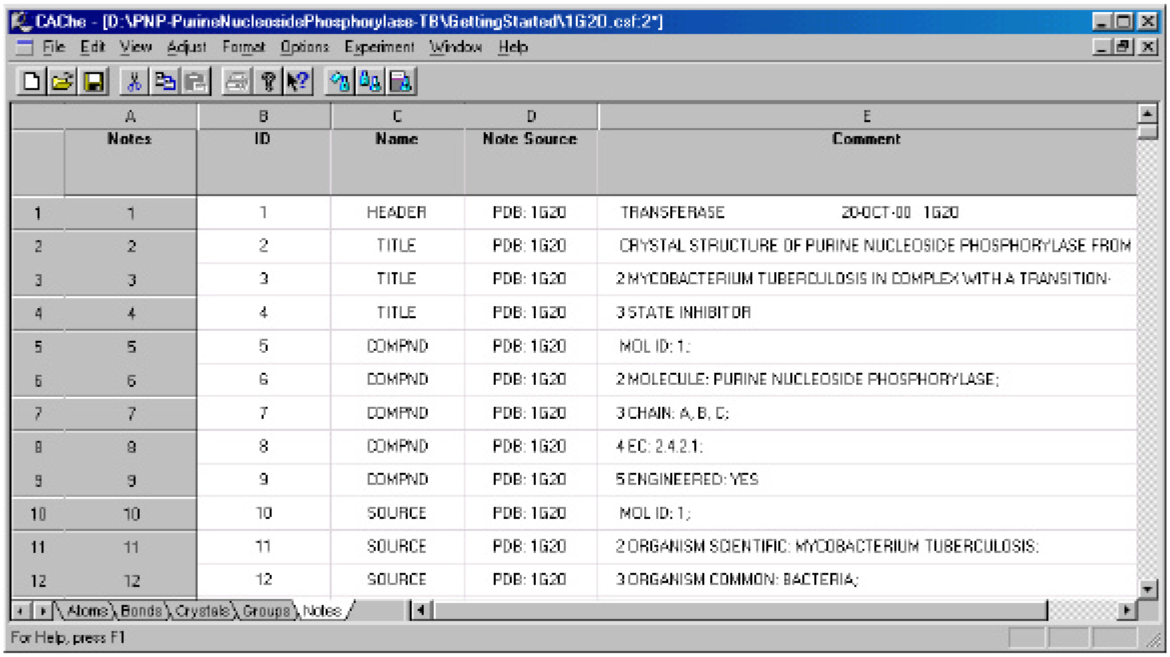
	A	B	C	D	E	F	G	H	I
	Atom List	ID	Formal Charge (charge_au)	X Coordinate (angstrom)	Y Coordinate (angstrom)	Z Coordinate (angstrom)	Partial Charge (charge_au)	Name	Temperature Factor
1	N1	1	1	65.481	35.981	63.049	0.000	N	32.670
2	C2	2	0	66.390	36.448	61.964	0.000	CA	32.600
3	C3	3	0	66.999	35.279	61.191	0.000	C	31.940
4	O4	4	0	67.744	34.469	61.746	0.000	O	31.650
5	C5	5	0	67.511	37.308	62.549	0.000	CB	33.340
6	C6	6	0	68.521	37.729	61.502	0.000	CG	34.010
7	O7	7	0	68.112	38.354	60.500	0.000	OD1	35.120
8	O8	8	-1	69.721	37.436	61.678	0.000	OD2	34.470
9	N9	9	0	66.690	35.184	59.889	0.000	N	31.280
10	C10	10	0	67.188	34.123	59.006	0.000	CA	30.670
11	C11	11	0	68.711	33.987	58.958	0.000	C	30.020

For Help, press F1

Note that the sample properties for a PDB molecule contain several worksheets:

- **Atoms** contains atoms and their properties
- **Bonds** contains bonds and their properties
- **Groups** contains residues and their properties
- **Electrons** worksheet keeps track of nonbonded electron pairs so that the valency can be checked. Electrons are added when you beautify a structure and do not appear when the PDB file is first opened.
- **Notes** contains PDB REMARK, NOTE, SOURCE, COMPND and HEADER records information.

Select the **Note** tab and expand the Comment column so that you see this:


CAChe - [D:\PNP-PurineNucleosidePhosphorylase-TB\GettingStarted\1G20.csf:2*]
File Edit View Adjust Format Options Experiment Window Help
Atoms Bonds Crystals Groups Notes
For Help, press F1

	A	B	C	D	E
	Notes	ID	Name	Note Source	Comment
1	1	1	HEADER	PDB: 1G2O	TRANSFERASE 20-OCT-00 1G2O
2	2	2	TITLE	PDB: 1G2O	CRYSTAL STRUCTURE OF PURINE NUCLEOSIDE PHOSPHORYLASE FROM
3	3	3	TITLE	PDB: 1G2O	2 MYCOBACTERIUM TUBERCULOSIS IN COMPLEX WITH A TRANSITION-
4	4	4	TITLE	PDB: 1G2O	3 STATE INHIBITOR
5	5	5	COMPND	PDB: 1G2O	MOL ID: 1;
6	6	6	COMPND	PDB: 1G2O	2 MOLECULE: PURINE NUCLEOSIDE PHOSPHORYLASE;
7	7	7	COMPND	PDB: 1G2O	3 CHAIN: A, B, C;
8	8	8	COMPND	PDB: 1G2O	4 EC: 2.4.2.1;
9	9	9	COMPND	PDB: 1G2O	5 ENGINEERED: YES
10	10	10	SOURCE	PDB: 1G2O	MOL ID: 1;
11	11	11	SOURCE	PDB: 1G2O	2 ORGANISM SCIENTIFIC: MYCOBACTERIUM TUBERCULOSIS;
12	12	12	SOURCE	PDB: 1G2O	3 ORGANISM COMMON: BACTERIA;

📖 TIP

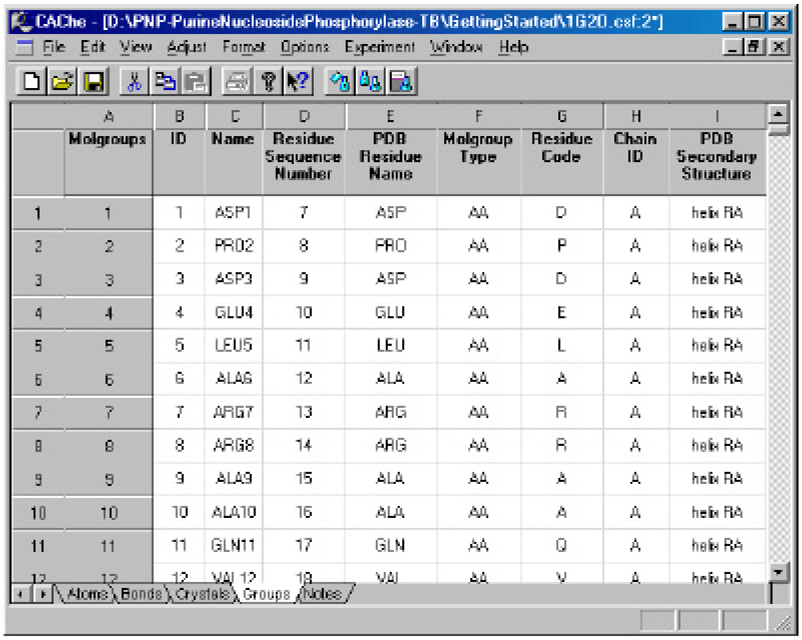
Click the **Comment** column header to highlight the column, then select **Format | Left** to left justify the comments.

Read through the comments to confirm that this is the structure for ImmH complexed with TB-PNP at 1.75 Angstrom resolution. Notice that the first six residues (Met-Ala-Asp-Pro-Arg-Pro) in each of the three chains TB-PNP were not resolved and are missing from the structure (row 249). Find the HET records and the HETNAM records. These records identify groups that are not standard amino acids such as ligands. There is one IMH and one PO4 for each of the three chains in TB-PNP. The names and molecular formulas for each HET group is given in the HETNAM and FORMUL records. You will use this information later when cleaning up the HET groups.

Simplifying the structure

Each of the chains in TB-PNP is the same and each has ImmH and PO4 complexed in its active site. In the next exercise you will analyze the binding of ImmH to the active site in TB-PNP and it is helpful to work with the simplest model first, a single chain and its complexed groups. We will reduce the structure to the monomer by removing chains B and C.

Choose the **Groups** tab in the workbook.The Groups worksheet appears


CAChe - [D:\PNP-PurineNucleosidePhosphorylase-TB\GettingStarted\1G20.csf:2*]
File Edit View Adjust Format Options Experiment Window Help
Atoms Bonds Crystals Groups Notes

	A	B	C	D	E	F	G	H	I
	Molgroups	ID	Name	Residue Sequence Number	PDB Residue Name	Molgroup Type	Residue Code	Chain ID	PDB Secondary Structure
1	1	1	ASP1	7	ASP	AA	D	A	helix RA
2	2	2	PRO2	8	PRO	AA	P	A	helix RA
3	3	3	ASP3	9	ASP	AA	D	A	helix RA
4	4	4	GLU4	10	GLU	AA	E	A	helix RA
5	5	5	LEU5	11	LEU	AA	L	A	helix RA
6	6	6	ALA6	12	ALA	AA	A	A	helix RA
7	7	7	ARG7	13	ARG	AA	R	A	helix RA
8	8	8	ARG8	14	ARG	AA	R	A	helix RA
9	9	9	ALA9	15	ALA	AA	A	A	helix RA
10	10	10	ALA10	16	ALA	AA	A	A	helix RA
11	11	11	GLN11	17	GLN	AA	Q	A	helix RA
12	12	12	VAL12	18	VAL	AA	V	A	helix RA

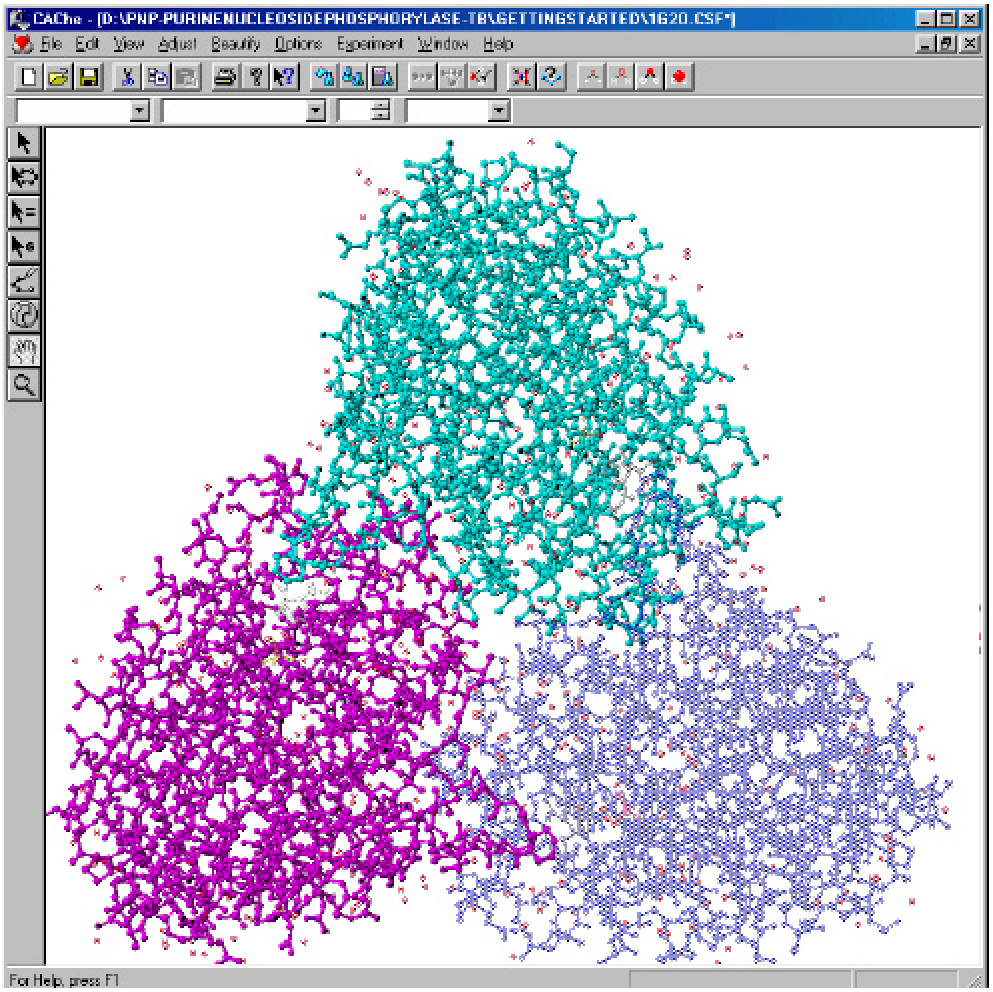
Do the following:

1. Scroll the window until you see the start of rows containing **Chain ID** B.
2. Select the first row with **Chain ID** B by clicking the row number on the left. (Row 263)

   Row 263 highlights.
3. Scroll until you see the last row containing **Chain ID** C.
4. Shift-click the row number on the left.

   All residues in chains B and C are highlighted indicating that they are selected.
5. Close the Sample Properties Window by clicking the close box (x) in the window’s upper right-hand corner.

The 3D Structure Window comes to the front and you see:


CAChe - [D:\PNP-PURINENUCLEOSIDEPHOSPHORYLASE-TB\GETTINGSTARTED\1G20.CSF*]
File Edit View Adjust Beautify Options Experiment Window Help
For Help, press F1

Select Tool

Select Molecule

Select Similar

Select Group

Atom Tool

Rotate

Translate

Zoom

6. Select **Edit | Delete**.

   Chains B and C disappear exposing the HET groups and water molecules associated with chains B and C.

7. Using the Select Tool click and drag to select a rectangular area containing water molecules and HET groups from chains B and C.

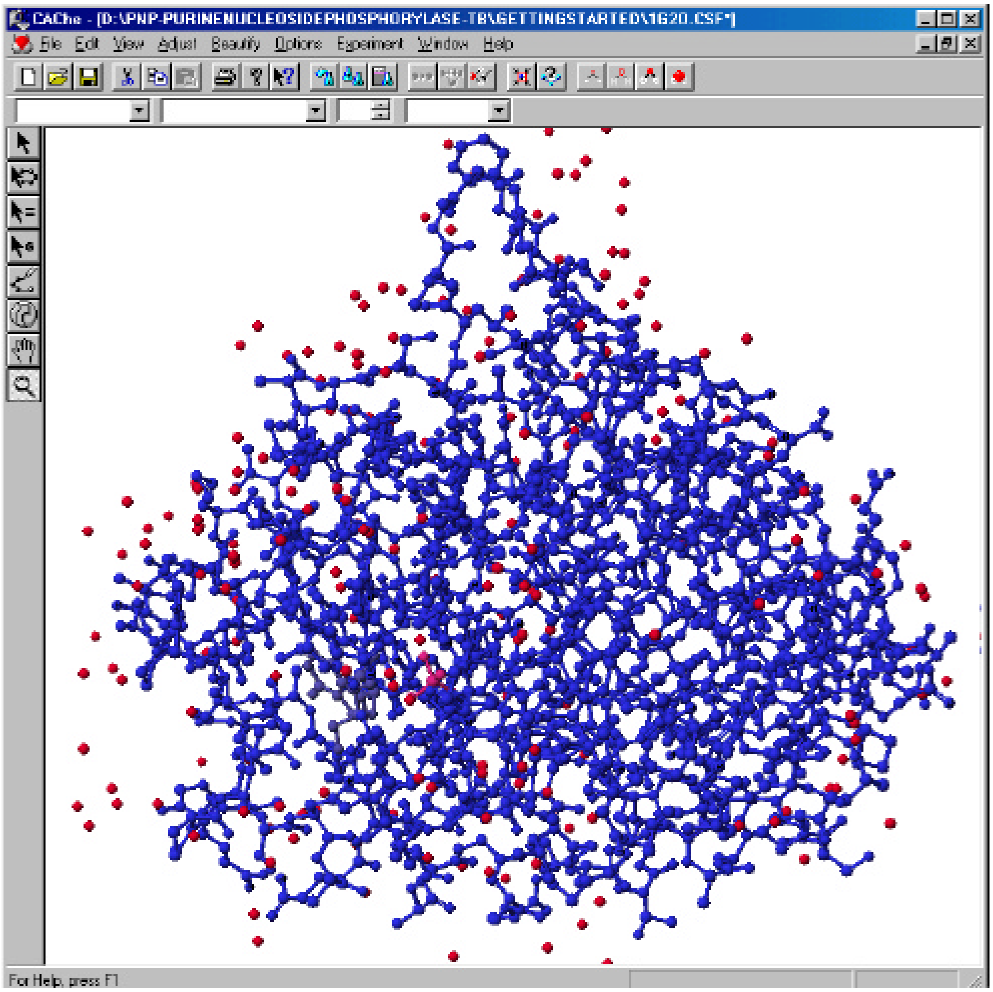
   The selected water molecules and HET groups are highlighted and chain A is greyed.

8. Select **Edit | Delete**.

   The excess water and HET groups disappear and the remaining structure is highlighted.

9. Repeat the drag selection and deletion until all water molecules and HET groups belonging to chains B and C are deleted.

   Don’t worry too much about getting all of the water molecules. Leave any water molecules that you think might be associated with chain A. Your monomer structure will look similar to this


CAChe - [D:\PNP-PURINENUCLEOSIDEPHOSPHORYLASE-TB\GETTINGSTARTED\1G20.CSF*]
File Edit View Adjust Beautify Options Experiment Window Help
For Help, press F1

10. Choose **File | Save As** and save this file as `TB-PNP-monomer.csf`.

Next you will prepare the structure for molecular modeling and analysis.

Cleaning protein structure

When a PDB file is opened in CAChe, some new information is automatically calculated and added. However, the information required for molecular modeling must still be added such as hydrogen atoms, atom hybridization and correct bond types for HET groups and non standard residues. If residues are missing or incomplete, it may be necessary to correct their structures.

Lock atoms at their crystallographic positions

1. Choose **Edit | Select All**.

   All atoms and bonds are highlighted.

2. Choose **Adjust | Lock**.

   Selected atoms are locked at their current position in space. This prevents us from accidentally moving an atom from its crystallographic position.

Check and correct the bonding in HET groups

1. Choose **View | Color by Molecule**.

   The protein chain, immucillin, and phosphate groups are displayed in different colors.

2. Choose the Select Molecule Tool and click on the ImmH ligand.

   ImmH is highlighted, the rest of the structure is dimmed.

3. Choose **View | Hide Unselected**.

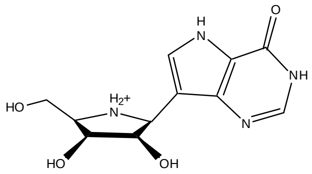
   All unselected atoms and bonds disappear.

4. Choose **View | Fit in Window**.

   ImmH zooms to fill the window.

5. Choose **View | Color by Element**.

   The ligand changes so that carbon atoms are grey, oxygen atoms are red, nitrogen atoms are blue and hydrogen atoms are white.

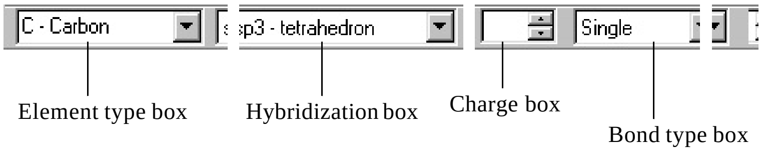


immucillin-H

Examine ImmH and change the bonding to agree with immucillin-H structure shown in the figure.

6. To change a bond order, use the Select Tool and click a bond to select it.
7. From the style bar, pull down the Bond type menu and select the new bond type.

Style bar



All selected bonds change to the new bond type.

Next examine ImmH and set the charges to agree with the figure

8. Click the nitrogen atom in the iminoribitol ring to select it
9. Change the charge in the Charge Box to 1 using the up arrow next to the charge box.

   A **+1** appears on all selected atoms.

TIP

If you type the charge in the charge box instead of using the scrolling arrows, press the Enter key to apply the new charge.

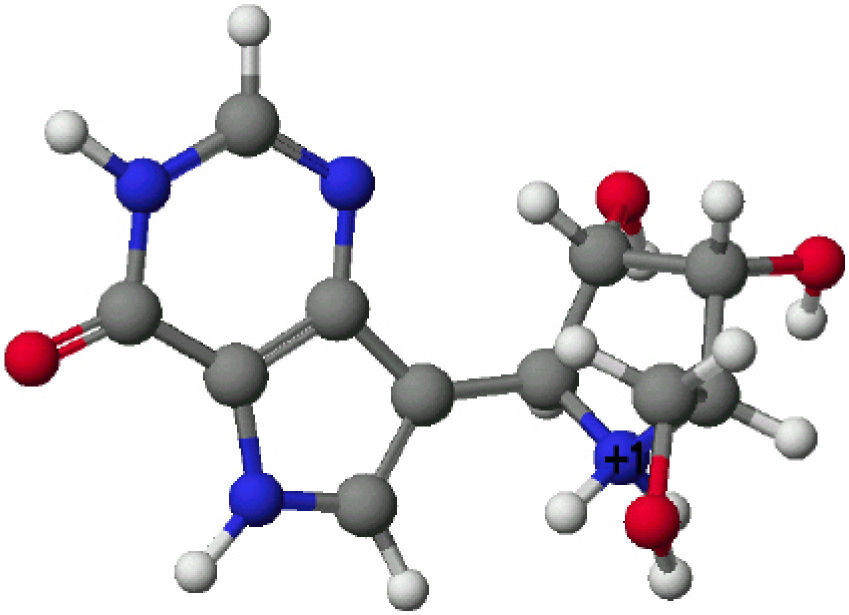
Add hydrogen atoms and define atom hybridization for ImmH

1. Choose the Select Molecule Tool and click on the ImmH ligand.

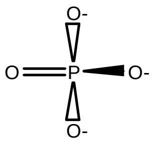
   ImmH is highlighted.
2. Choose **Beautify | Valence**.

   Hydrogen atoms, electrons and the atom hybridization are added.

ImmH should look like this now



Add hydrogen atoms and define atom hybridization for PO_4

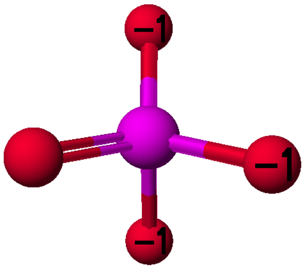


1. Choose **View | Show Hidden**.

   The full structure is displayed.

Repeat the steps to clean the PO_4^{3-} ligand.

For this exercise, don't be concerned about which oxygen atoms have the negative charge. The cleaned PO_4^{3-} should look like this:



Add hydrogen atoms and atom hybridization to the protein

1. Choose **View | Show Hidden**.

   All the atoms and bonds in the chemical sample appear.

2. Choose **Edit | Select All**.

   All the atoms and bonds are highlighted.

3. Choose **Beautify | Valence**.

   Hydrogen atoms, electrons and hybridization are added to the protein and the water molecules.

4. Choose **Beautify | H-Bonds**.

   Hydrogen atoms in -OH and water molecules are rotated to maximize hydrogen bonding.

Only the geometry of hydrogen atoms are changed with these commands.

📖 NOTE

Depending on the speed of your computer, it may take a minute to beautify the full protein and hydrogen bonds.

Save your work as `TB-PNP-cleaned.csf`.

1. Choose **File | Save As** and save the chemical sample as `TB-PNP-cleaned.csf`.

2. Choose **File | Close** to end your work with this chemical sample.

At this point you would normally refine the position of the hydrogen atoms using molecular mechanics.