

Building a low-resolution 3D model from high-resolution data

In summary, a coordinate file (e.g. *.pdb file) is prepared, H atoms are added, structure factors (SF) are computed, SF format is changed, and the 3D model is computed. This outline is only for icosahedral particles in PDB format but other symmetries or formats could, of course, be used after some changes.

Unless stated otherwise, see /work3/people/belnap/Xplor/ for example files. This file is calc_dens.doc in that directory.

1) Get a PDB file from Brookhaven data base or elsewhere.

- a) Determine if a PDB file exists in the Brookhaven National Laboratory's Protein Data Base (PDB). One method to quickly search the PDB files utilizes Brookhaven's website (<http://www.pdb.bnl.gov/>).

NOTE: be specific in the search. The keyword "polio" yielded no result while the more descriptive "poliovirus" located many entries.

Once the proper file is located, you can retrieve it from the data base with your WWW browser program. If you want to use FTP to retrieve it then make note of the Data Base's identification number for the specific file. It should be located at the start of the description of the file (e.g. 1pvc for POLIOVIRUS TYPE 3, SABIN STRAIN).

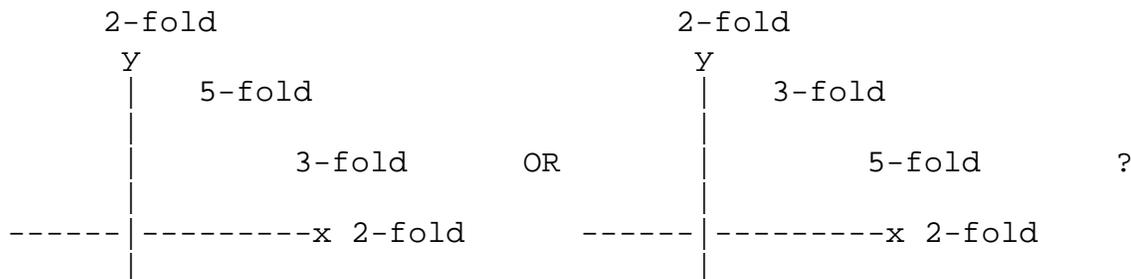
b) Use the following procedure to FTP the file from Brookhaven.

```
>FTP [cr]
>open pdb.pdb.bnl.gov [cr]
user>anonymous [cr]
password>your name [cr]
-Enter the directory with the desired PDB file.
  >cd /fullrelease/uncompressed_files/** [cr]
-Replace ** with the two middle letters found in the id
number obtained in step one. Examine the contents of this
directory with the list command.
  >ls [cr]
-Locate the file and download it to your disk.
  >ascii [cr]      (tells the computer the file is text)
  >get pdb1pvc.ent [cr]      (for example)
```

2) Prepare PDB coordinates. Atomic coordinates are given in the *.pdb file. The structure must be oriented in the correct position and the file must be edited. Steps should be followed in the order given. (Example file: polyoma_OT.pdb, which is the X-ray structure of polyoma virus VP1 capsid protein [Stehle & Harrison (1996) Structure 4:183-194].) You can use Abhinav Kumar's program "DEAL_PDB" (also called DOPDB), or some other program that will edit PDB-format files, for many of the operations.

a) Are coordinates in standard (i.e. 2-fold axes on x,y,z) or crystal orientation? If crystal, ask to have them in standard orientation or rotate them yourself. The center of the particle must be at the origin (0,0,0).

b) Is orientation in x-y plane,



The left case is the correct convention. If orientation is the right case then rotate coordinates by 90 deg about any axis (x, y, or z). Use ROTATE option in DEAL_PDB program. The particle center must be at the origin (prior to the 90 deg. rotation).

c) Edit PDB file to remove extraneous material and to renumber the atoms and residues. The following gives DEAL_PDB commands:

-Eliminate all lines that do not contain information regarding atomic coordinates. This can be achieved in DEAL_PDB by reading in the PDB file with the "read" option. (See last substep for keeping the extra information out of output file.)

```
>deal_pdb [cr]
>read filename.pdb [cr] (for example)
```

-Remove water and other non-amino acid molecules (if any are present) by deleting the associated residue numbers in DEAL_PDB

```
>del_res [cr]
```

or simply delete by use of a text editor. Beware of any atoms with "strange" names such as "OXT" for the terminal oxygen.

-Renumber the atoms and residues with the following commands:

```
>renum_atm [cr]
```

```
>renum_res [cr]
```

-Remove the chain id number (if any) with a command found in the "set_value" sub-option page.

```
>set_value [cr]
```

```
>chain_id [cr]      (this sets all chain id's to " ")
```

-Remove the text found at the end of each line.

```
>set_value [cr]
```

```
>text/segid [cr] -this sets text description to " "
```

-Save the changes made to the file with the "Write" option. Use "noheader" option to keep only coordinates in output file.

```
>write new_filename.pdb noheader [cr]  (for  
example)
```

d) Now shift particle center from 0,0,0 to 1/2,1/2,1/2 (i.e. the middle of a unit cell that begins at the origin). Use TRANSLATE option in DEAL_PDB.

e) Edit PDB file, add OT and END to the end of the file. OT signifies the terminal oxygen and should be the final atom in the last residue. Simply copy the last O atom's coordinates, change the atom number (n+1), and alter the coordinates slightly. "END" should be the last line in the file. For example, in polyoma the last residue is PHE:

```
ATOM 20852 C   PHE  5373      252.950 402.164 482.294  1.00136.36  
ATOM 20853 O   PHE  5373      252.000 401.978 483.069  1.00136.95  
ATOM 20854 OT  PHE  5373      251.990 401.968 483.079  1.00136.95  
END
```

NOTE: Some pdb files list the atoms in different sequences. It is important that the OT follow a coordinate for an oxygen atom. This means one may have to cut the last oxygen atom and paste it before the created "OT" line.

3) Add hydrogen atoms to high-resolution structure. (Example file is gen_H.inp.) (In XPLOR jargon, gen_H.inp actually builds "molecular structure".)

a) Edit gen_H.inp. Change "polyoma_OT.pdb" to your PDB filename. You can also change the "polyoma" on line 40 or the output filenames to something else.

- b) Files `tophcsdx.pro`, `parhcsdx.pro`, and `toph19.pep` have standard Xplor parameters essential for adding the H atoms. Do not alter them.
 - c) Execute the command `"xplor < gen_H.inp > gen.out"` from the UNIX prompt or as part of the `xpl.sh` batch file. Output files are `gen.out`, `genfit.pdb`, and `genfit.psf`. If errors are bad enough, `genfit.pdb` and `genfit.psf` are not written out. Even if they are, please check the `gen.out` file for errors (search for "ERR").
- 4) Compute structure factors. (Example file is `fcalc.inp`.)
- a) Edit `ncs_center`, a file containing the 60 icosahedral matrices. On line 6, change translation values to $1/2, 1/2, 1/2$ of your unit cell. (Same value as in step 2d above.) The "skew" on line 2 means the translation will apply to all other matrices.
 - b) Edit `fcalc.inp`, input parameters for SF calculation.
 - Unit cell values (a,b,c) in Angstroms.
 - "symmetry = (x, y, z)" refers to crystallographic symmetry only
 - "nreflections" should overestimate the number of reflections
 - "resolution" is "INFINITY" (low) to <high> (e.g. 25.0 Angs.)
 - "!grid = X" means program will use default grid size which is 1/3 of resolution
 - set temperature factor, b. The value 2000.0 works well.
 - scale `fcalc` output at line "do amplitude (`fcalc=0.001*fcalc`)" (in this example the scale factor is 0.001), you can wait until step 6 or 7 unless an amplitude is too big for the output file format.
 - c) Execute command `"xplor < fcalc.inp > fcalc.out"` from prompt or batch file. Input files are `genfit.psf`, `parhcsdx.pro`, `ncs_center`, and `genfit.pdb`. Output is `fc.xpl` and `fcalc.out`.
- 5) Change SF format to one compatible with EMSF3DBT.
- a) Run program EMFORMAT. It adds a header and other parameters essential to the EMSF3DBT program. Input `fc.xpl`, output `*.sf`.
 - b) You will be asked to enter "a, b, c (in pixels), alpha, beta, gamma (in deg.)" in EMFORMAT. Alpha, beta, and gamma

equal 90. The unit cell parameters, a, b, and c, are the dimensions of the unit cell in pixels. This should be an even number and a=b=c. The unit cell size you entered in step 5b divided by the number you enter here equals the pixel size in Angstroms.

- c) The program may need to be compiled. If so, execute "f77 emformat.f", which gives an executable file named "a.out".
 - d) In EMFORMAT, you may want to downweight the SF amplitudes. You can wait until step five unless the amplitudes are too large for the ASCII-formatted *.SF file.
- 6) Compute 3D model
- a) Port *.sf file to VMS computer.
 - b) Edit a *.HED parameter file. Typically, the entire unit cell will be computed but a smaller portion can be. See example files UNIT_CELL.HED (full unit cell) and PART_OF_CELL.HED (part of cell) in TSB\$BATCH. For changes to the *.HED file, see documentation for EMSF3DBT program TSB\$DOC:EMSF3DBT.DOC).
 - c) Run EMSF3DBT program. Input *.HED and *.SF, output is 3D model *.SFM.
 - d) Trim off first column, row, and section in EMMAP program. This is done by using option O and reading in a "PORTION" of the file: 2,max_num for cols., rows, secs. Use option W (write) to save file.
 - e) If necessary, scale the densities to be in the range -100 to +100 or less (but greater than -10 to +10). Very large intensities are usually seen without scaling and these high values can be a problem for programs like CMP (used to compare two 3D maps).
 - f) In EMMAP, rotate the map by 90 deg. and multiply the densities by -1.0 if you plan to use the map as input for EMPFT or to compare to image reconstructions.