





Joint Workshop of CERMM and PROTEO

CHARMM: Chemistry at HARvard Molecular Mechanics

Friday, March 12th and Saturday, March 13th, 2010

Richard J. Renaud Science Complex, Loyola Campus, Concordia University

7141 Sherbrooke Street West, Montreal, Quebec, Canada

Day 2 (Saturday, March 13th)

9:30 -13h00: TUTORIAL: System construction in CHARMM

Patrick Lagüe, Laval University

Tutorial Outline

- **First Step**: Getting to know the protein
- **Second step**: Getting ready for system building
- **Third Step**: System building with Charmm-gui:
  - STEP1: read biomolecular coordinate
  - STEP2: Solvate the biomolecule
  - STEP3: Setting periodic boundary conditions
  - STEPS4 and 5: Equilibration and Production
- **Fourth step**: Manual tweaking to include heme

Getting to know the protein

For the tutorial: PDB ID 3K9Z (*Nature* 462, p. 1079 (2009))

This is a **Mutated Myoglobin**, engineered to catalyse a Nitric Oxyde Reduction:

two-electron reduction of NO to N_2O

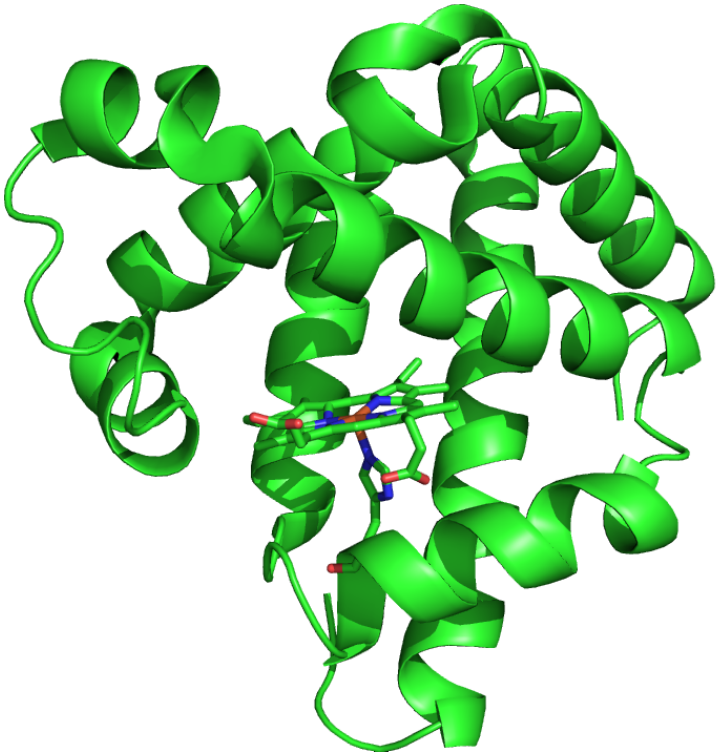
Protein contains:

153 aa (one subunit),

heme (42 heavy atoms,

Linked via HIS93),

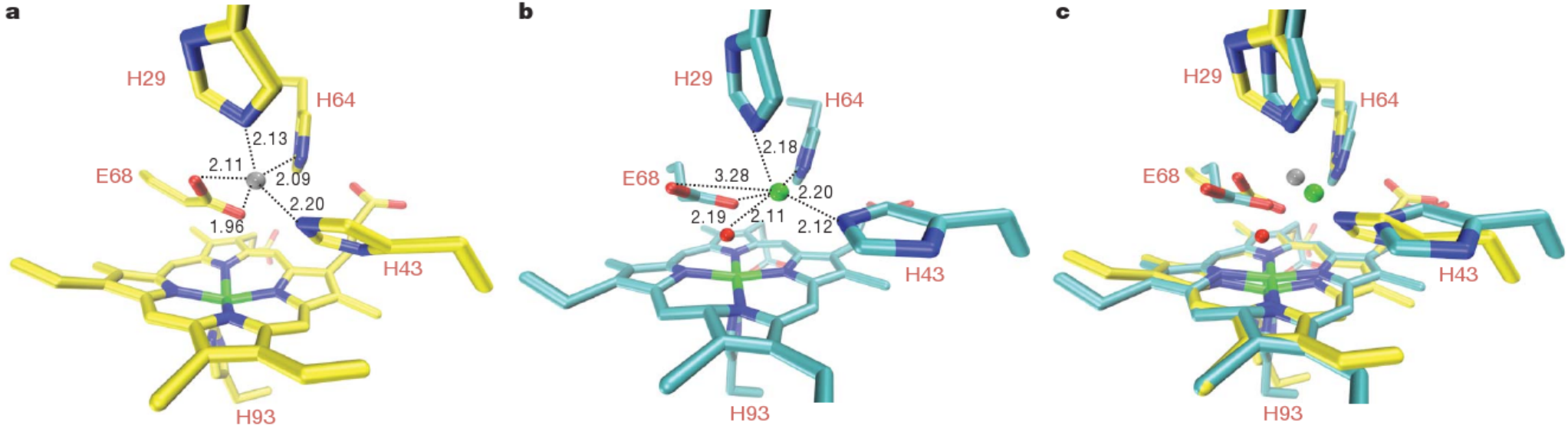
2 FE



Getting to know the protein

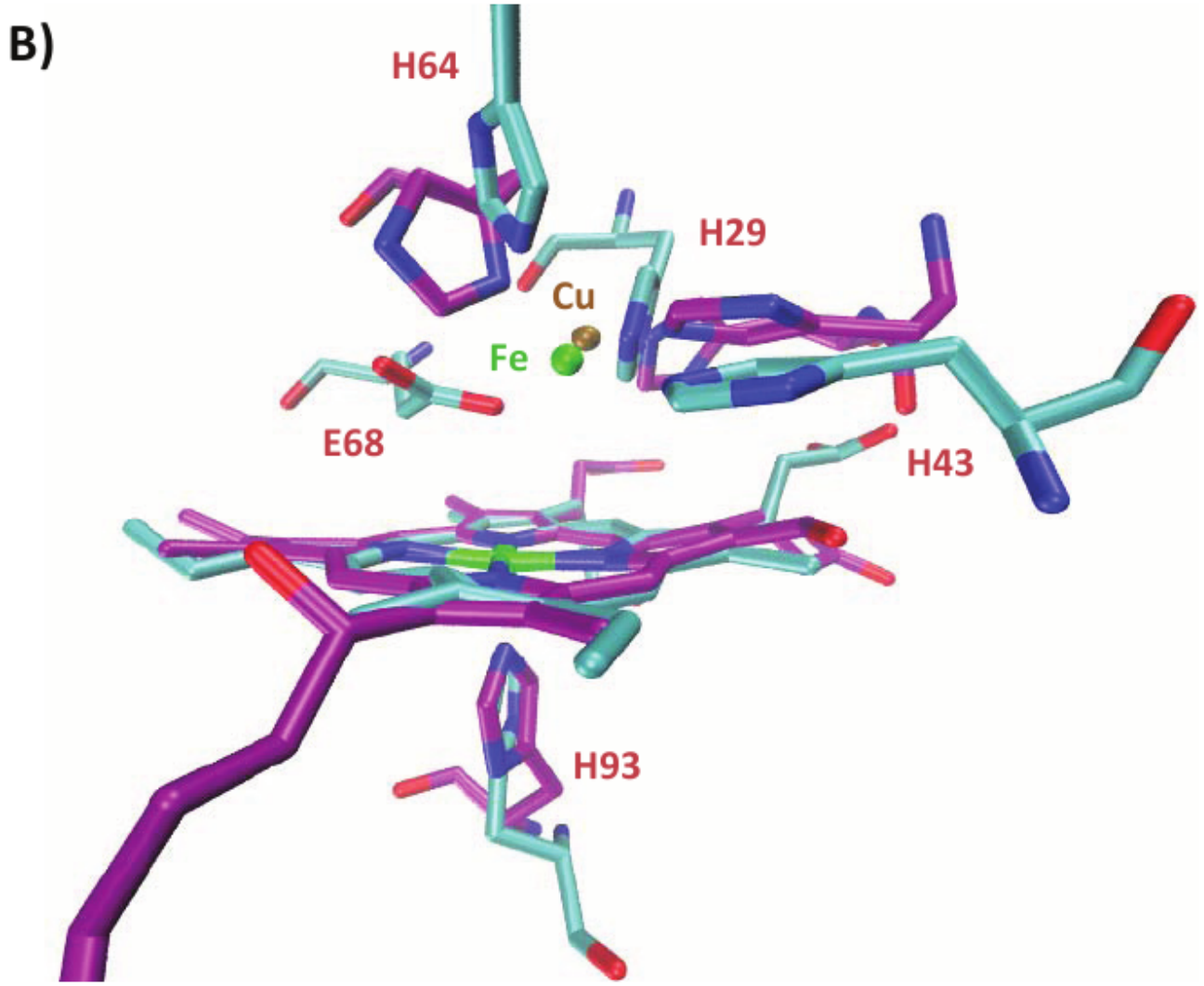
Mutated residues: L29H, F43H, V68E

H64 conserved, involved in reaction



Model Crystal Superposition

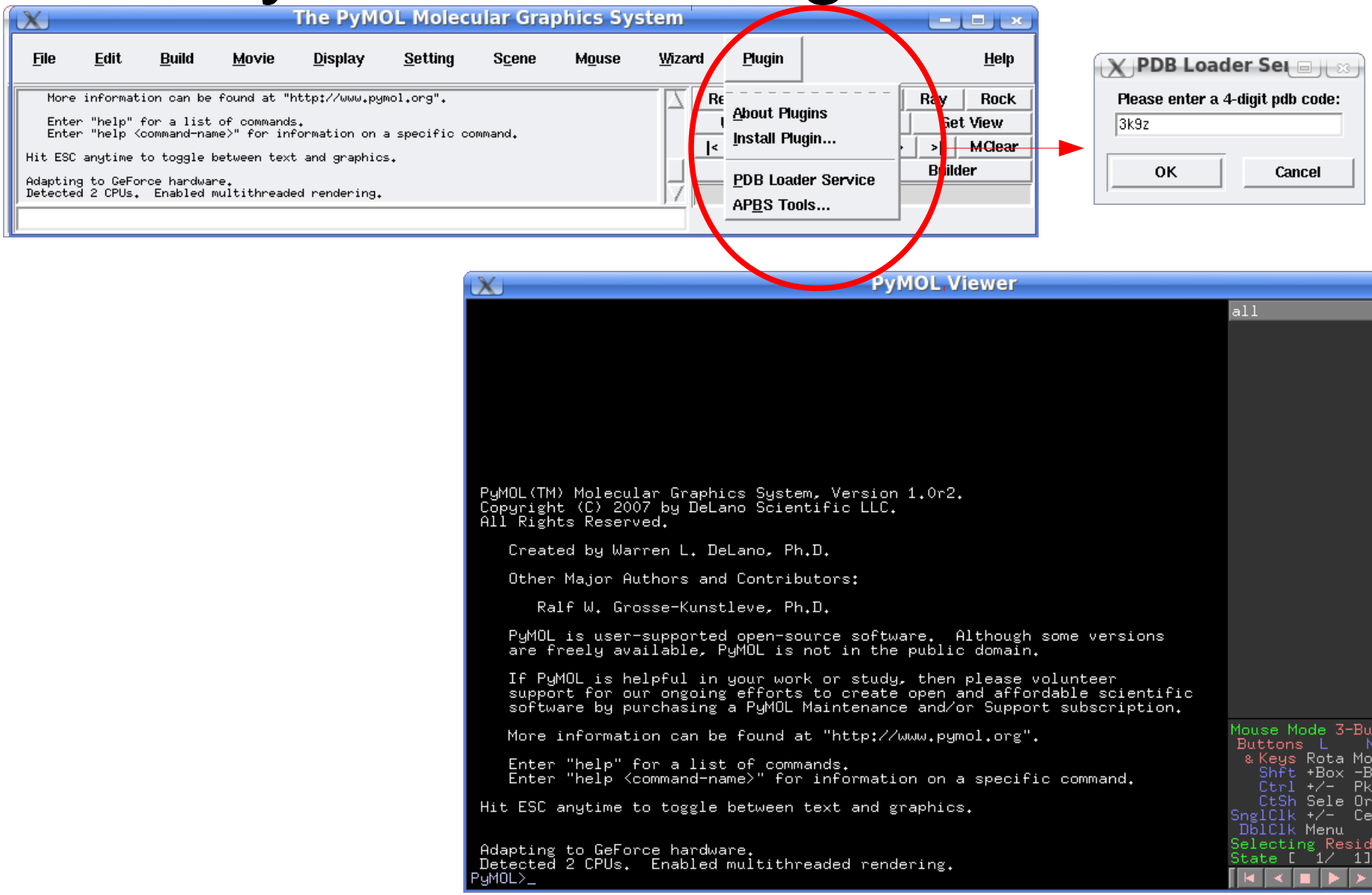
Similarity with Bovine Cyt C Oxidase



PYMOL - www.pymol.org

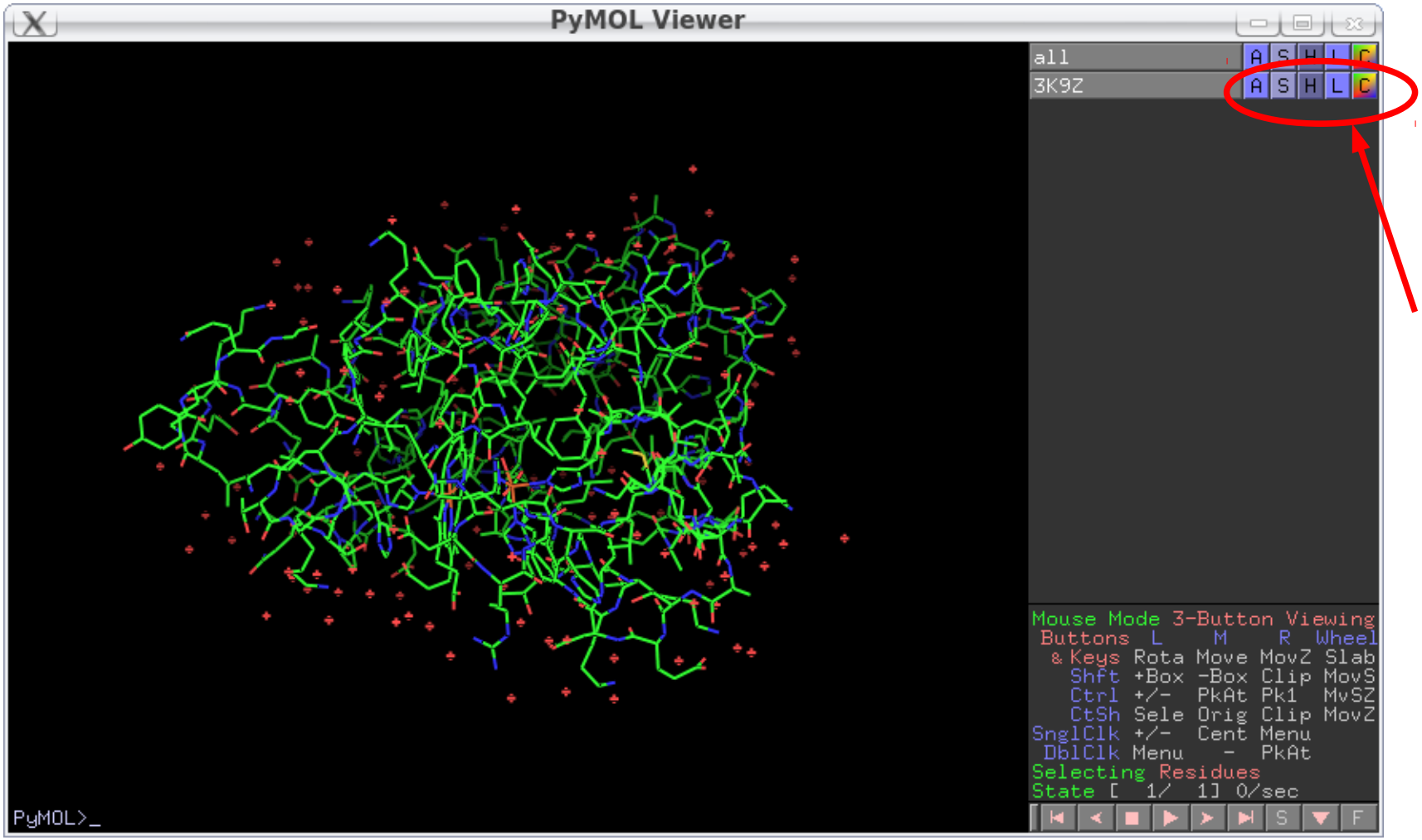
- Open source software – Molecular viewer
- Work on Linux, Unix, Mac and Windows
- Good community-supported documentation
- Publication-quality figures and Movies
- Scripting (powerful) - multithreaded
- Free academic licence
- On opening: 2 windows :
  - Viewer and Command

Pymol – Loading a molecule



To load a molecule, 2 options: File-Open OR Plugin – PDB Loader Service

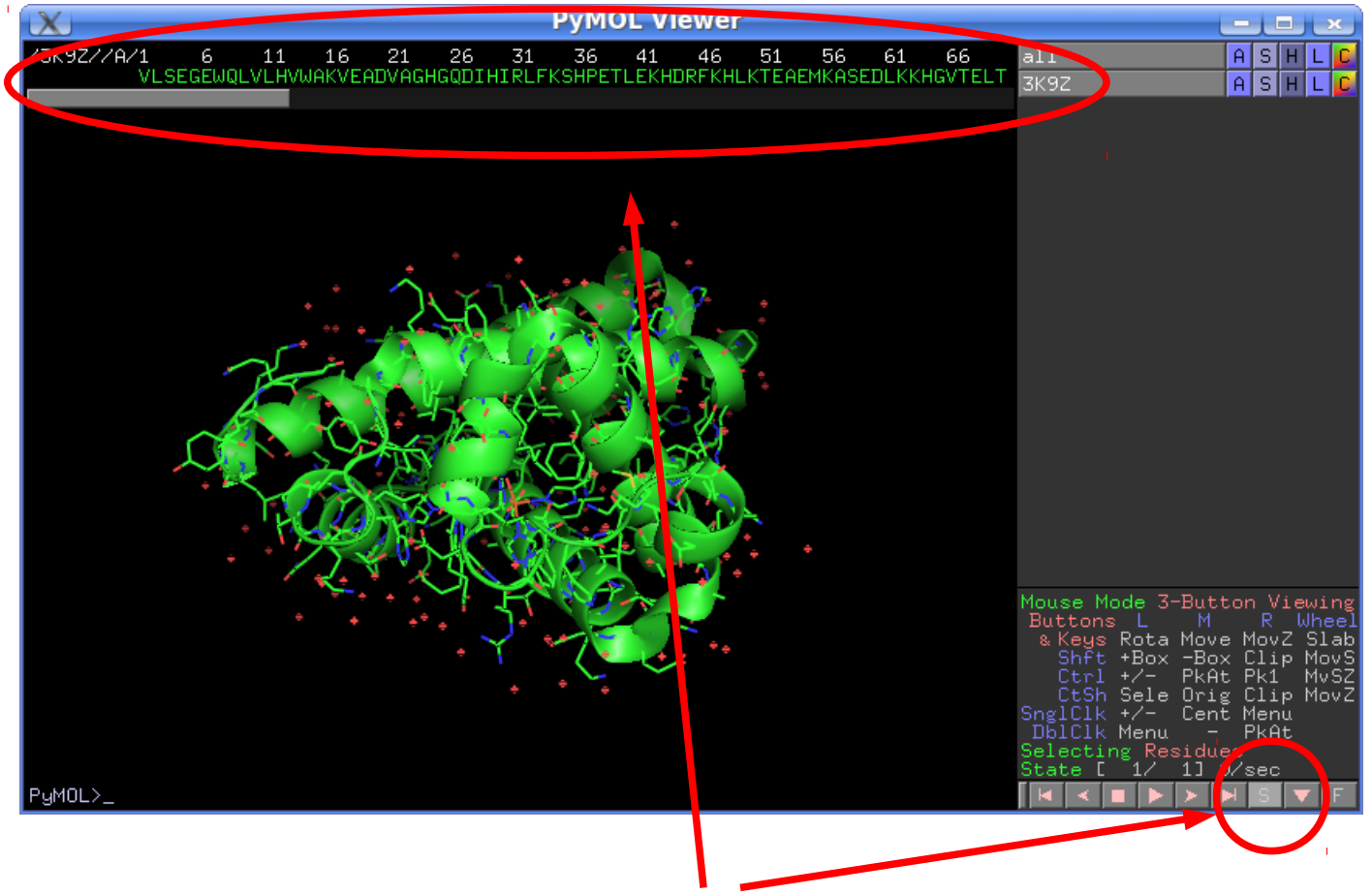
Pymol - Objects



Action
Show
Hide
Labels
Colors

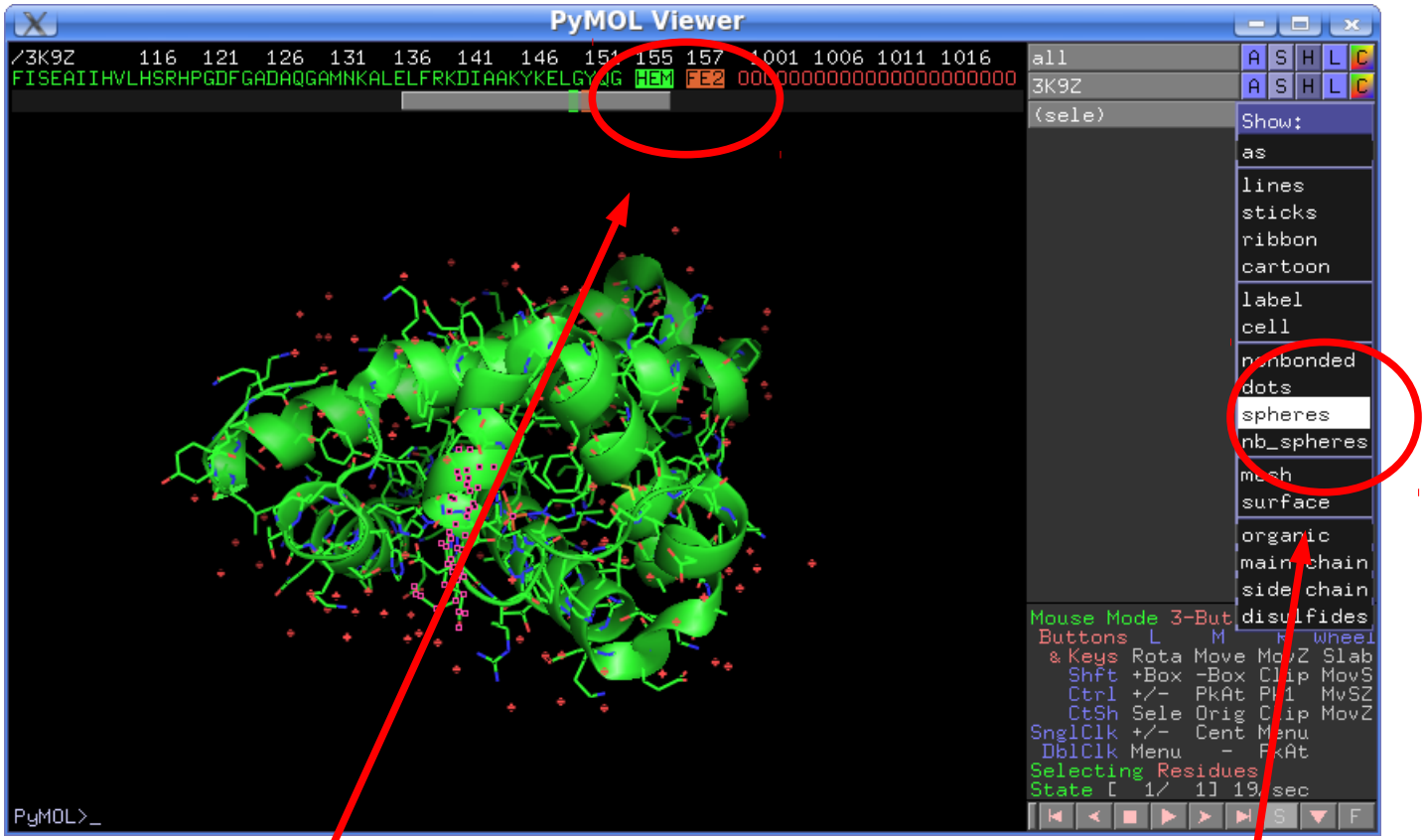
You can rotate, translate and Zoom with the mouse (right, middle and left buttons)
Try: Show - cartoon on 3K9Z object

Pymol - Sequence



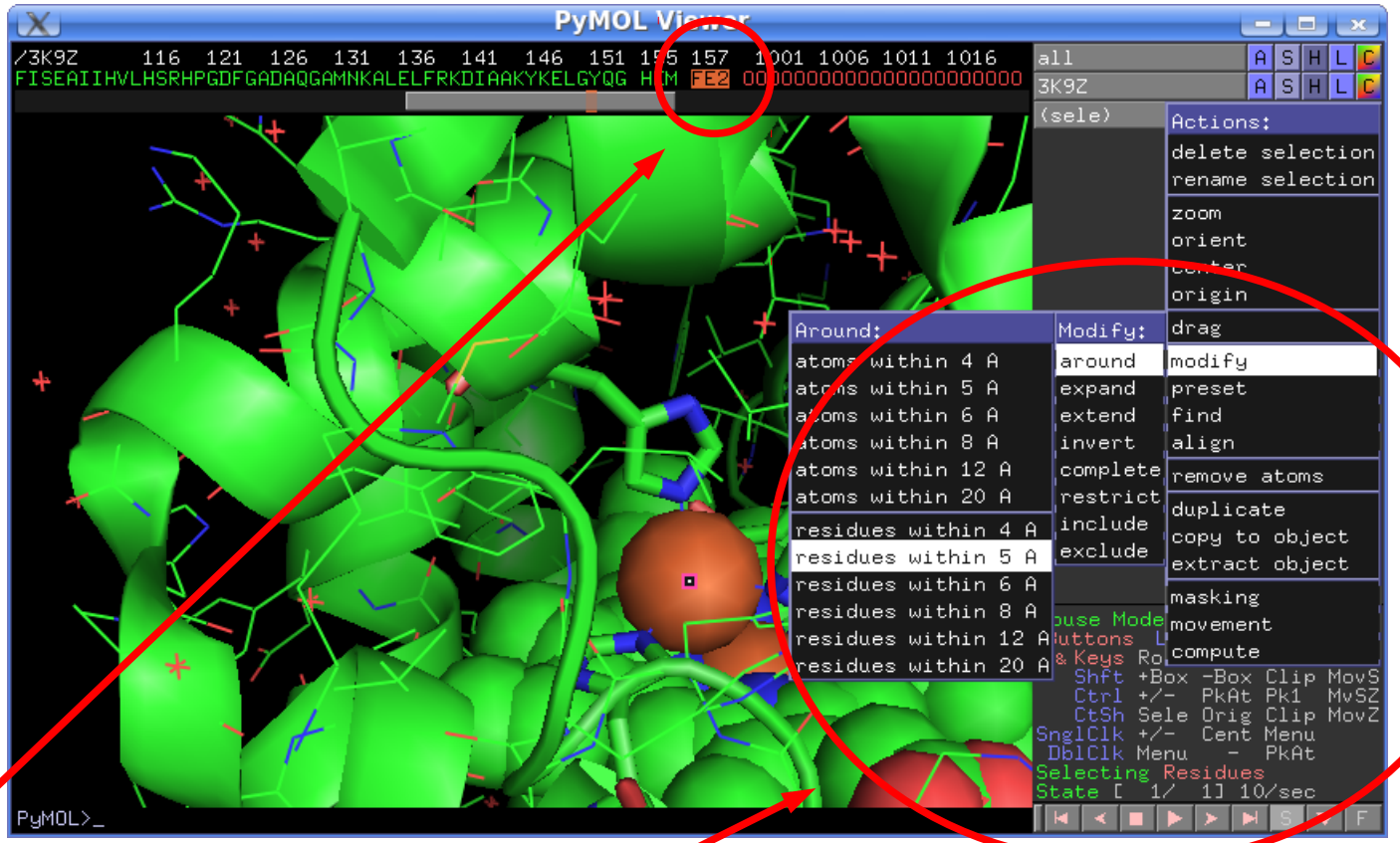
The Sequence button makes the molecule sequence appears.
You can then select one or more residues from the sequence.

Pymol Selection



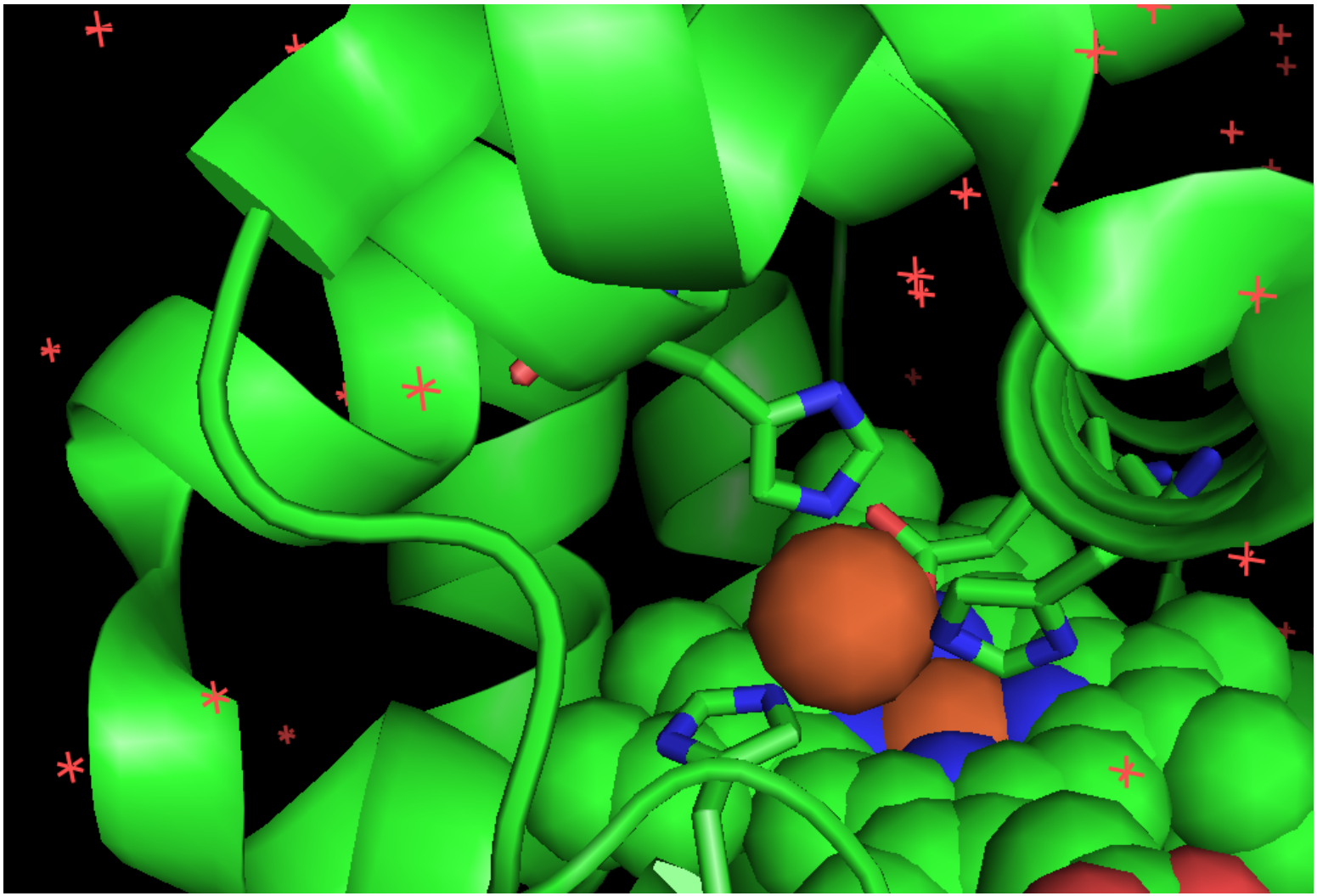
Select HEM and FE2 molecules, and try « Show Spheres »
On the selection to change the appearance of these residues

Pymol – Finding neighbors



Select FE2 molecule only, and « Action – modify – around – residues within 5A»
Then, on this selection, make a « Show Sticks » to hilight neighbors
(and a « Hide - Lines » on 3K9Z Object to make it clearer)

Pymol – Residues around FE2



For nice images: « ray » (command-line) and « png name.png, dpi=1000 »

What are the goals of the simulations?

- Identify the questions to answer from theoretical work:
  - What theoretical approach to use (classical MD, implicit solvent, etc...)
  - To determine what simulations are needed, what is included in the simluations (water, ligands, etc...)
  - How many trajectories, and how long will they be (equilibrium, number of events to observe, statistics, etc...)
  - What will the analysis consist of?
- For the present work: determine if the mutations will be structurally viable for NOR

Reproduce the simulations from the publication

- Fe2 is replaced with a Zn(II)
  - force field parameters are not available for this atom
  - We will use the crystal structure from the work instead of a modified Wild-type myoglobin structure

- Overview of System building: 2 steps

  1) Automated step: only the protein, solvant and ions

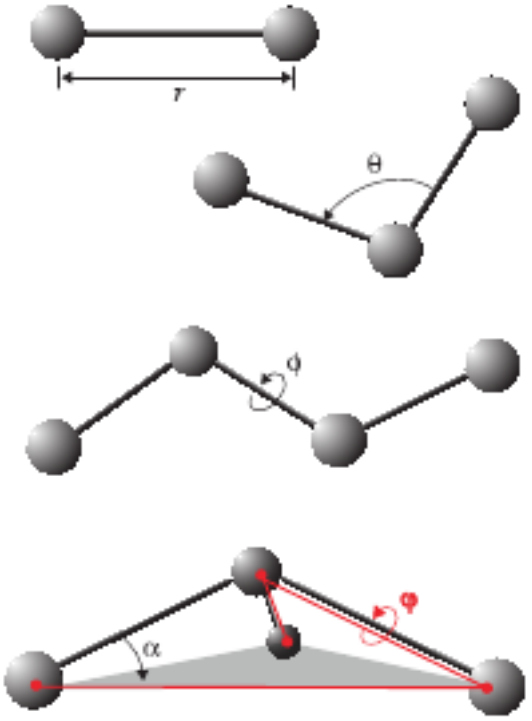
  2) Manual edition: addition of ZN(II) and heme to the system

What is needed to start MD simulations

- Molecular Topologies and force field parameters.
- Protein Structure File (PSF)
  - Built from topologies
- Protein coordinates CRD
- Simulation Conditions:
  - Solvent and Ionic strength
  - Temperature
  - Constant volume or Pressure? (Mostly constant pressure)

CHARMM Topology

Defines which atoms are connected to one another through chemical bonds, including **bonds**, **angles**, **dihedrals** and **impropers**:



Topology example

Alanine Topology Entry

```
RESI ALA          0.00
GROUP
ATOM N    NH1    -0.47  !     |
ATOM HN   H       0.31  !   HN-N
ATOM CA   CT1     0.07  !     |     HB1
ATOM HA   HB      0.09  !     |    /
GROUP                   !   HA-CA--CB-HB2
ATOM CB   CT3    -0.27  !     |    \
ATOM HB1  HA      0.09  !     |     HB3
ATOM HB2  HA      0.09  !    O=C
ATOM HB3  HA      0.09  !     |
GROUP                   !
ATOM C    C       0.51
ATOM O    O      -0.51
BOND CB CA   N  HN   N  CA
BOND C  CA   C  +N   CA HA   CB HB1   CB HB2   CB HB3
DOUBLE O  C
IMPR N -C CA HN   C CA +N O
DONOR HN N
ACCEPTOR O C
IC -C   CA   *N   HN    1.3551 126.4900  180.0000 115.4200  0.9996
IC -C   N    CA   C     1.3551 126.4900  180.0000 114.4400  1.5390
IC N    CA   C    +N    1.4592 114.4400  180.0000 116.8400  1.3558
IC +N   CA   *C   O     1.3558 116.8400  180.0000 122.5200  1.2297
IC CA   C    +N   +CA   1.5390 116.8400  180.0000 126.7700  1.4613
IC N    C    *CA  CB    1.4592 114.4400  123.2300 111.0900  1.5461
IC N    C    *CA  HA    1.4592 114.4400 -120.4500 106.3900  1.0840
IC C    CA   CB   HB1   1.5390 111.0900  177.2500 109.6000  1.1109
IC HB1  CA   *CB  HB2   1.1109 109.6000  119.1300 111.0500  1.1119
IC HB1  CA   *CB  HB3   1.1109 109.6000 -119.5800 111.6100  1.1114
```

New redisu with the name ALA and total charge of 0.00

Atoms following are part of a group carrying an integer charge

New atom named N, atom type NH1 and partial charge -0.47 (electrostatic) (! is a comment signal)

Indicates sets of 2 atoms connected by a single bond

Indicates sets of 4 atoms implied in an improper bond

Donor and Acceptor (obsolete)

Internal coordinates used to build missing coordinates (optional)

From Topology Tutorial www.ks.uiuc.edu/Training/tutorials

CHARMM Parameters

$$
\begin{aligned}
V = & \sum_{\text{bonds}} k_b \left(b - b_0\right)^2 + \sum_{\text{angles}} k_\theta \left(\theta - \theta_0\right)^2 + \sum_{\text{dihedrals}} k_\phi \left[1 + \cos\left(n\phi - \delta\right)\right] \\
& + \sum_{\text{impropers}} k_\omega \left(\omega - \omega_0\right)^2 + \sum_{\text{Urey-Bradley}} k_u \left(u - u_0\right)^2 \\
& + \sum_{\text{nonbonded}} \epsilon \left[\left(\frac{R_{\min_{ij}}}{r_{ij}} \right)^{12} - \left(\frac{R_{\min_{ij}}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{\epsilon r_{ij}}
\end{aligned}
\tag{1}
$$

The parameter file contains all the force field parameters for a given set of atom types

Sources of coordinates

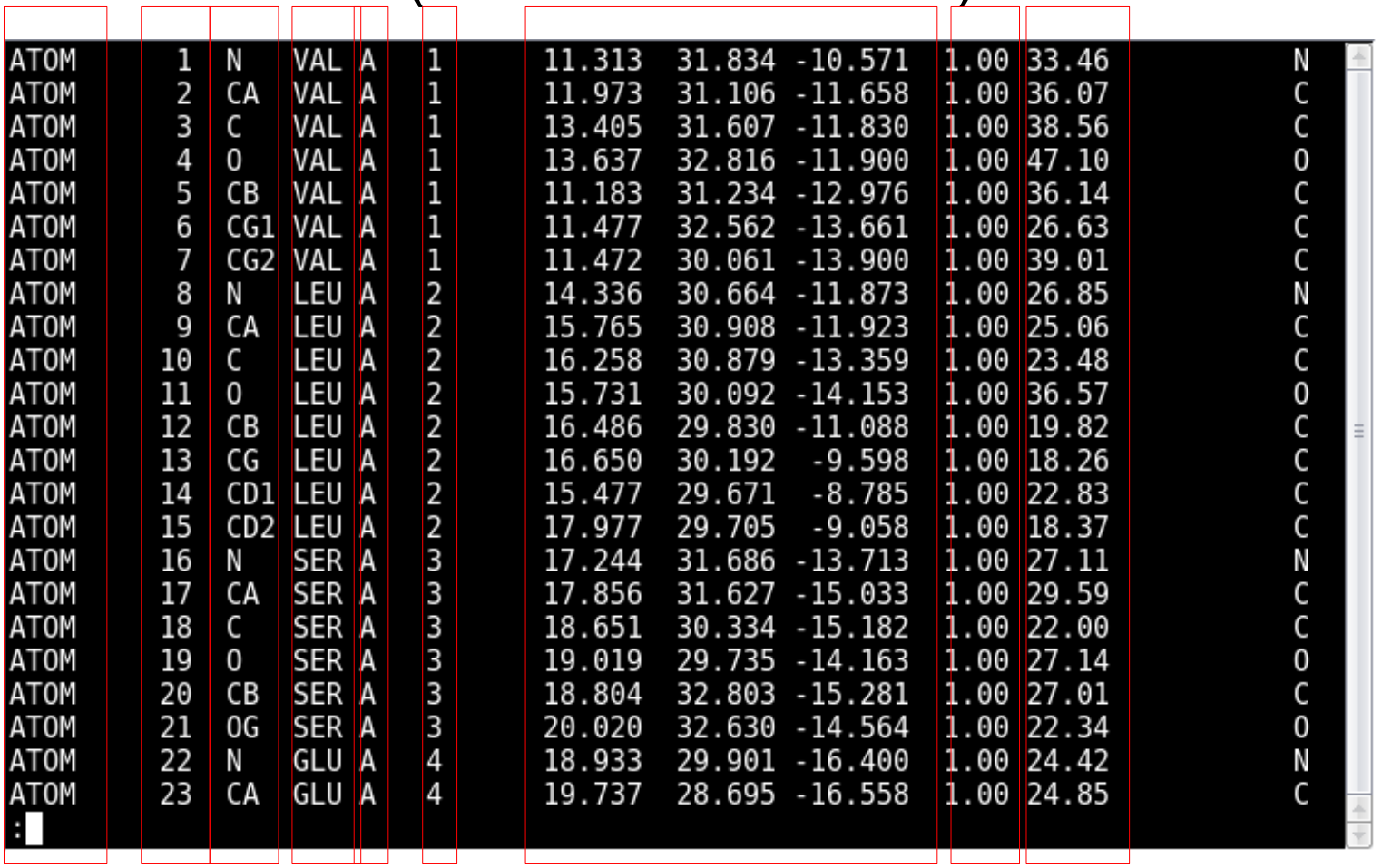
- The **Protein DataBank** (PDB)
  - Web site: PDB.org
  - Proteins (63956 structures) and some ligands
- Homology modeling (structure prediction)

For protein-membrane complexes:

- Charmm-gui to build systems
- Orientation of Proteins in Membranes: OPM http://opm.phar.umich.edu/

PDB files contain coordinates

(and much more...)



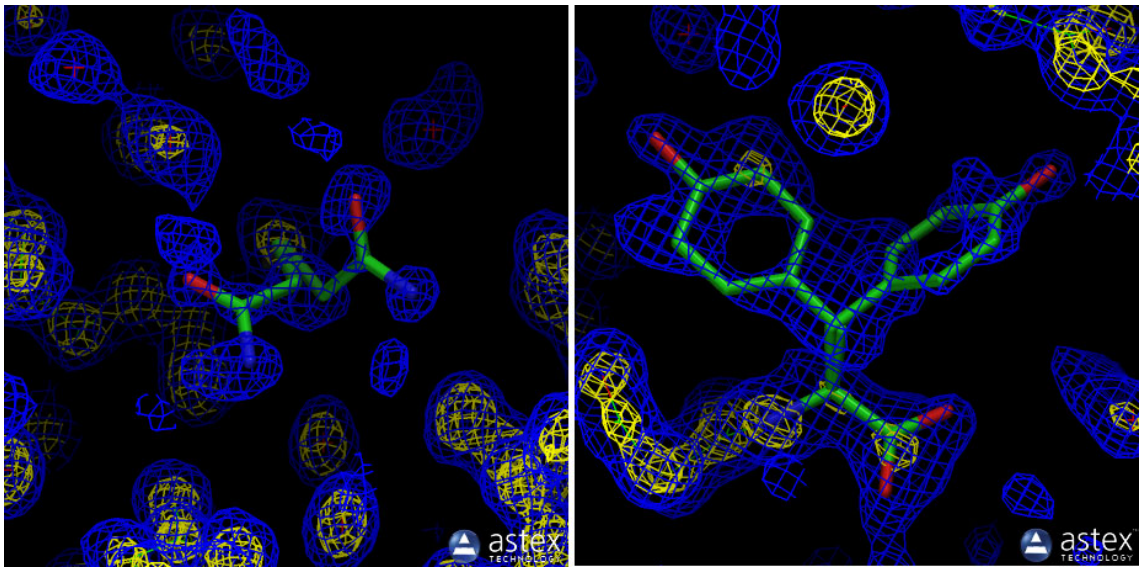
```
ATOM      1  N   VAL A   1      11.313  31.834 -10.571  1.00 33.46           N
ATOM      2  CA  VAL A   1      11.973  31.106 -11.658  1.00 36.07           C
ATOM      3  C   VAL A   1      13.405  31.607 -11.830  1.00 38.56           C
ATOM      4  O   VAL A   1      13.637  32.816 -11.900  1.00 47.10           O
ATOM      5  CB  VAL A   1      11.183  31.234 -12.976  1.00 36.14           C
ATOM      6  CG1 VAL A   1      11.477  32.562 -13.661  1.00 26.63           C
ATOM      7  CG2 VAL A   1      11.472  30.061 -13.900  1.00 39.01           C
ATOM      8  N   LEU A   2      14.336  30.664 -11.873  1.00 26.85           N
ATOM      9  CA  LEU A   2      15.765  30.908 -11.923  1.00 25.06           C
ATOM     10  C   LEU A   2      16.258  30.879 -13.359  1.00 23.48           C
ATOM     11  O   LEU A   2      15.731  30.092 -14.153  1.00 36.57           O
ATOM     12  CB  LEU A   2      16.486  29.830 -11.088  1.00 19.82           C
ATOM     13  CG  LEU A   2      16.650  30.192  -9.598  1.00 18.26           C
ATOM     14  CD1 LEU A   2      15.477  29.671  -8.785  1.00 22.83           C
ATOM     15  CD2 LEU A   2      17.977  29.705  -9.058  1.00 18.37           C
ATOM     16  N   SER A   3      17.244  31.686 -13.713  1.00 27.11           N
ATOM     17  CA  SER A   3      17.856  31.627 -15.033  1.00 29.59           C
ATOM     18  C   SER A   3      18.651  30.334 -15.182  1.00 22.00           C
ATOM     19  O   SER A   3      19.019  29.735 -14.163  1.00 27.14           O
ATOM     20  CB  SER A   3      18.804  32.803 -15.281  1.00 27.01           C
ATOM     21  OG  SER A   3      20.020  32.630 -14.564  1.00 22.34           O
ATOM     22  N   GLU A   4      18.933  29.901 -16.400  1.00 24.42           N
ATOM     23  CA  GLU A   4      19.737  28.695 -16.558  1.00 24.85           C
:
```

1 2 3 4 5 6 7 8 9

1. ATOM (protein)
   HETATM (all others)
2. Number in sequence
3. Atom names
4. Residu name
5. Chain classification (many subunits)
6. Residu number
7. x,y,z coords
8. Occupancy
9. Temperature factor

PDB file Pitfalls

- Missing portions of the molecule and hydrogen atoms
  - Not observed experimentally (usualy extremities)
- Multiple conformations (quantified by occupancy):
  - Sidechain mobility
  - Substrate may bind in two conformations
  - Metal ion may be bound to only a few of the molecules



PSF file

- Protein Structure File (PSF) contains all of the molecule-specific information needed to apply a particular force field to a molecular system
- The PSF file contains six main sections of interest: **atoms, bonds, angles, dihedrals, impropers, and cross-terms**
- CHARMM and XPLOR formats

PSF file example

The title and atom records:

```
PSF CMAP

       6 !NTITLE
 REMARKS original generated structure x-plor psf file
 REMARKS 2 patches were applied to the molecule.
 REMARKS topology top_all27_prot_lipid.inp
 REMARKS segment U { first NTER; last CTER; auto angles dihedrals }
 REMARKS defaultpatch NTER U:1
 REMARKS defaultpatch CTER U:76

    1231 !NATOM
       1 U    1    MET  N    NH3   -0.300000       14.0070           0
       2 U    1    MET  HT1  HC     0.330000        1.0080           0
       3 U    1    MET  HT2  HC     0.330000        1.0080           0
       4 U    1    MET  HT3  HC     0.330000        1.0080           0
       5 U    1    MET  CA   CT1    0.210000       12.0110           0
       6 U    1    MET  HA   HB     0.100000        1.0080           0
       7 U    1    MET  CB   CT2   -0.180000       12.0110           0
```

The fields in the atom section are **atom ID, segment name, residue ID, residue name, atom name, atom type, charge, mass, and an unused 0**

PSF file example (2)

The covalent bond section lists four pairs of atoms per line:

```
1237 !NBOND: bonds
   1       5       2       1       3       1       4       1
   5       6       7       5       7       8       7       9
  10       7      10      11      10      12      13      10
```

The angle section lists three triples of atoms per line:

```
2257 !NTHETA: angles
   1       5       6       1       5      18       2       1       5
   2       1       4       2       1       3       3       1       5
   3       1       4       4       1       5       5      18      19
```

PSF file example (3)

The dihedral and improper sections list two quadruples of atoms per line:

```
    3293 !NPHI: dihedrals
       1       5       7      10       1       5       7       8
       1       5       7       9       1       5      18      20
       1       5      18      19       2       1       5       7

     204 !NIMPHI: impropers
      18       5      20      19      20      18      22      21
      30      32      27      31      30      27      32      31
      32      30      33      34      32      30      34      33
```

The cross-term sections list two quadruples of atoms per line:

```
    74 !NCRTERM: cross-terms
      18      20      22      35      20      22      35      37
      35      37      39      54      37      39      54      56
      54      56      58      74      56      58      74      76
```

Automated tools to build PSF

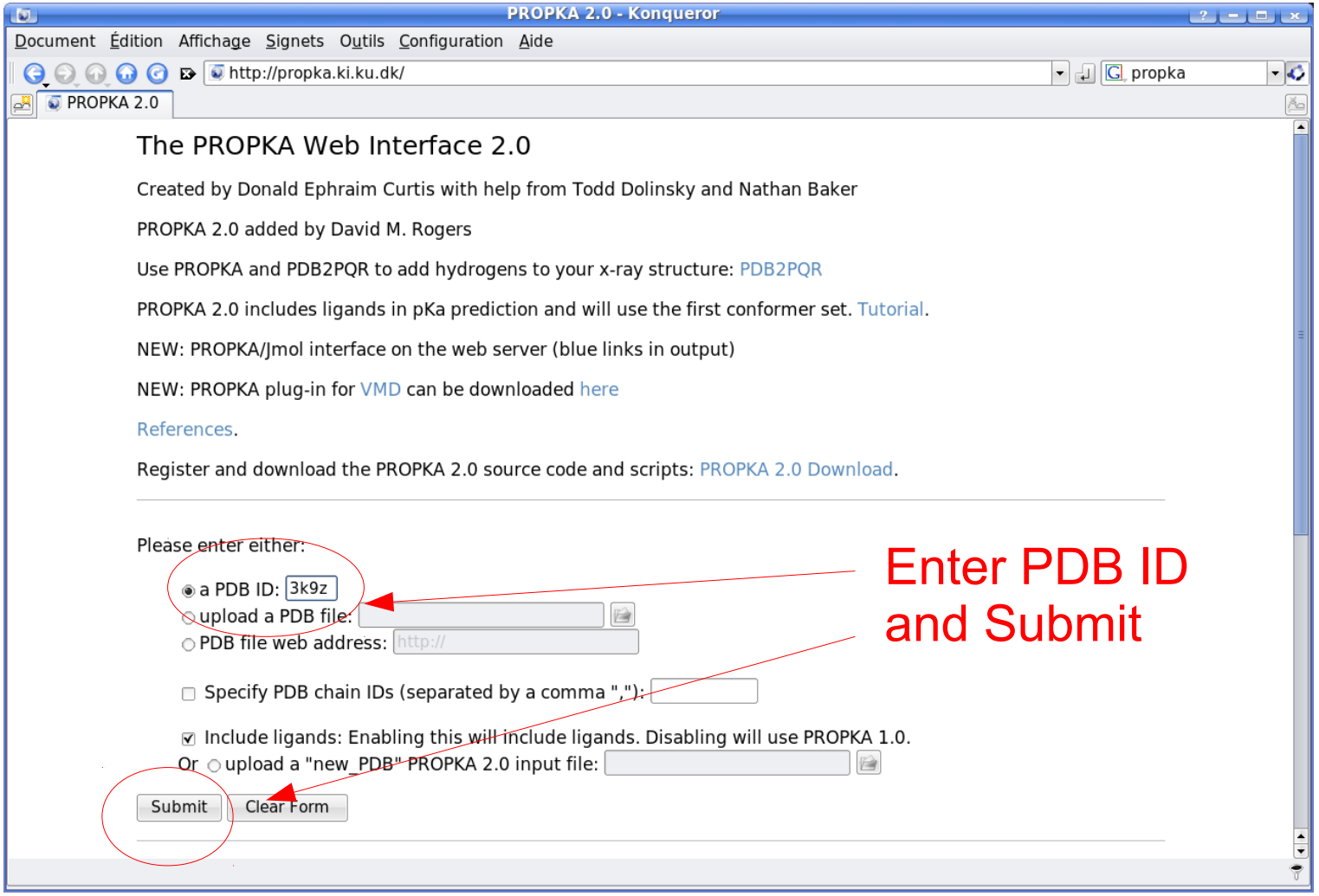
- Easy the task (compared to CHARMM) but have limited features

Include:

- Visual Molecular Dynamics (VMD) (free for academics)
- Charmming (need more development)
- Charmm-gui

ProPKa

- Web interface: http://propka.ki.ku.dk
- A tool to predict the amino acids protonation state, and detect cys-cys disulfide bonds



ProPka results

PROPKA Results for 3k9z

To save these results right click and hit Save As...

Results

PROPKA 2.0 Input

```
 44 ligand atoms:
HEM 1220  C2  CHA  13.908  14.945   9.307
HEM 1221  C2  CHB  14.607  17.338   5.211
HEM 1222  C2  CHC  14.949  21.422   7.647
```

The section of interest:

```
---------------------------------------------------
SUMMARY OF THIS PREDICTION
    RESIDUE    pKa   pKmodel   ligand atom-type
   ASP  20A   2.61      3.80
   ASP  27A   3.60      3.80
   ASP  44A   3.82      3.80
   ASP  60A   3.42      3.80
   ASP 122A   3.33      3.80
```

ProPka results of interest

These residues are neighbors to the FE2:

HIS	29A	3.72	6.50
GLU	68A	-10.03	4.50
HIS	43A	-2.04	6.50
HIS	64A	6.20	6.50

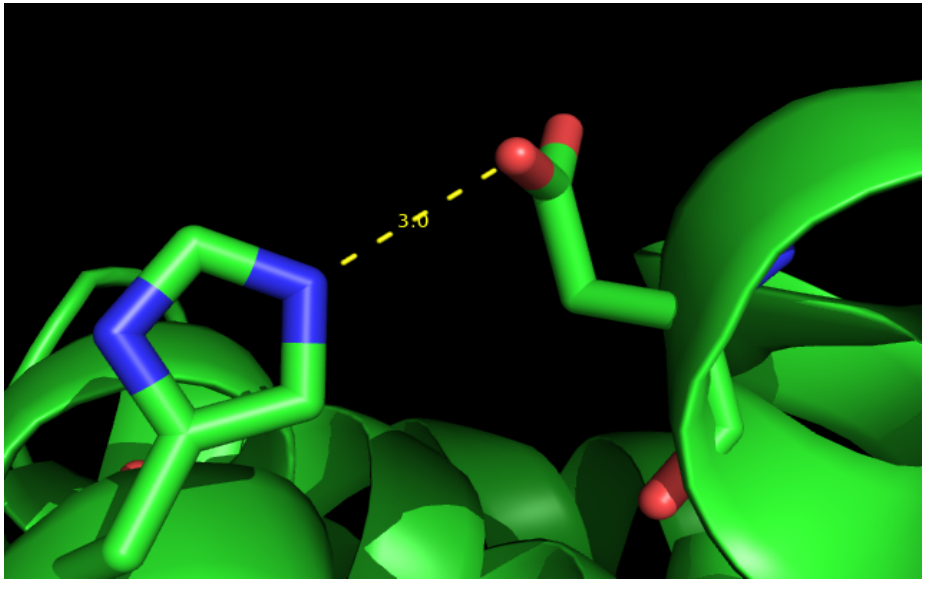
This residue is bound to heme FE:

HIS	93A	-9.61	6.50

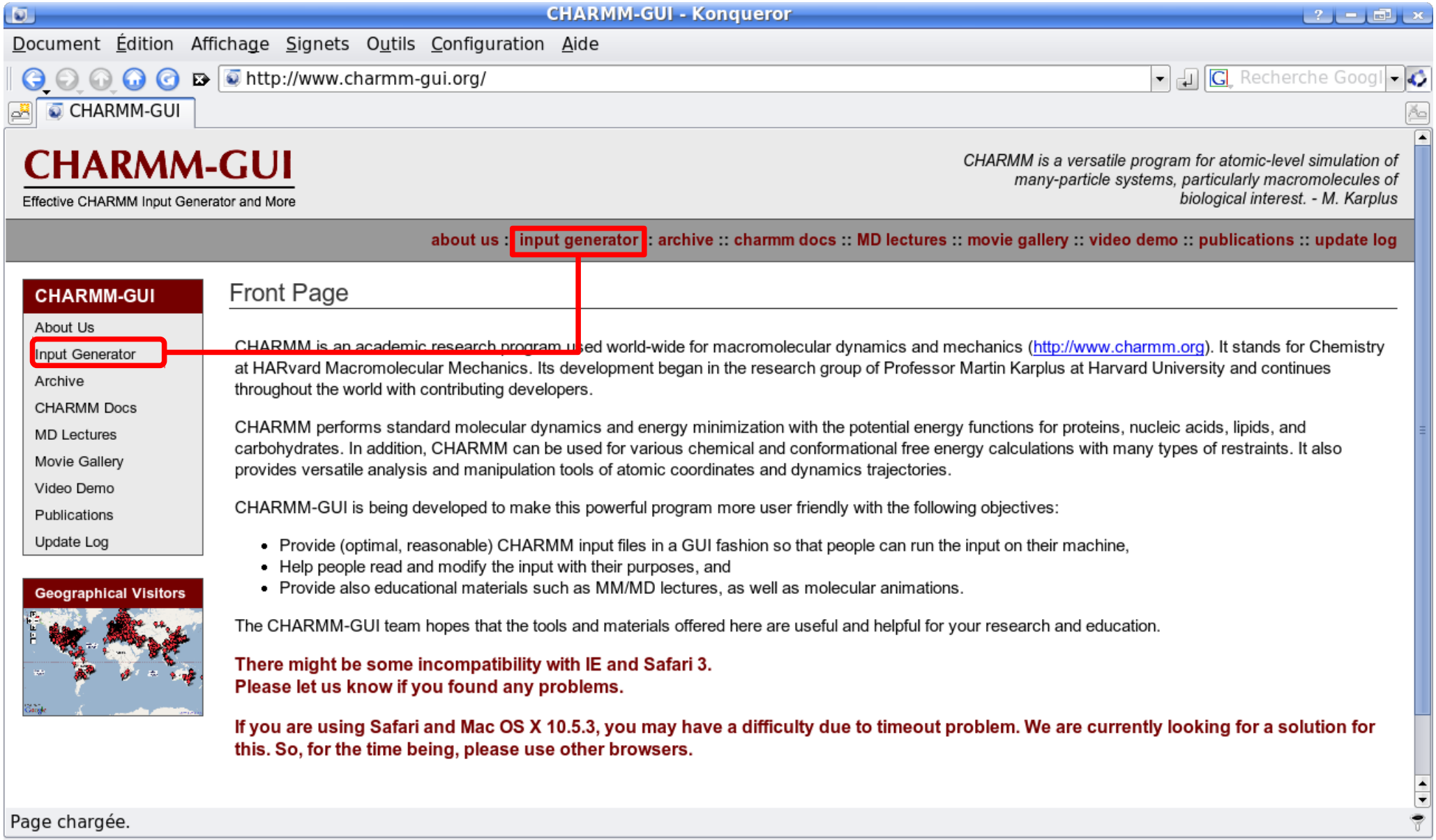
This residue is predicted to be protonated at pH 7:

HIS	12A	7.29	6.50

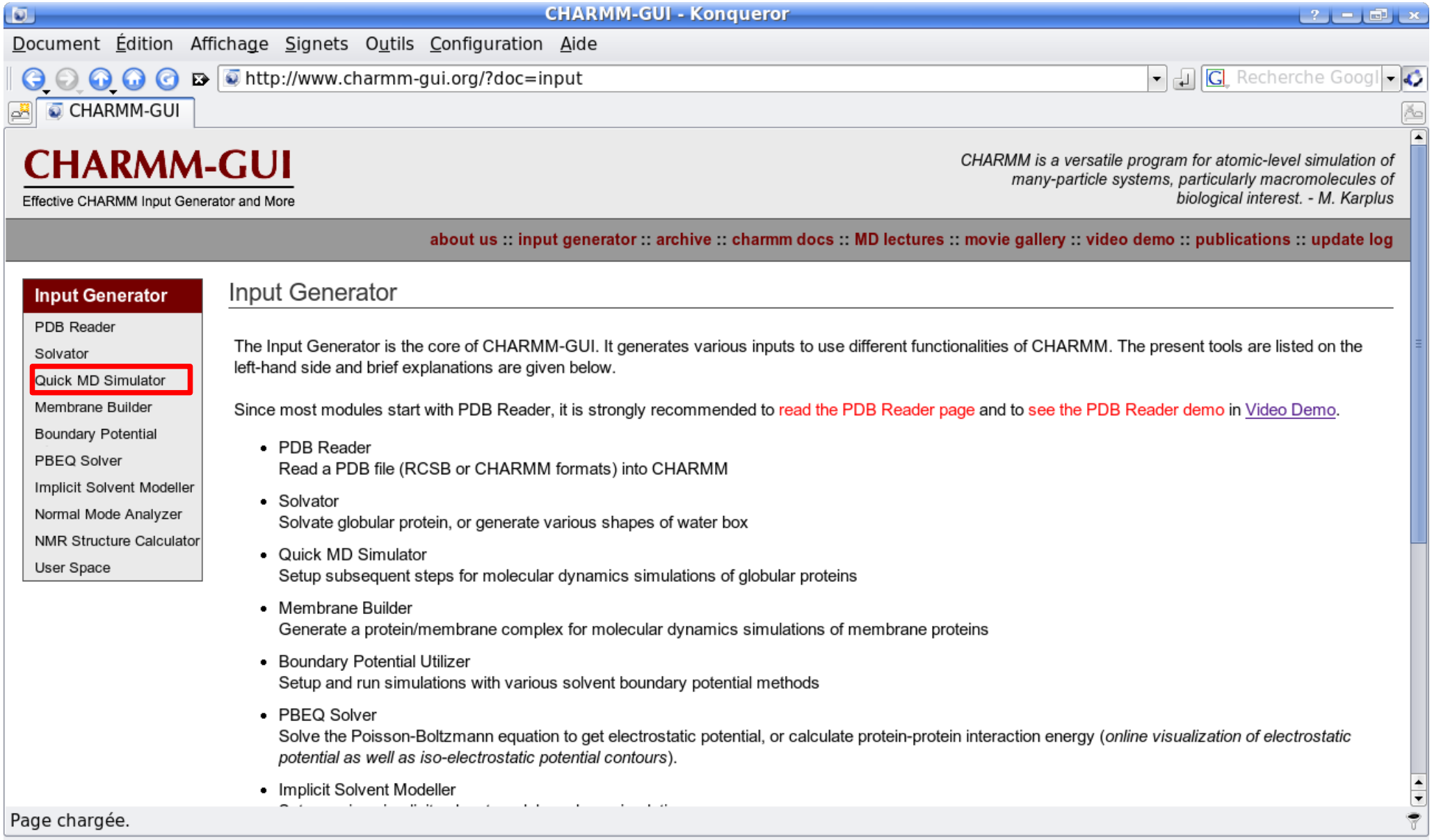
H-bond with Asp122



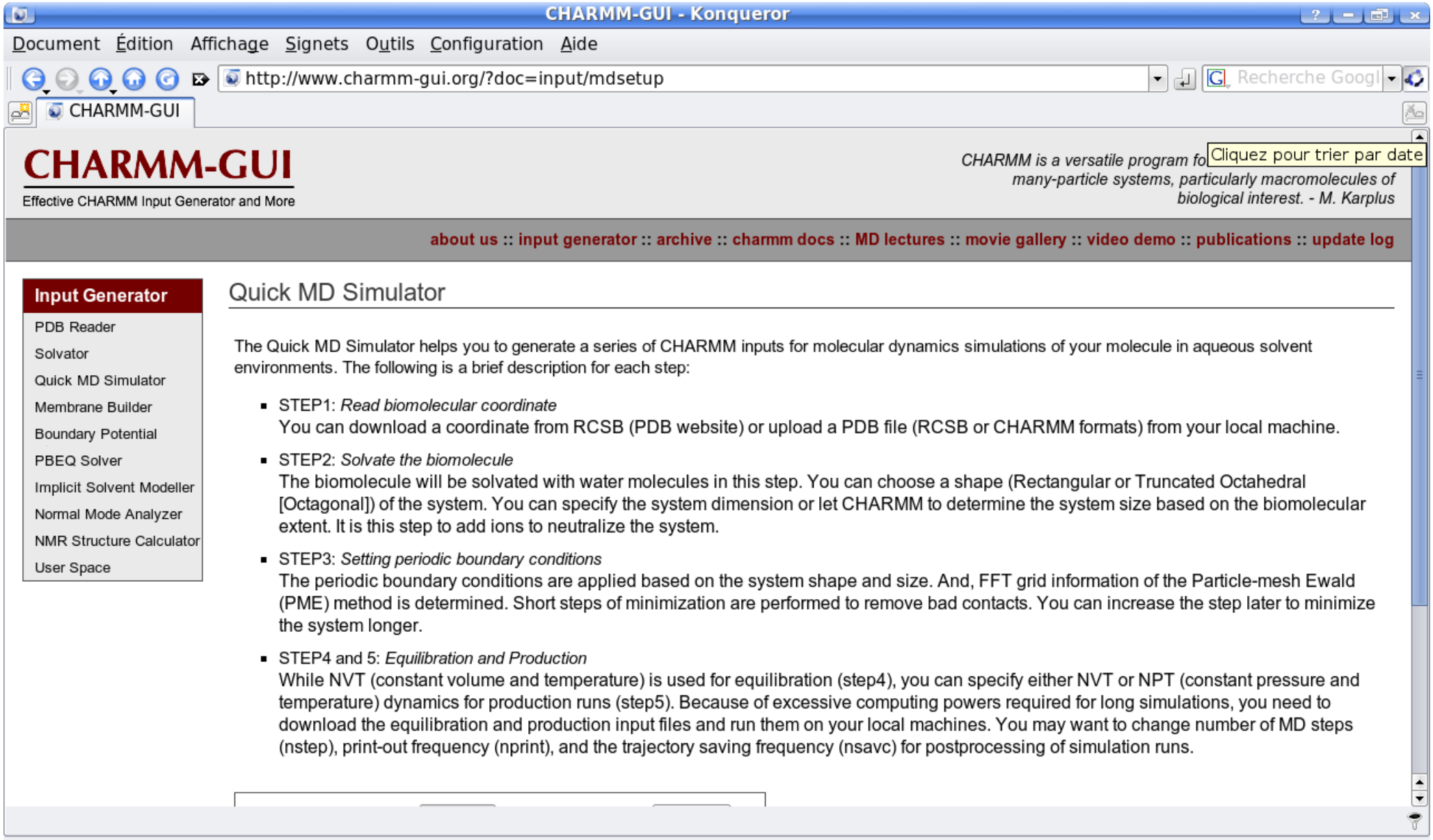
CHARMM-GUI.ORG



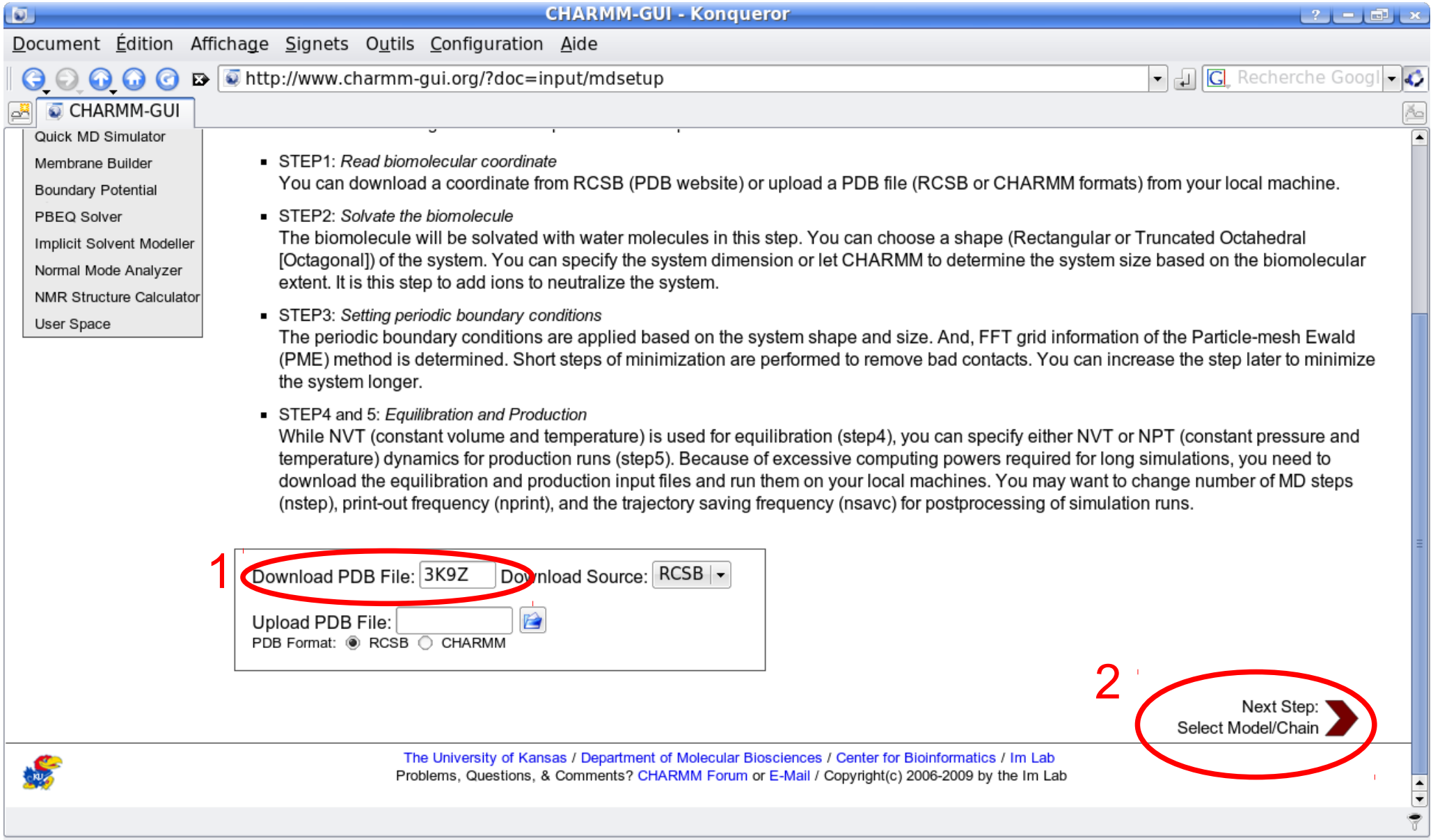
Input Generator



Quick MD Simulator



Read PDB File



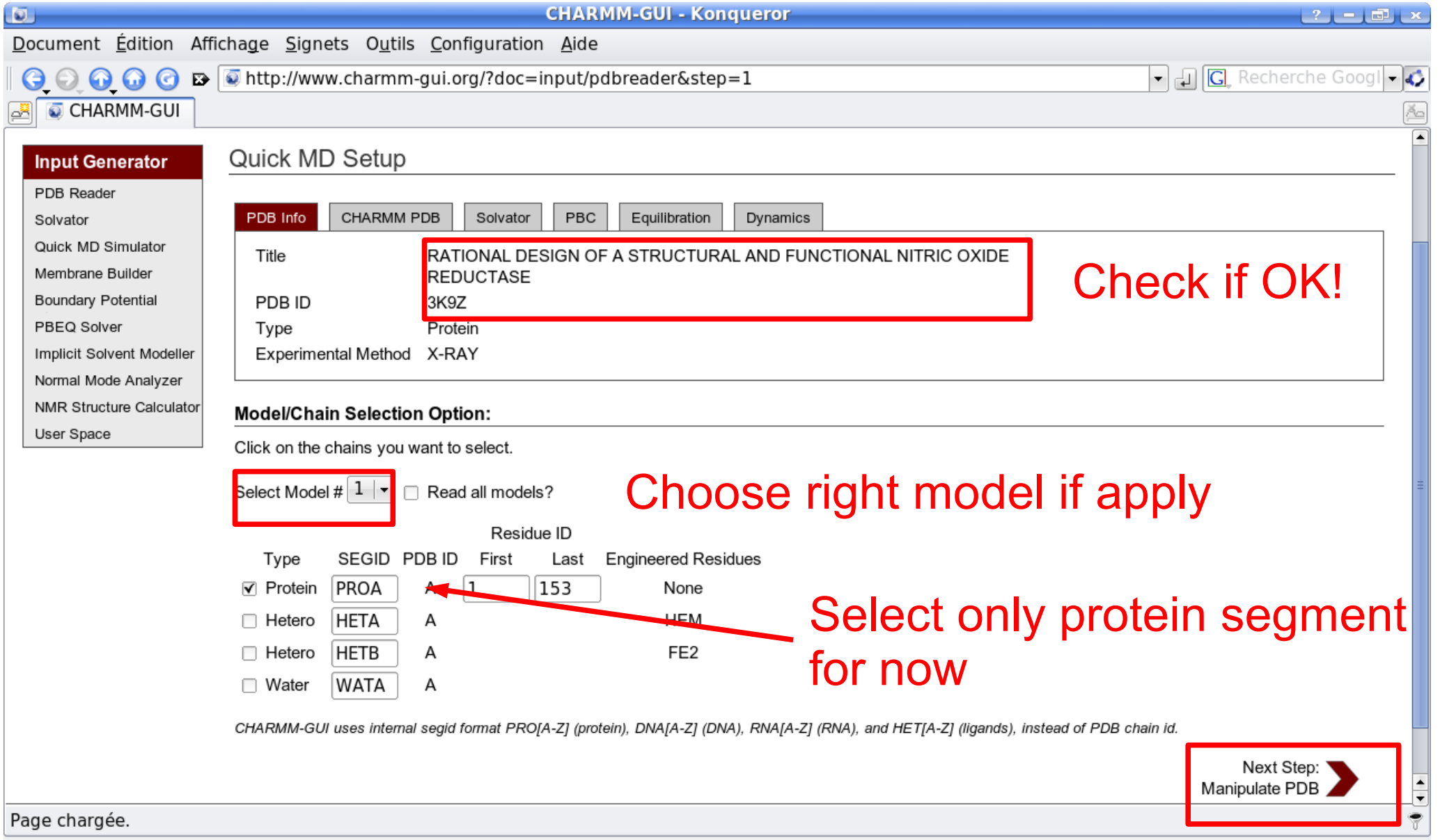
CHARMM-GUI Server load

To preserve the server from High load charges:

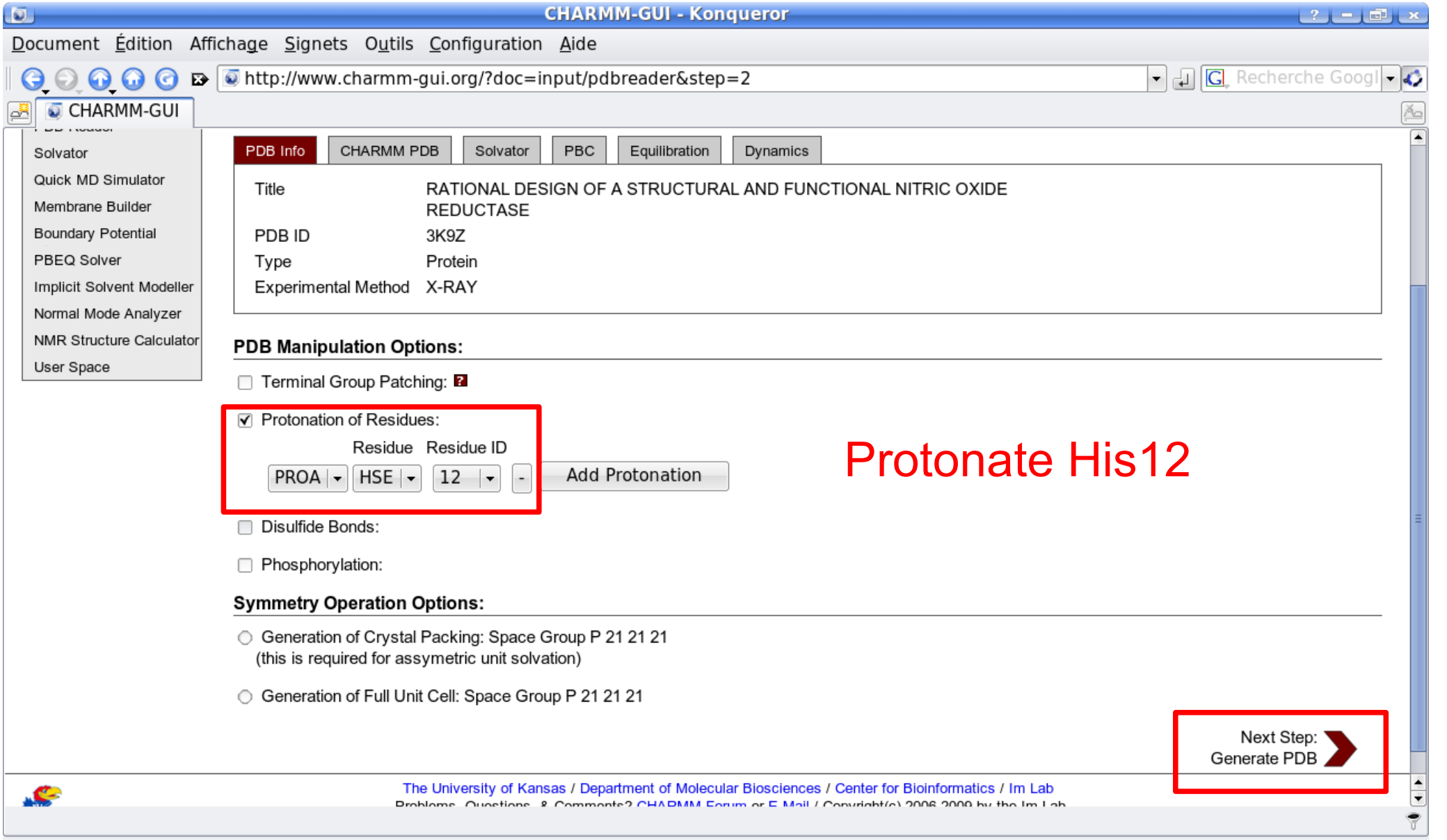
Only 3-5 users should submit a job

at the same time

Quick MD Setup

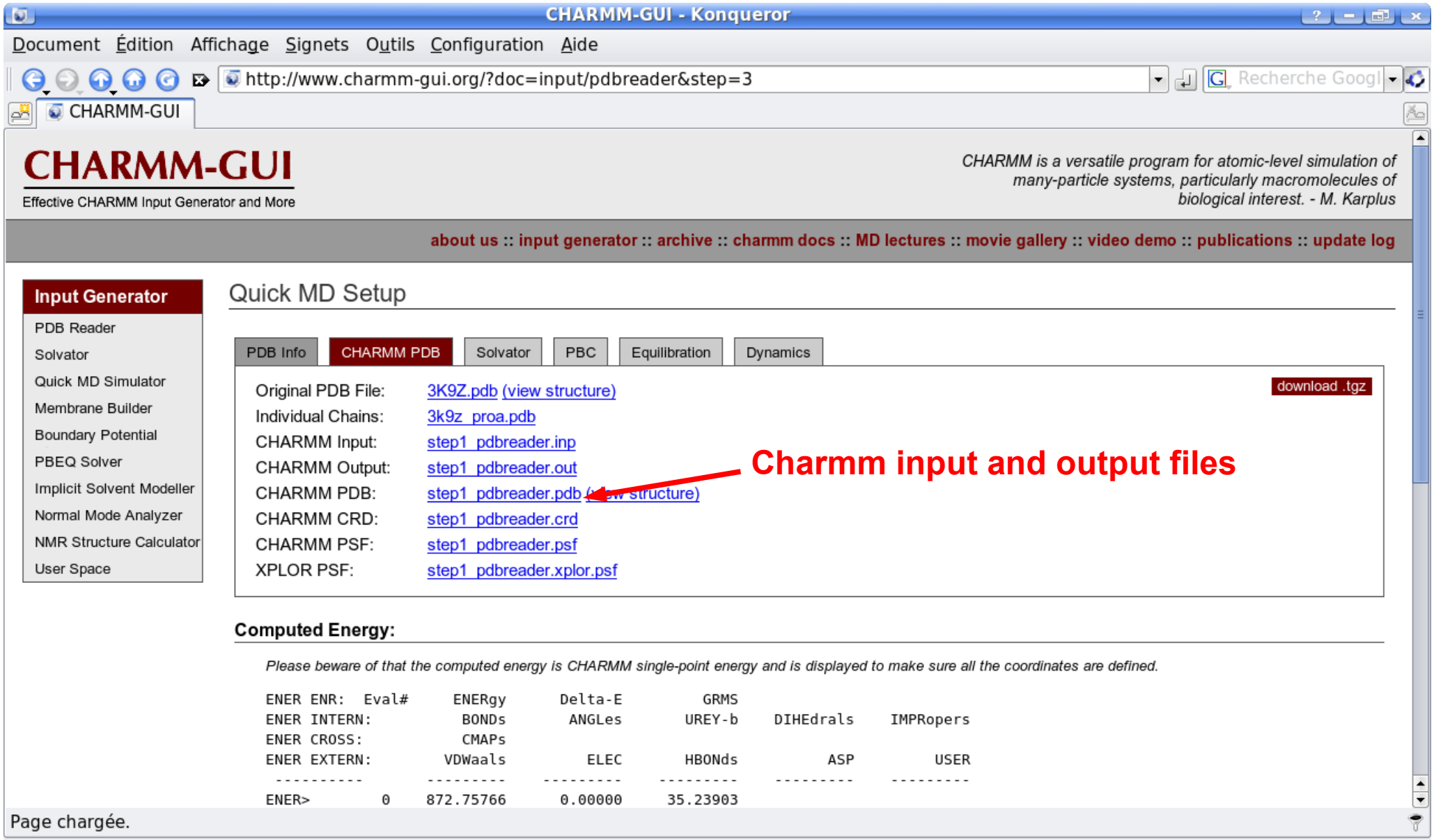


Quick MD Setup 2



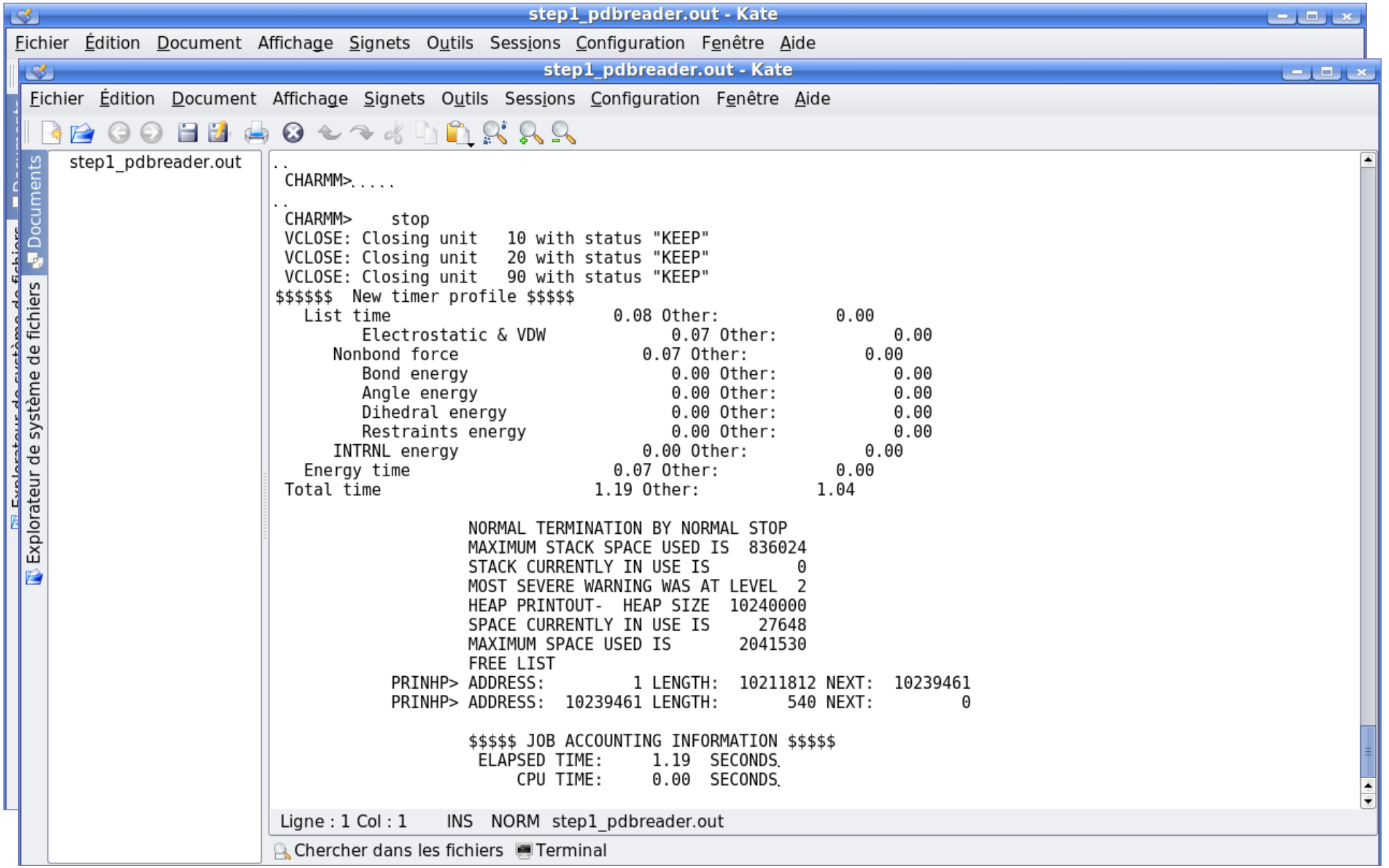
Protonate His12

PSF and coordinates are ready

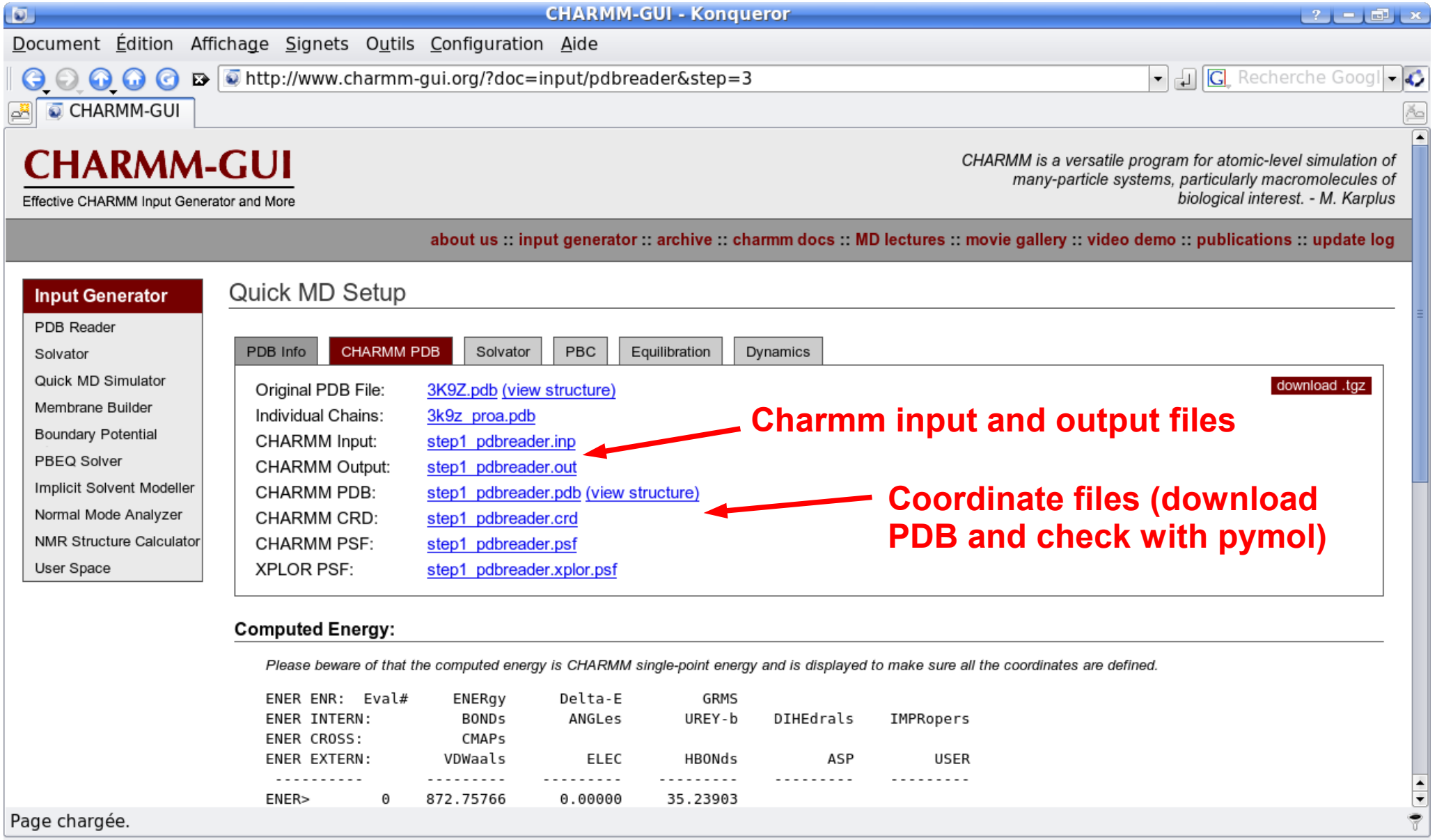


Have a look at the input file to get familiar with CHARMM commands

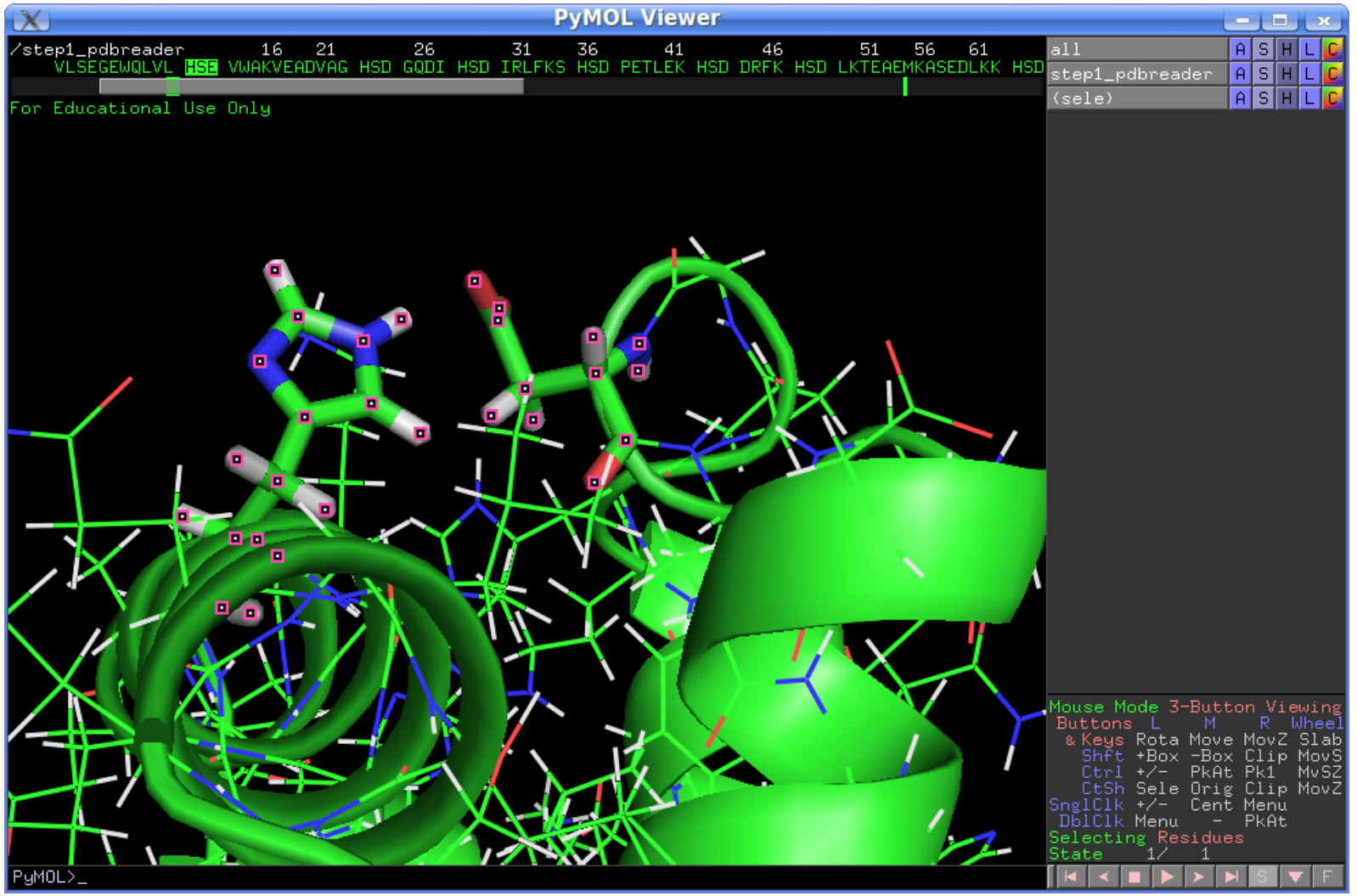
Always check the output files!



PSF and coordinates are ready

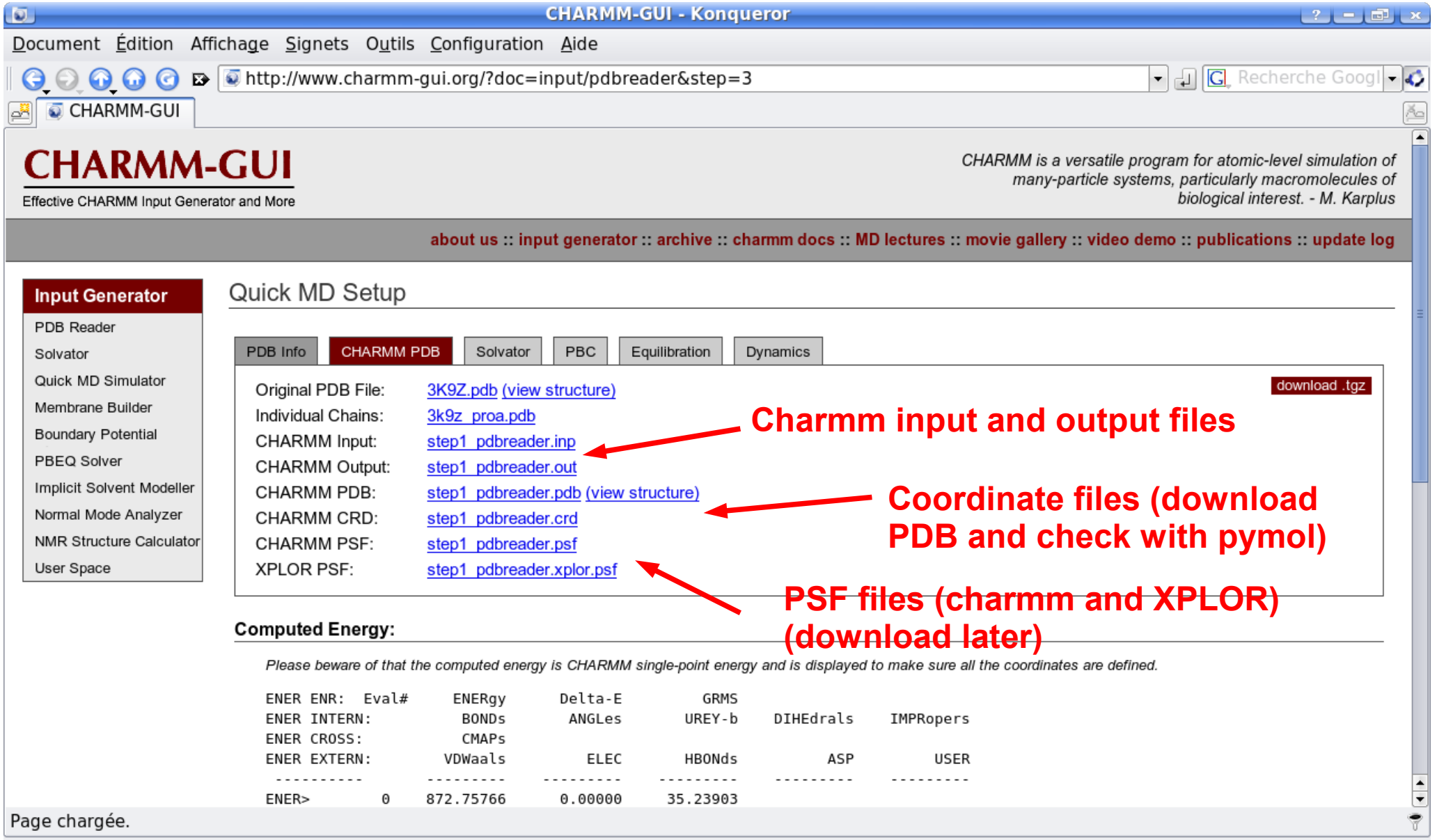


Check with Pymol

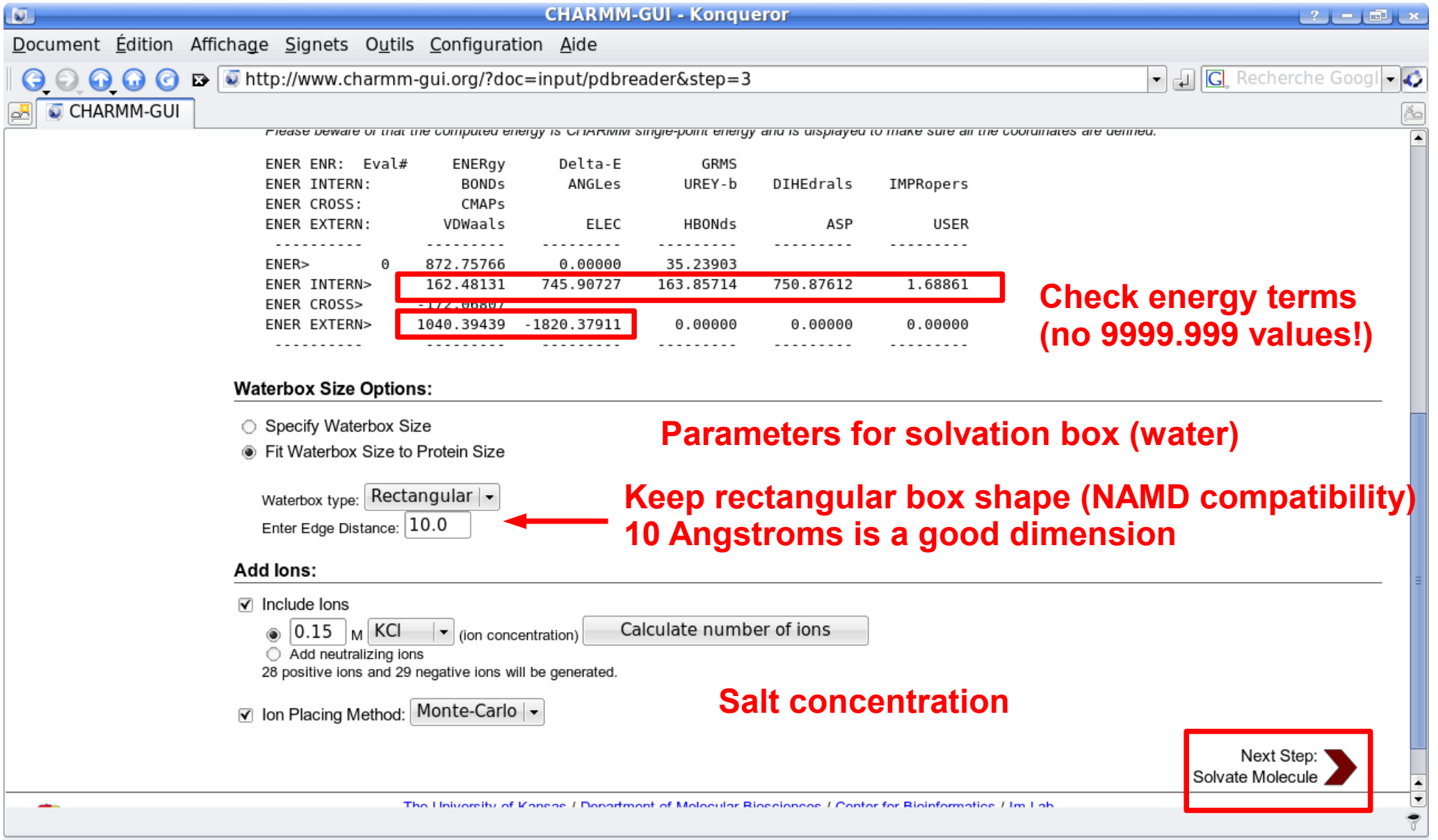


His12 is rightly protonated (HBond with Asp122)

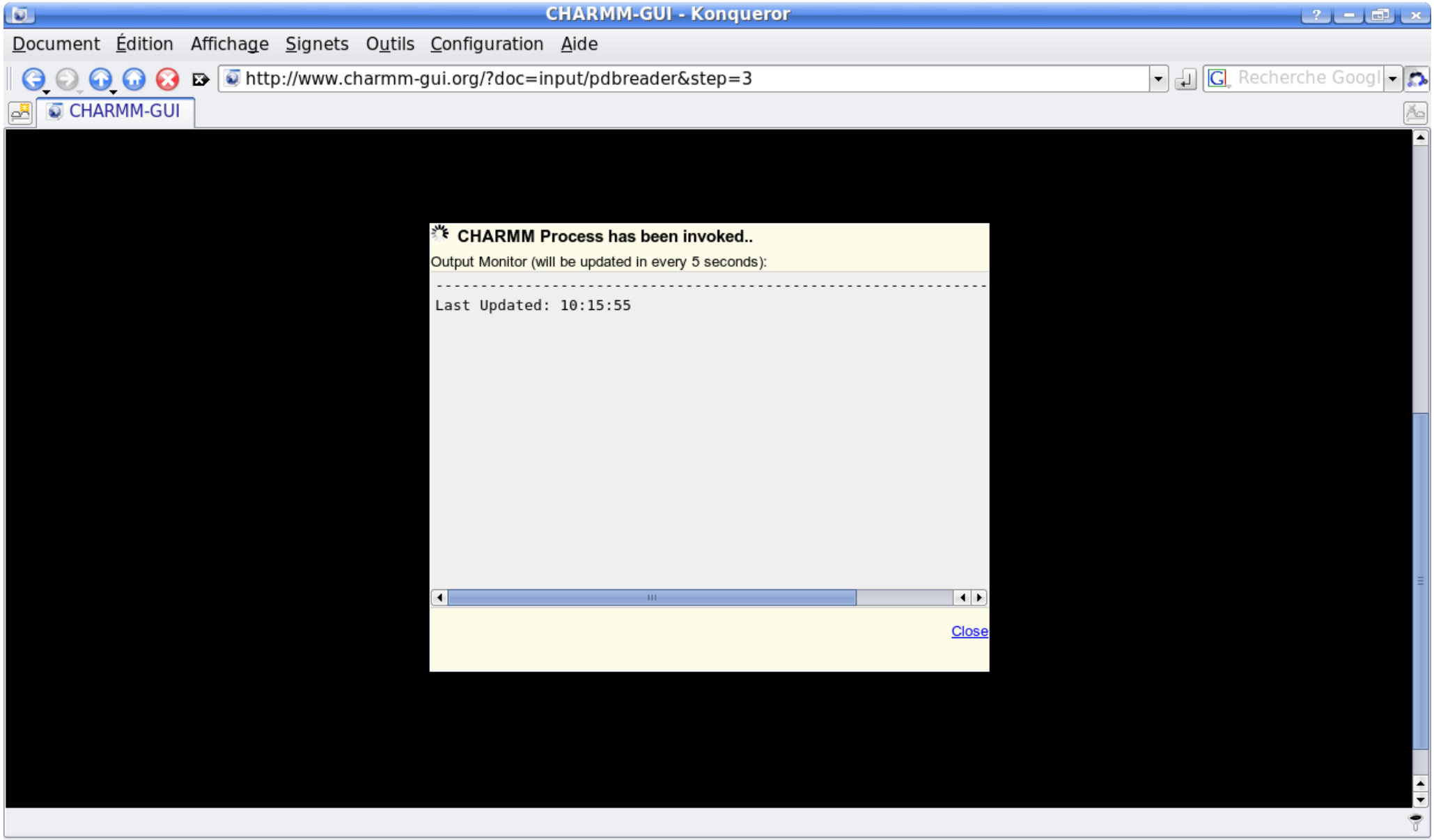
PSF and coordinates are ready



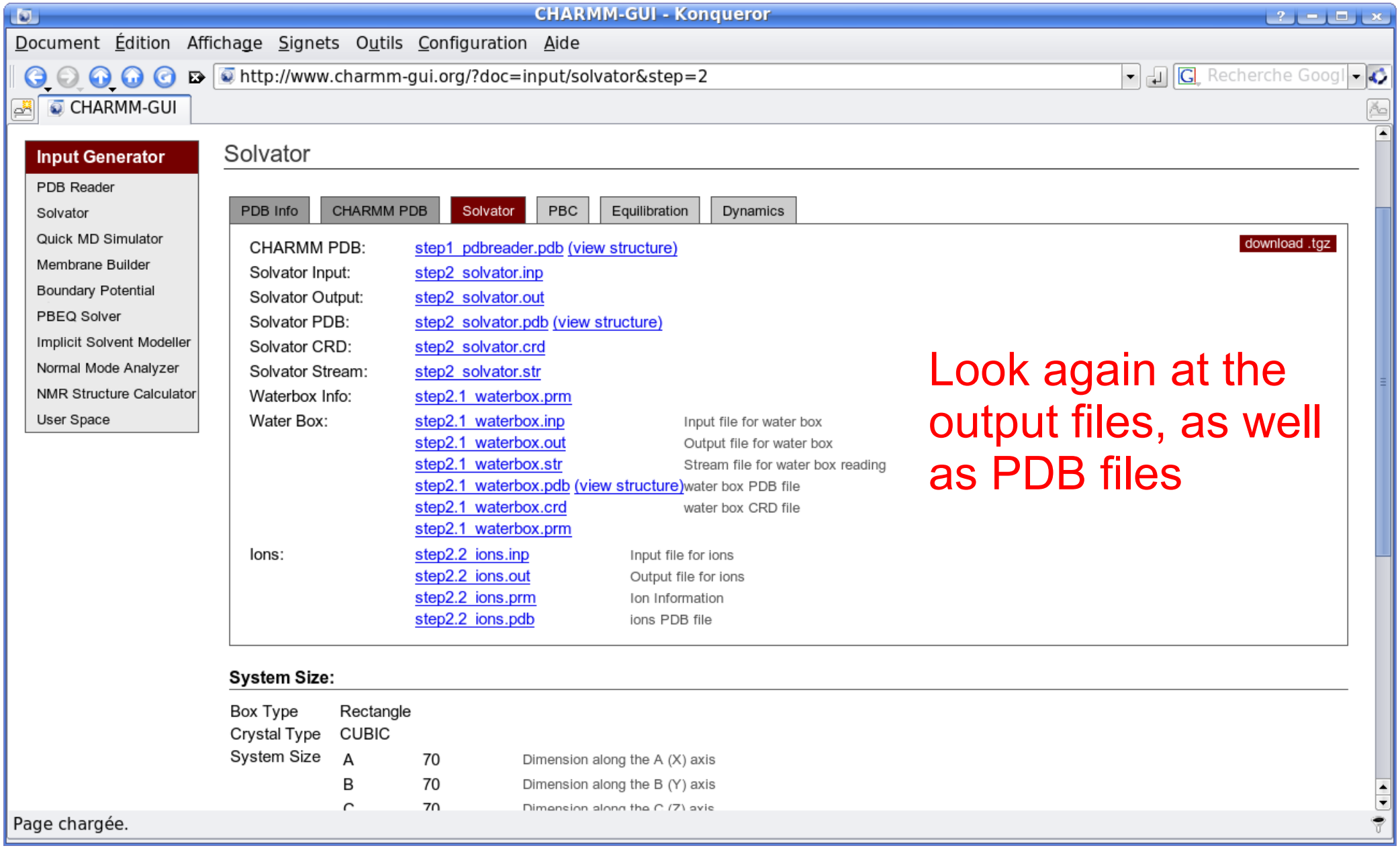
STEP2: Adding solvent



It takes a little time...

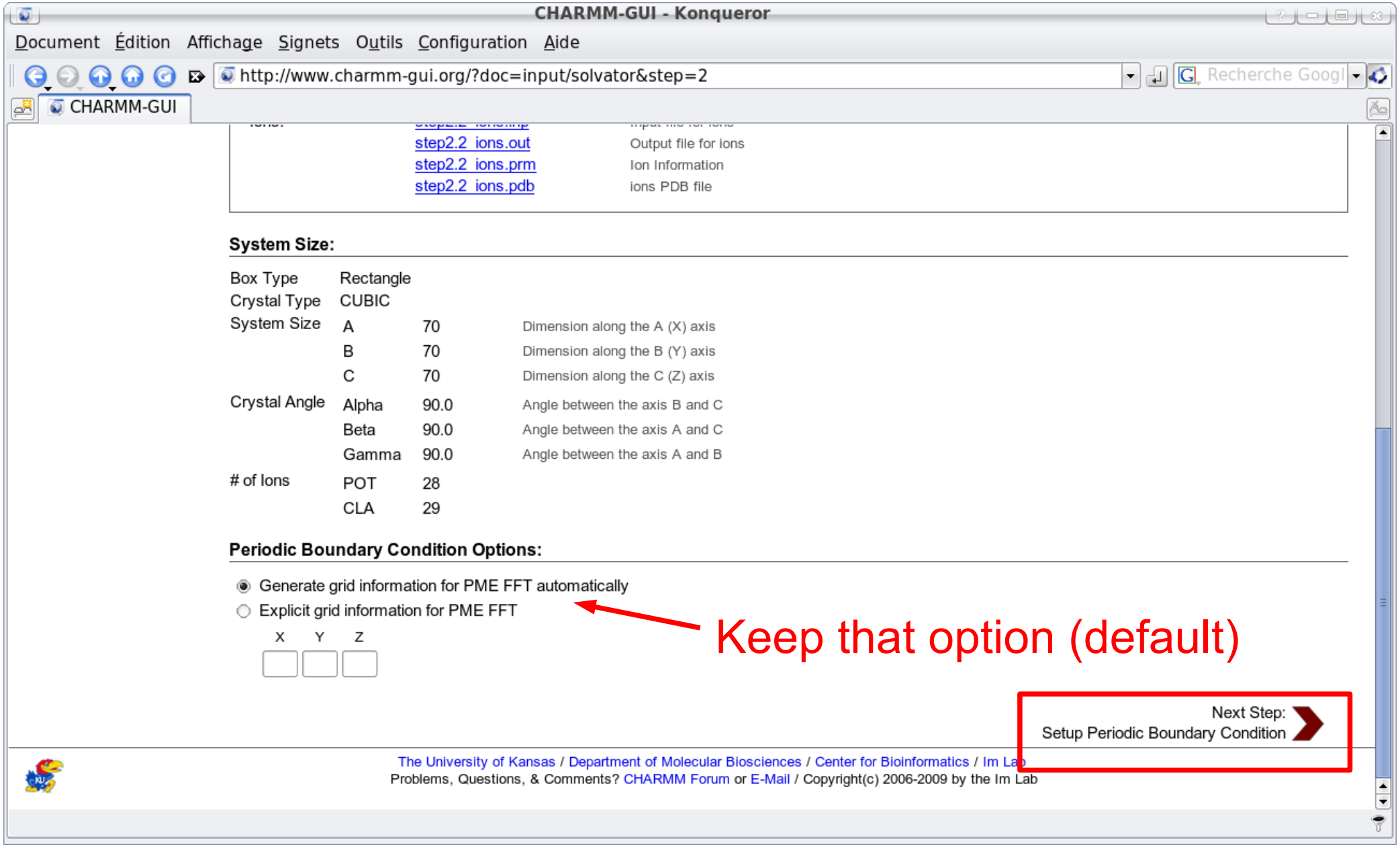


Solvent added

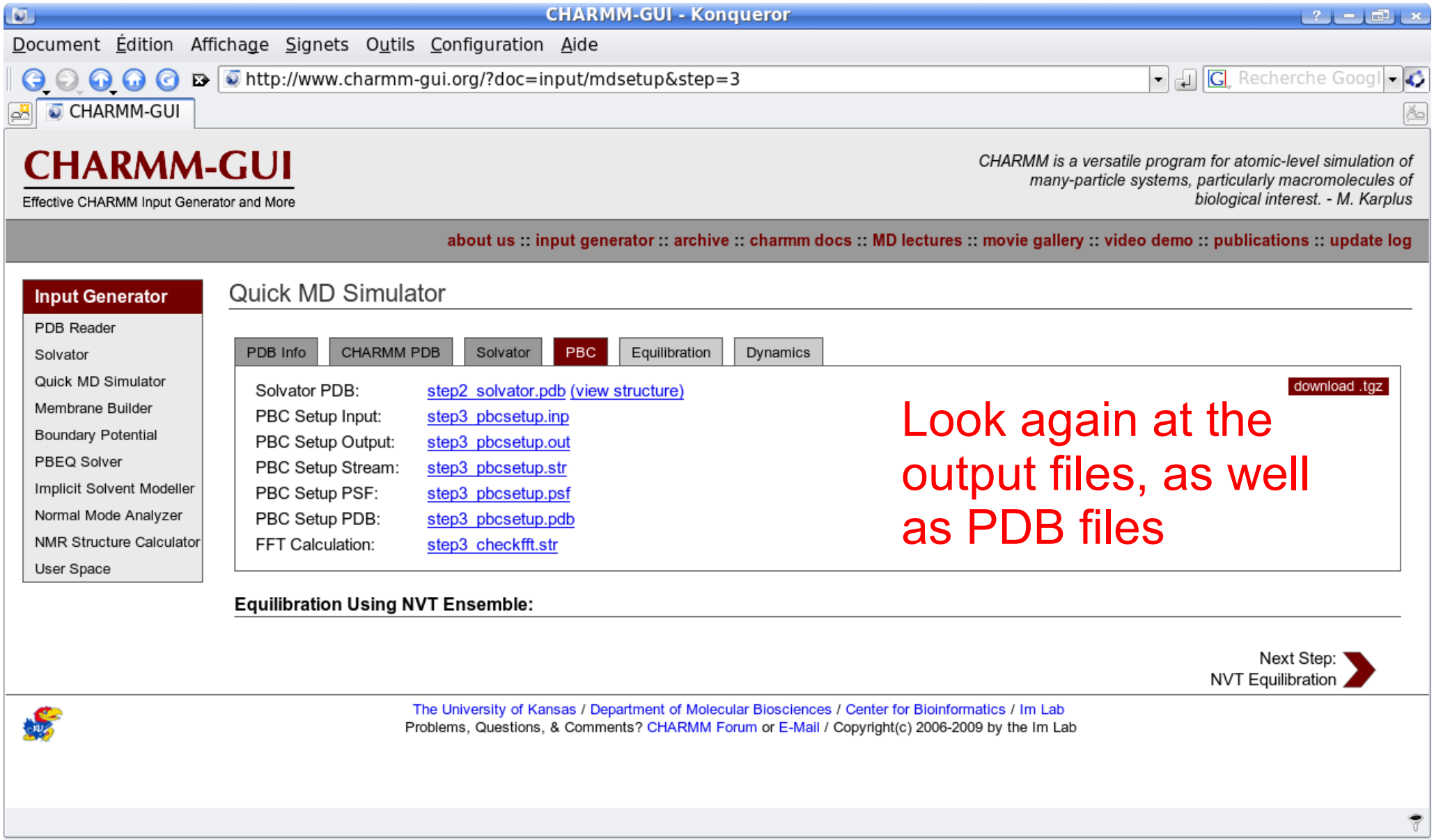


Look again at the output files, as well as PDB files

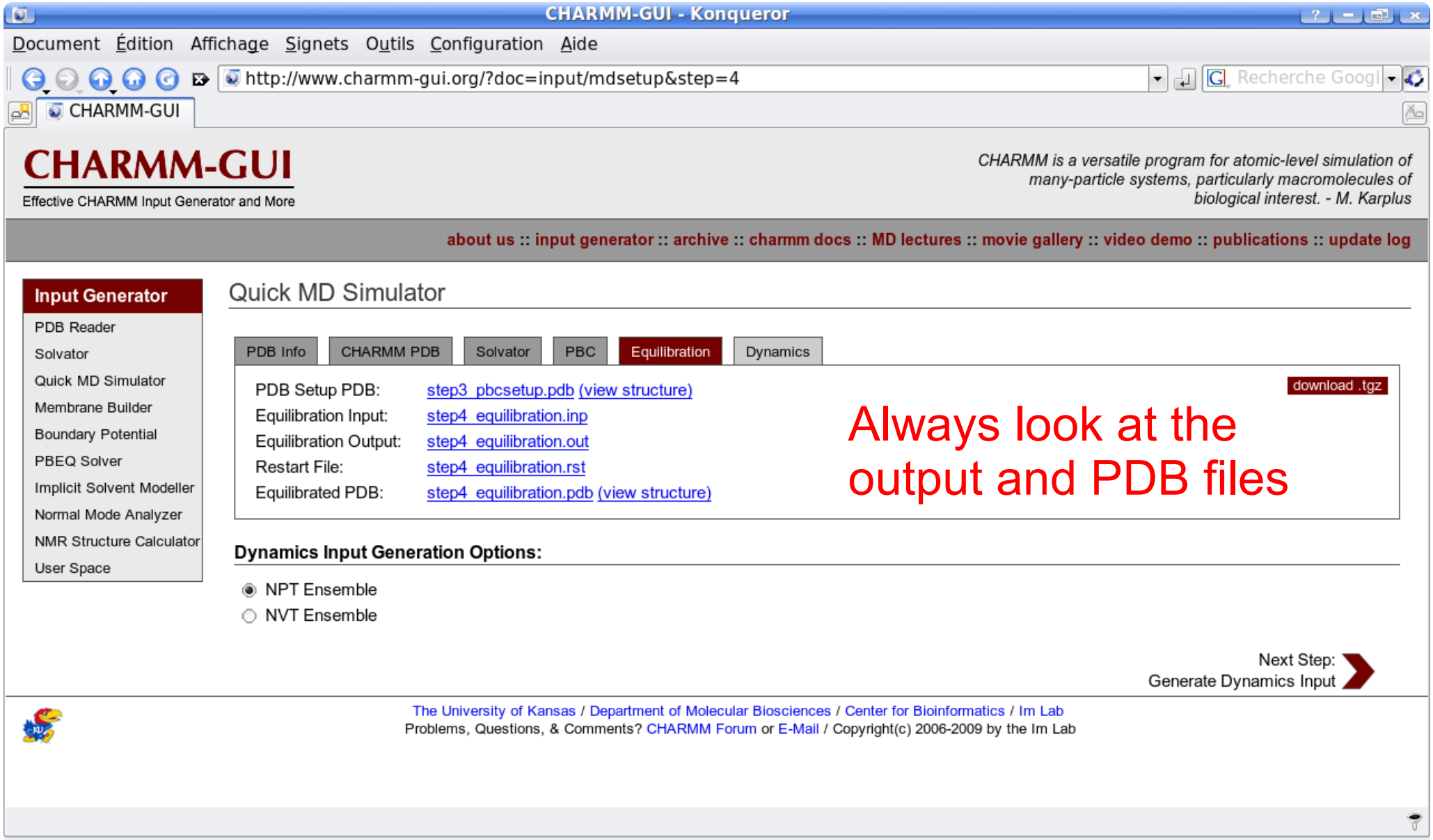
Getting ready for minimisation



Minimised system



System's equilibration



Good chances that the equilibration was not enough!

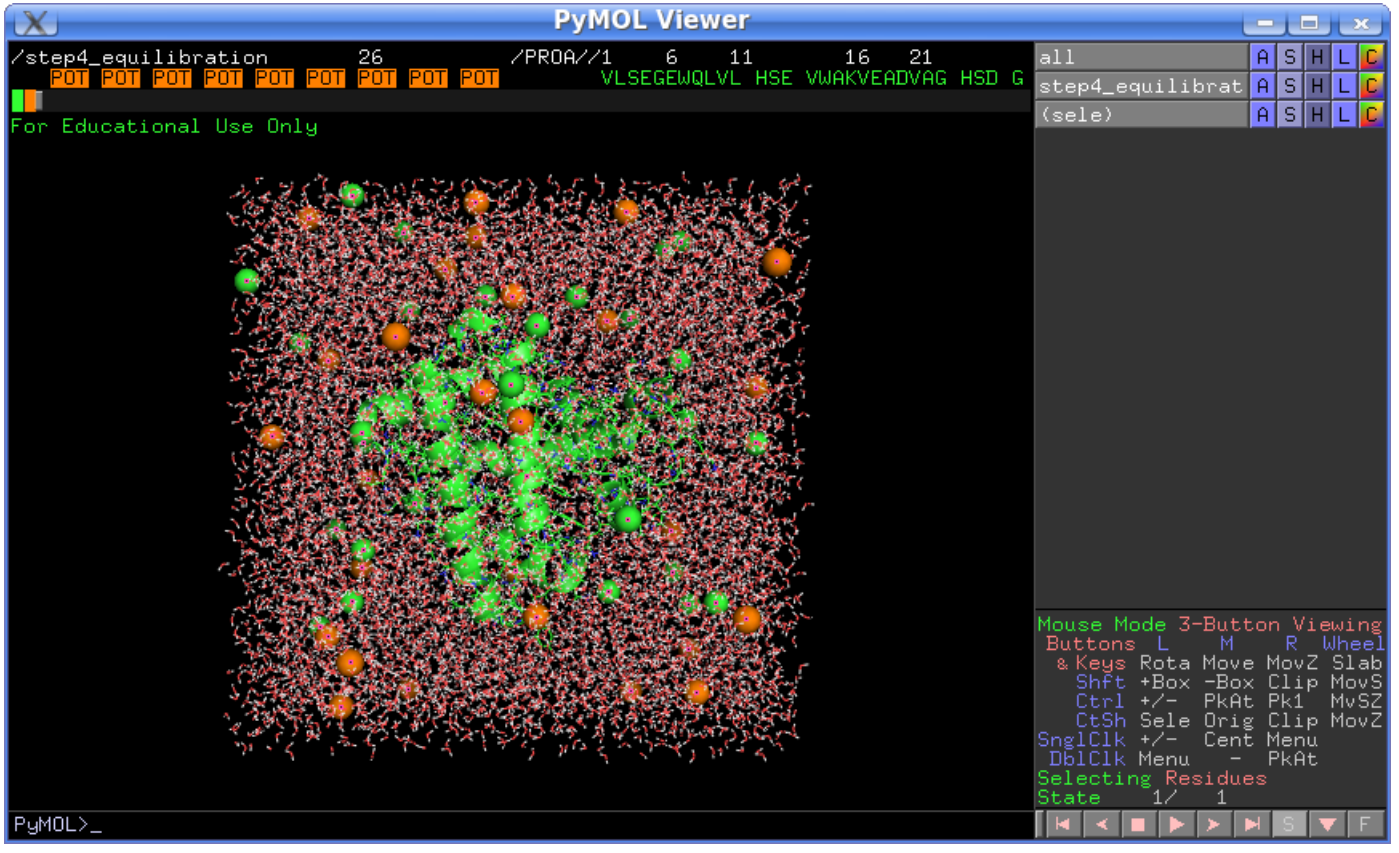
Get input files for MD simulation



Final system without heme

In a terminal session:

% tar zxvf charmm-gui.tgz (extract archive)
% mv charmm-gui.tgz charmm-gui-noheme.tgz
% mv charmm-gui charmm-gui-noheme
% cd charmm-gui-noheme
% pymol step4_equilibration.pdb



Status

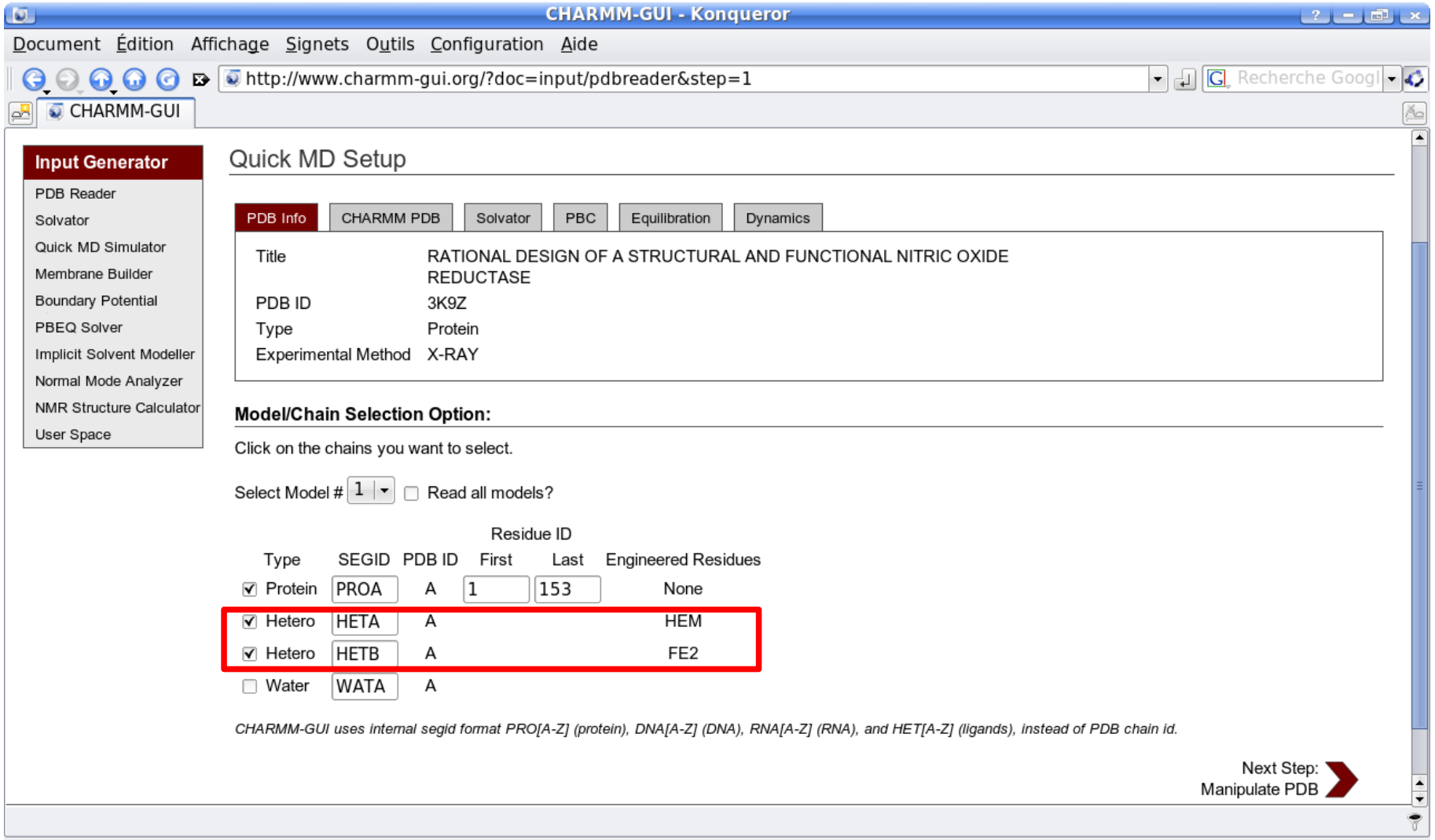
- At this point, you have everything to run a MD simulation for the built system (no heme...)
- 2 files:
  - step5.1_production.inp to start the simulation
  - step5.2_production.inp to continue the simulation
- Parameters you might want to adjust:
  - time 0.002 (timestep of 2fs)
  - nstep 100 (trajectory of only 2fsx100=200fs)
  - nsavc 500 (frequency to save coordinates for analysis)
- To run charmm:

charmm < step5.1_production.inp > step5.1_production.out

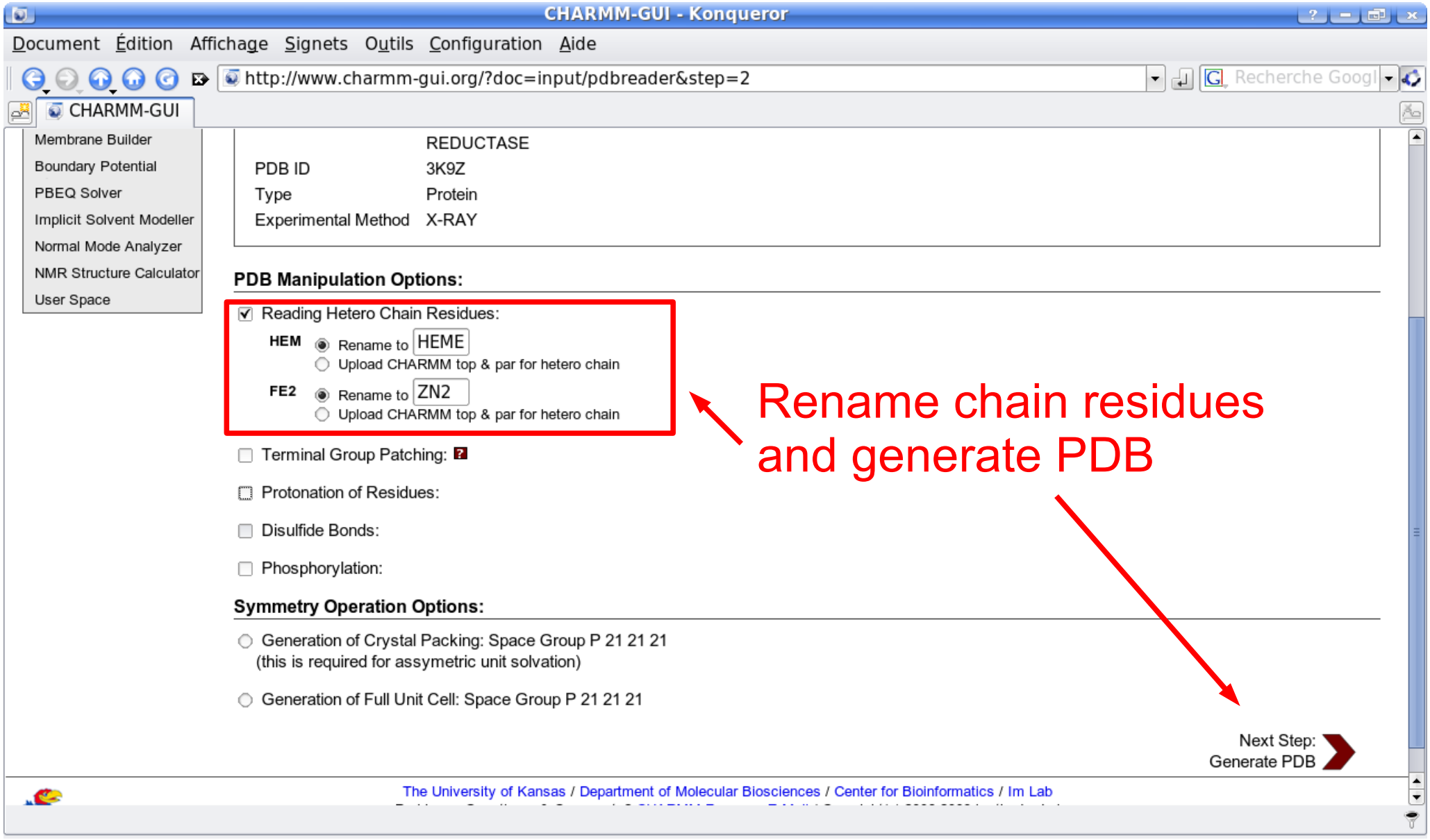
charmm < step5.2_production.inp > step5.2_production.out

Next step: including Heme and ZN(II)

Start the process from the beginning, but include the Heme and the Fe2 molecules:

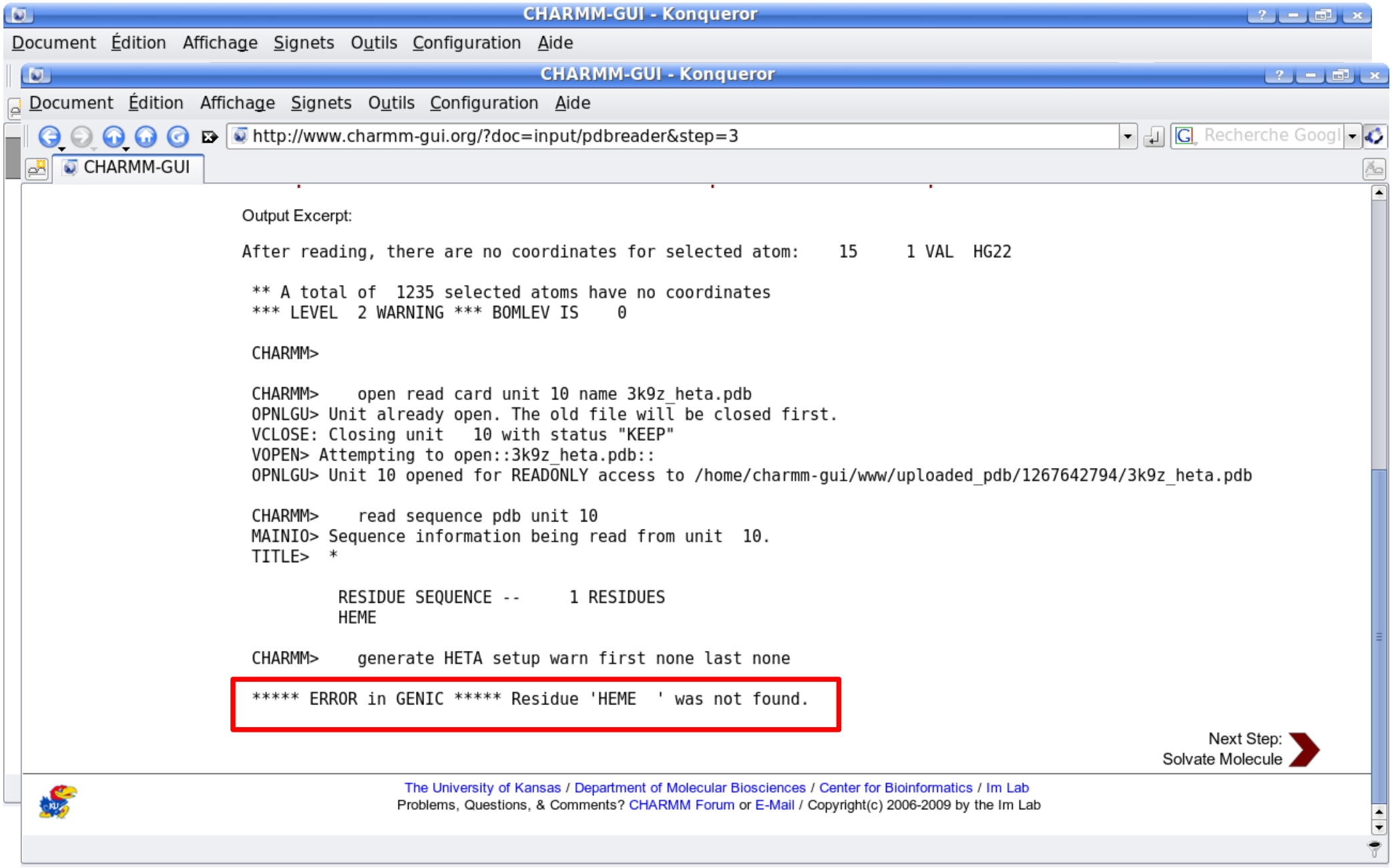


Rename chain residues

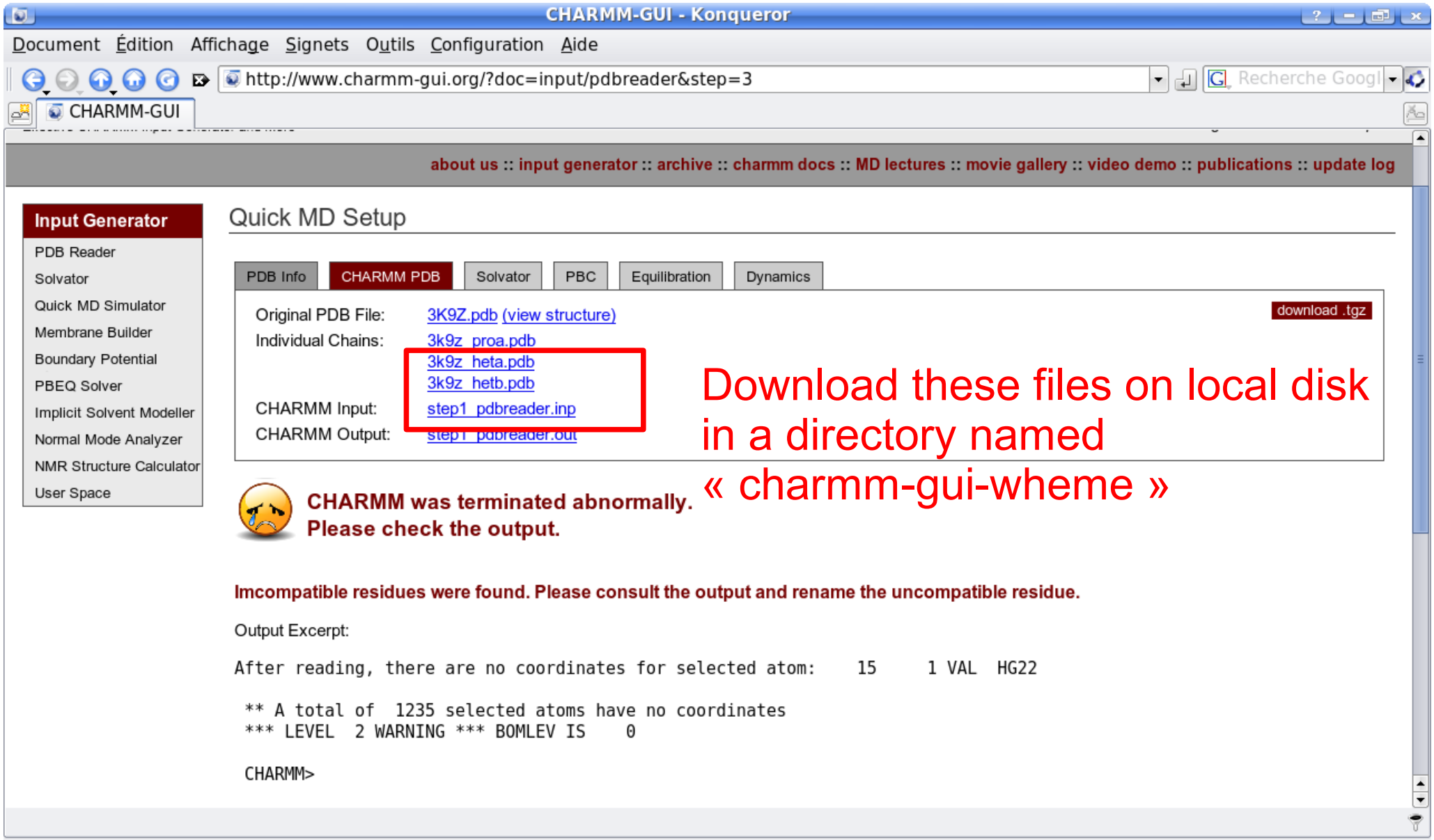


Rename chain residues and generate PDB

Generate PDB failed



Generate PDB failed



Manual STEP1 – get toppar files

Get the current topology and parameter files form Alex Mackerell's web page (file toppar_c35b2_c36a2.tgz)

Save the file in the workshop directory, and untar the file:

```
% tar zxvf toppar_c35b2_c36a2.tgz
```

In the charmm-gui-wheme directory:

- Copy the topology and parameter files from charmm toppar directory

```
% cp ../toppar/top_all27_prot_na.rtf .
% cp ../toppar/par_all27_prot_na.prm .
```

- Copy the topology and parameters for the heme:

```
% cp ../stream/toppar_all22_prot_heme.str .
```

Manual STEP1 – create stream file

Create the file toppar.str, that contains the following lines:

```
* Topology and Parameter Stream File
*

! Read topology and parameter files
open read card unit 10 name top_all27_prot_na.rtf
read  rtf card unit 10

open read card unit 20 name par_all27_prot_na.prm
read para card unit 20

stream toppar_all22_prot_heme.str

return
```

Manual STEP1 – edit input file

Edit the file step1_pdbreader.inp:

Replace the following lines:

```
open read card unit 10 name top_all27_prot_na.rtf
read  rtf card unit 10

open read card unit 20 name par_all27_prot_na.prm
read para card unit 20
```

With this line:

```
stream toppar.str
```

STEP 1: edit input file

Add the following lines (in red):

```
open read card unit 10 name 3k9z_heta.pdb
read coor pdb unit 10 resid

! Bind heme to protein
PATCH PHEM PROA 93 HETA 155
…
open read card unit 10 name 3k9z_hetb.pdb
read coor pdb unit 10 resid

! join segments
join PROA HETA renumber
join PROA HETB renumber
```

Manual STEP1 – edit pdb file

Edit the file 3k9z_hetb.pdb:

ATOM 1264 FE ZN2 A 157 18.065 18.840 11.081 1.00 36.74 HETBFE

for

ATOM 1264 ZN ZN2 A 157 18.065 18.840 11.081 1.00 36.74 HETBFE

Copy the file 3k9z_proa.pdb, from the charmm-gui-noheme directory, to the working directory:

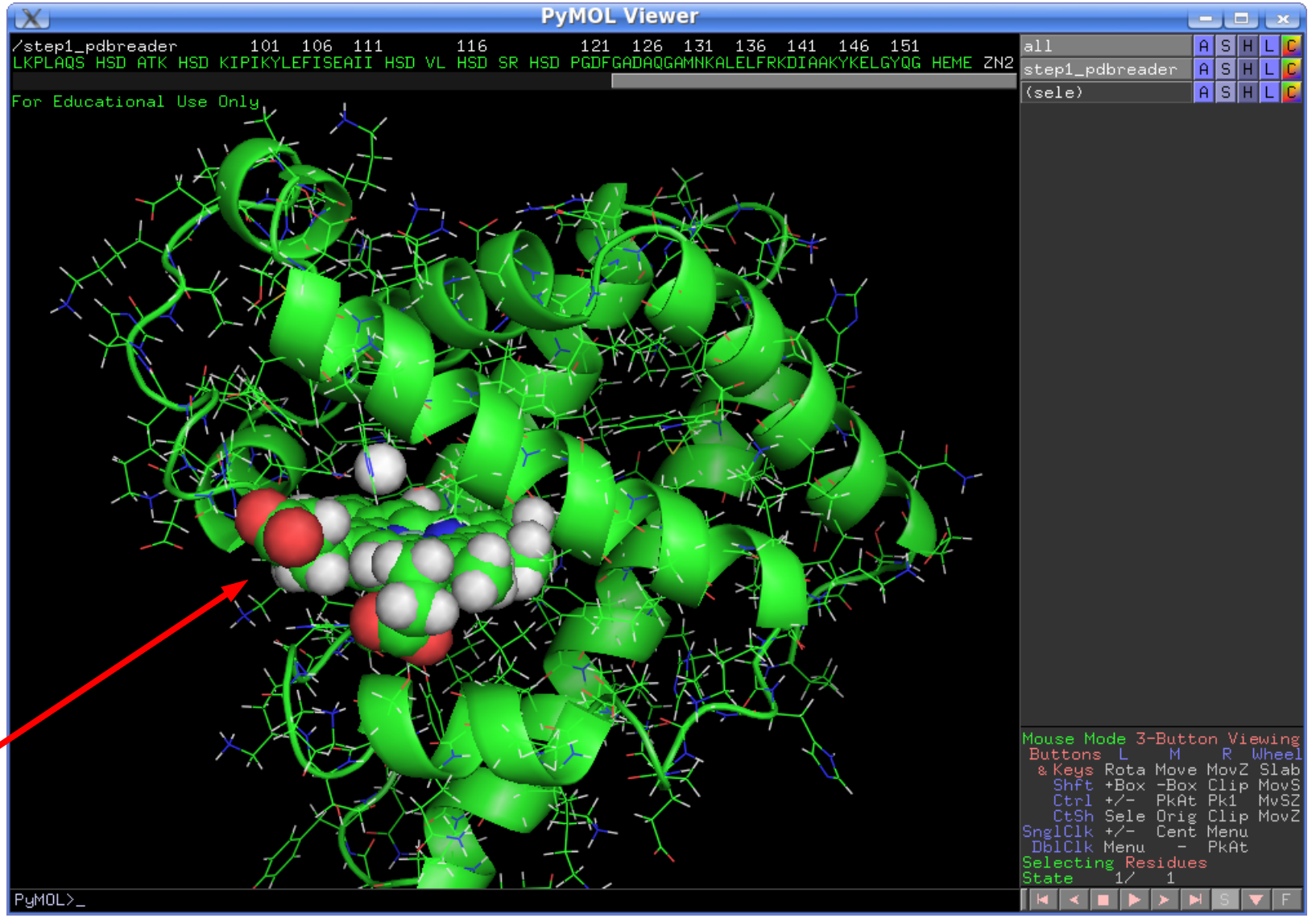
% cp ../charmm-gui-noheme/3k9z_proa.pdb .

and run charmm:

charmm < step1_pdbreader.inp > step1_pdbreader.out

Check the output and the PDB files created.

STEP 1: final result (PDB file)



Heme
+
ZN

Step 2: Solvation

Copy the following files, from the previous charmm-gui directory, to the working directory:

- All the files beginning with step2*

Edit the file step2_solvator.inp, replace the following lines:

```
open read card unit 10 name top_all27_prot_na.rtf
read  rtf card unit 10

open read card unit 20 name par_all27_prot_na.prm
read para card unit 20
```

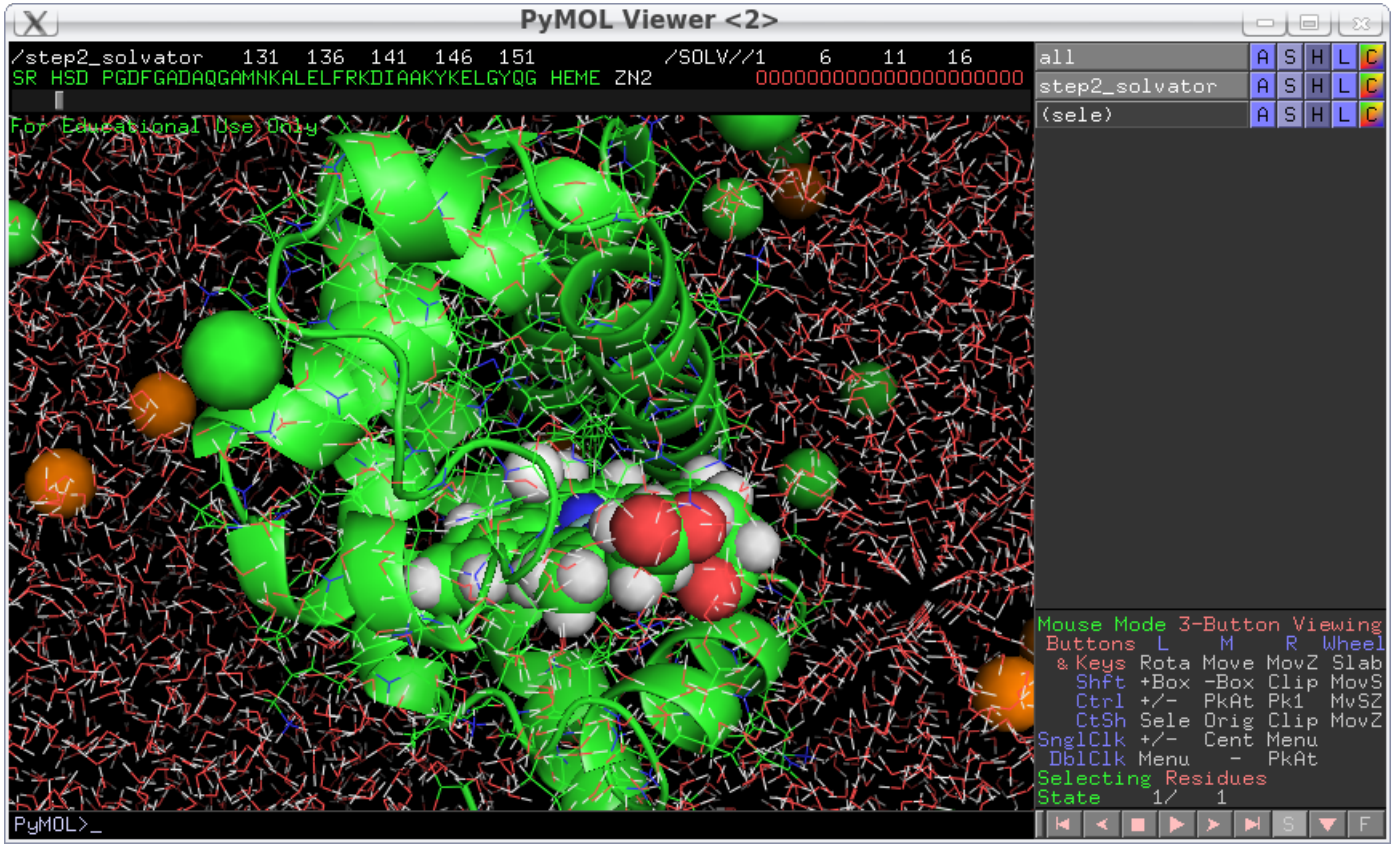
With this line:

```
stream toppar.str
```

STEP 2 : SOLVATION

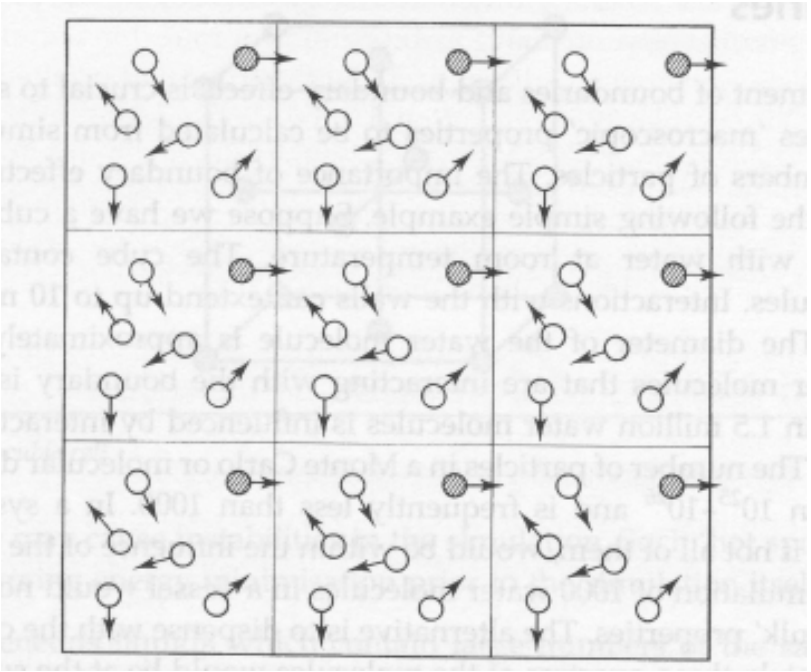
Run charmm using the input file step2_solvator.inp (the water box with ions are recycled from the first run).

Verify the final pdb file (step2_solvator.pdb):



Periodic boundary conditions

- Necessary to avoid frontiers effects
- Use the actual system to mimics molecules at the frontiers:



STEP 3: PBC Setup

Copy the following files, from the previous charmm-gui directory, to the working directory:

- All the files beginning with step3* and the file crystal_image.str

Edit the file step3-pbcsetup.inp, replace the following lines:

```
open read card unit 10 name top_all27_prot_na.rtf
read  rtf card unit 10

open read card unit 20 name par_all27_prot_na.prm
read para card unit 20
```

With this line:

```
stream toppar.str
```

STEPS 3 AND 4

Run charmm using the input file step3_pbcsetup.inp

Verify the final output file for Normal termination.

Finally:

Repeat the process for STEP4: Equilibration

The system is now ready for the production of the trajectory if you modify step5 input files.

Production and analysis

- GLOBAL RMSD over time
- Volume, pressure, temperature over time
- Is the heme still bound to HIS93
- What happen to the mutated residues?
- Comparison with WT-myoglobin crystal

Time for your Protein?

Documentation: charmm.org

- Examples of input files
  (building special systems, analysis, etc)
- Forum to ask questions, and to read answers
  already answered questions
- Getting latest developments of charmm