Assignment A02: Geometry Definition: File Formats, Redundant Coordinates, PES Scans

In Assignments A00 and A01, you familiarized yourself with *GaussView* and *G09W*, you learned the basics about input (GJF) and output files (LOG, CHK), and you learned how to setup simple GJF files with *GaussView* and from scratch. In Assignment A02, you will learn about various file formats conveniently used for crystal structures, you will learn how to extract a pertinent system (either a molecule or a molecular aggregate) from a crystal structure file, and you will perform entry-level ab initio computations on this model system. We will focus on two crystal structure formats: CIF and PDB files. You will study a molecular aggregate found in a crystal structure of a "small molecule" CIF file, and you will study an active site transition metal complex found in a "macromolecular" CIF file (a.k.a., mmCIF) or a PDB file of an enzyme. *GaussView* can read PDB files (small molecule and macromolecular) and small molecule CIF files, but it cannot read macromolecular mmCIF files.

You can study a **molecular aggregate** found in a "small molecule" CIF file linked to the assignment page (A02_FF.cif, ASCII, hydrated diphenylalanine zwitterion). Tasks will be described for this specific example. Feel free to study a dimeric molecular aggregate found in a "small molecule" CIF file of your choice and perform the analogous tasks.

You can study an **active site transition metal complex** starting from the crystal structure **5CJH** reported on the PDB website which is linked to the assignment page (Crystal Structure of Eukaryotic Oxoiron MagKatG2 at pH 8.5). All tasks will be described for this specific example. Feel free to study an active site transition metal complex found in a "macromolecular" CIF file or a PDB file of a protein of your choice and perform the analogous tasks. To do so, your protein must contain an oxoiron-heme system coordinated by a protein side-chain ligand.

The specific tasks include (a) the construction of a model system from a "small molecule" CIF file, (b) the construction of a model system from a "macromolecular" CIF file or from a PDB files, (c) the optimization and frequency computation at the RHF/6-31G* level of dimeric molecular aggregate starting from the crystal structure, and (d) the exploration of the RHF/6-

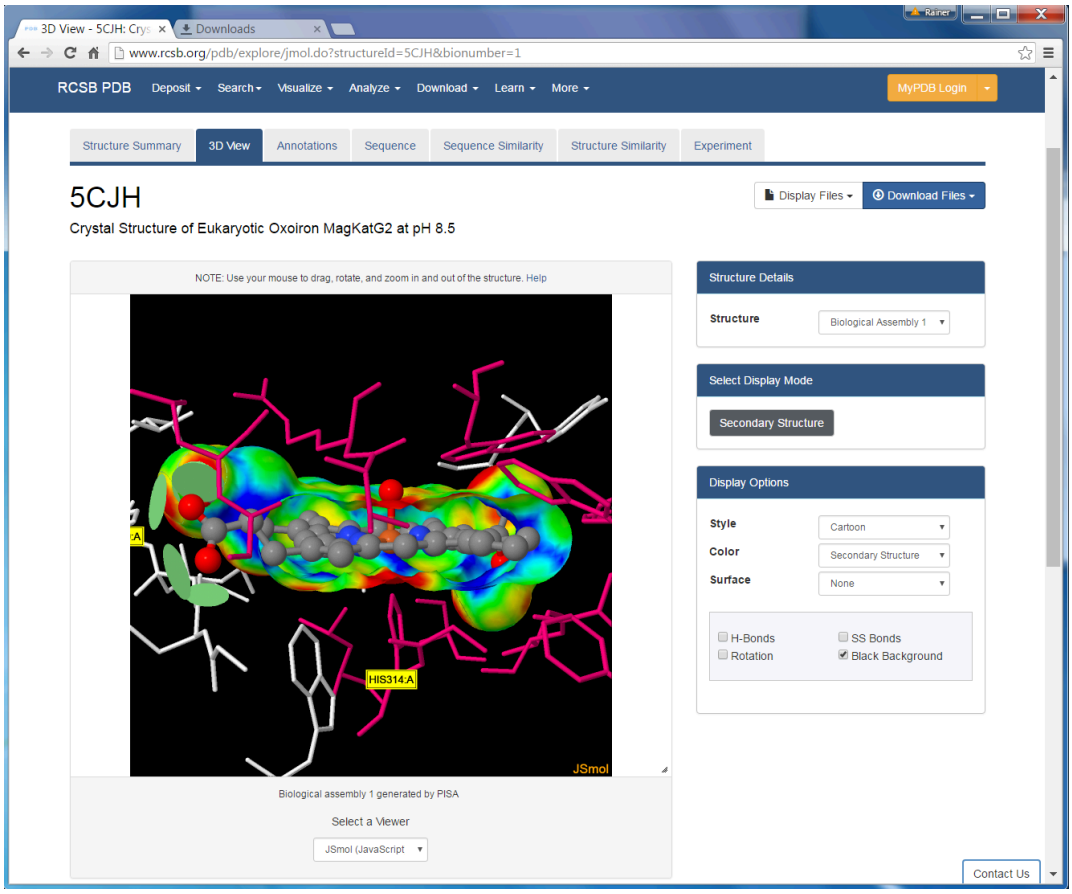
31G* potential energy surface of an active site transition metal complex starting from an enzyme crystal structure using relaxed scanning techniques.

(a) Construct Model System from "Small Molecule" CIF File. Obtain the CIF file (as ASCII file). Your **molecular aggregate** must contain two molecules and only one molecule can be water. Select an FF-water aggregate or an FF-dimer as you see fit. Select an aggregate that can reasonable be expected to maintain its integrity in isolation. Delete all other atoms, and generate a GJF file that contains the crystal structure of your aggregate. Use this GJF file in part (c). This task can be performed with *Gaussview*. But you might invest some time to edit the crystal structure with *Mercury*.

(b) Construct Model System from "Macromolecular" CIF File or PDB File. On the structure summary page of your protein structure (i.e., http://www.rcsb.org/pdb/explore/explore.do?structureId=5CJH), note the menus "Download Files" and "Display Files". You can download the structure in various formats including PDB, PDBc/mmCIF, and CIF; do that. Note that the extension of an mmCIF file also is CIF; that may cause some confusion. The file 5cjh.pdb is an ASCII file: you can open the file with *GaussView* and you can view the file content through "View File" in the "Results" menu, or you can open the file with a word processer (i.e., *WordPad*) to view its content. The files 5cjh.cif and 5cjh-sf.cif are ASCII files: you cannot open the files with *GaussView* but you can open the files with a word processer (i.e., *WordPad*) to view their content.

Your **active site transition metal complex** must contain an oxoiron-heme system coordinated by a protein side-chain ligand. Find that system in the crystal structure **5CJH**. On the structure summary page of your protein structure (i.e., http://www.rcsb.org/pdb/explore/explore.do?structureId=5CJH), select "3D View" and display the heme pocket (see image on following page). Keep the oxoiron-heme system and keep the side-chain ligand L, that is the imidazole of His314. Delete all other atoms, and generate a GJF file that contains the crystal structure of your

model system. This can be done in various ways, and the process will take some time. Use this GJF file in part (d).



About the **Hydrogen Count, Charge and Multiplicity** of the Model: It is important to note that graphical user interfaces (GUI) come with different degrees of sophistication when it comes to adding hydrogen atoms to crystal structure data. Crystal structure data frequently do not contain information about the positions of hydrogen atoms. In such cases the GUI will add these hydrogen and it is not always done right. Hence, you must think about this issue, you need to know what the correct structure should look like (you may learn that from reading the paper in which the crystal structure is discussed), and you may need to correct the GUI-suggested model if the GUI makes this necessary. For example, the bridges between the five-membered rings in

porphine are CH units (containing sp^2-C), not CH_2 units (containing sp^3-C). Take a look at the structure of porphine (google it), the parent molecule of all hemes. Note that neutral porphine contains two five-membered rings with "=N–" and two five-membered rings with "–NH–". The ligand in heme is the dianion generated by deprotonation of the two five-membered rings that contains "–NH–" units. The heme frequently has side chains attached such as methyl groups, vinyl groups, alkyl groups containing carboxylates, and so on. For the purpose of this exercise, you can replace some or all of those side chains with hydrogen atoms.

Sometimes the hydrogen count is correct in the GUI but the hydrogens are at the wrong places. For example, the Fe-coordinated imidazole ring contains one "=N–" unit and one "–NH–" unit, and it is the "=N–" unit that is used to ligate the iron. *GaussView* has produced models in which the "–NH–" unit points to the metal instead of the "=N–" unit.

Once you have the structure you want, think carefully about the charge and the multiplicity. This might not be trivial for a transition metal complex in the active site of an enzyme. For the **5CJH** system, consider iron to be 3+, the oxygen to be neutral, and the porphine ligand is a dianion (see above). Hence, the "FeO(Por)" unit has an overall charge of +1. The imidazole ligand is neutral. Fe(3+) is a d^5-metal and can have 1, 3 or 5 unpaired electrons. Use the lowest multiplicity (unless you have reasons to use higher spin states). Hence, the basic system will have a charge of +1 and multiplicity of 2. If you left any carboxylate-containing side chains in the large ligand, then you must decide whether you want to compute them as carboxylate anions or as neutral carboxylic acids and adjust the overall charge of the system accordingly.

(c) Optimizations and Frequency Computation at RHF/6-31G* Level of Molecular Aggregate Starting From the Crystal Structure. In part (a), you have generated a GJF file that contains the molecular aggregate with the geometry found in the crystal structure. Edit the GJF file so that it contains the correct commands (opt freq RHF/6-31G*), will run on two processors with 256 MW of memory, and will create a suitably named CHK file. You can edit

with *GaussView* ("Calculate" menu) and submit to *G09W* from there. Or you can edit the GJF file with a word processor and submit the final GJF file directly to *G09W*.

(d) RHF/6-31G* Level Exploration of an Active Site Transition Metal Complex Starting From an Enzyme Crystal Structure. In part (b), you have generated a GJF file that contains your active site transition metal complex with the geometry found in the crystal structure. Starting with that GJF file, you now need to generate GJF files for **three tasks**: As a warm-up, optimize the structure RHF/6-31G* level and compute vibrational frequencies. Starting with that optimized structure, you should then generate GJF files for two the following two scan tasks: To scan the potential energy surface (PES) as a function of the Fe-O bond and to scan the PES as a function of the Fe-L distance. Both tasks should be performed at the RHF/6-31G* level, on at least two processors, and with at least 128 MW of memory. Edit with *GaussView* ("Calculate" menu) and select "Scan" as "Job Type". You now need to add a scan coordinate using the "Redundant Coordinate Editor". Select "Redundant Coordinates" in the "Edit" menu. Click "Add" to add a coordinate, select "Bond" to select a bond coordinate, and input the numbers of atoms Fe and O (or N of the imidazole L) by clicking in those atoms in the active window with the cursor in the appropriate field. The select "Scan Coordinate" and take *n* steps of size *m* Angstroms. Make reasonable choices for (*n*,*m*); small (long) steps for strong (weak) bonds. Write the GJF file; note the additions of "opt=modredundant" to the command line and of a line at the end containing scanning information. Run the scans. The LOG file of each task will contain the optimized structures at every step and you can display the total energies as a function of the scan coordinate using "Scan…" under "Results". Click on the "Scan of Total Energy" graph to see options to print, save, export…

(e) Write-Up. Submit one Word file "A02_'your_last_name(s)'.docx". The file must contain three Figures (see below); each Figure with its legend on a separate page. Use page breaks.

Generation of images of molecular models: Display molecule with "Ball & Bond Type", scale radii to 65%, generate image as TIF files with "Save Image File…" in the "Edit" menu (enlarge 2x, "White Background"). Crop TIF files. Insert image into Word file as "enhanced metafile".

In **Figure 1**, show an image of your molecular aggregate in the crystal structure on the left and of the optimized structure of your molecular aggregate on the right. Use the Table feature of Word (a two cell table). Write a figure legend.

In **Figure 2**, show an image of your active site transition metal complex in the crystal structure on the left and an image of the optimized structure of your active site transition metal complex on the right. Use the Table feature of Word (a two cell table). Write a figure legend.

In **Figure 3**, show plots of relative energies (kcal/mol, relative to optimized minimum) as a function of the scanned internal coordinate of your active site transition metal complex. Fe–O scan on the left, Fe–L scan on the right. Use the Table feature of Word (a two cell table). Use Excel to make "marked scatter" plots; use minor and major tick intervals, label axes. Insert XL graphs into Word file as "enhanced metafile" or as "PDF". Write a figure legend.

Submission & Deadlines: Submit "A02_'your_last_name(s)'.docx" as attachment to email on Tuesday, 09/20/16 by midnight. Bring one (stapled) hardcopy to class on Wednesday, 09/21/16, for evaluation by peer review.

Global: All Word files must contain the author name(s) in the header. Pages numbered, bottom center. Times New Roman, font size 12 pt, double-spaced (24 pt). 1-Inch margin on all sides.
Global: Color printing is accessible to all graduate and undergraduate students associated with research groups. Undergraduate students without access to free color printing may submit an additional PDF file for printing by DoC staff. Only PDF files will be printed. Submit this PDF file along with the electronic submission and request "please print PDF" in the email.