

HADDOCK LABORATORY EXERCISE

BCMB/CHEM 8190, 2014

Introduction.

Characterizing the structural aspects of how ligands bind to proteins and how proteins assemble in multi-protein complexes or protein-DNA/RNA complexes is important to drug discovery as well as the understanding of basic biological processes. NMR can make a contribution here, but it is more often in the form of sparse pieces of data (identification of ligand epitopes from STD or protein chemical shift perturbation on addition of ligand) than complete NOE/RDC based structure determination. Docking programs often can use this type of data to facilitate the generation of molecular models. We attempt to illustrate the use of one such program in the following exercise.

HADDOCK (High Ambiguity Driven biomolecular DOCKing) is one approach that makes use of biochemical and/or biophysical interaction data such as chemical shift perturbation data resulting from NMR titration experiments, mutagenesis data or bioinformatic predictions. It uses several modules originally introduced in the programs ARIA and CNS. Much of the data used is introduced as Ambiguous Interaction Restraints (AIRs) to drive the docking process. An AIR is defined as an ambiguous distance between all residues shown to be involved in the interaction. It can also make use of more quantitative restraints between pairs of nuclei as one obtains from NOEs and RDCs. Since the original 2003 JACS publication, HADDOCK has been extended to deal with a large variety of data and complexes. In addition to protein-protein docking, HADDOCK has been applied to the modeling of protein-DNA, protein-RNA, protein-oligosaccharides and other protein-ligand complexes.

The program is available as both through a webserver and as a downloadable program for local installation. Information can be found at: <http://www.nmr.chem.uu.nl/haddock/>. Descriptive articles include: Cyril Dominguez, Rolf Boelens and Alexandre M.J.J. Bonvin (2003). "HADDOCK: a protein-protein docking approach based on biochemical and/or biophysical information." J. Am. Chem. Soc. 125, 1731-1737 and S.J. de Vries, M. van Dijk and A.M.J.J. Bonvin. "The HADDOCK web server for data-driven biomolecular docking." Nature Protocols, 5, 883-897 (2010).

Laboratory Exercise:

HADDOCK has been installed on the computers you are using (or is accessible by opening a remote shell for a server having HADDOCK installed). The instructions that follow are intended to supplement the manual available on the HADDOCK website by leading you through an exercise that demonstrates docking a protein-oligosaccharide complex. The protein is the carbohydrate recognition domain of the mammalian lectin DC-SIGN (approximately 20 kDA) and the ligand is a trisaccharide composed of N-acetyl-glucosamine, fucose and galactose residues, Lewis^X. The AIR restraints used are from chemical shift perturbations for the protein and STD data for the ligand. You can view these in the file: `home/your_user_name/haddock/DC-SIGN-Lex-air_new.tbl`. Distance constraints come from NOEs and conserved fucose to calcium ion contacts found in other DC-SIGN complexes. You can view these in the file: `/home/your_user_name/haddock/DC-SIGN-Lex-noe_long_Ca.tbl`. These files are identified as program input in Step 1 below. The program also allows the designation of certain protein residues as flexible and semi-flexible, with the designation differing primarily in whether residues are flexible throughout steps of docking and refinement, or only in the later refinement steps. You will designate these residues in Step 2 below. The program then proceeds in three steps involving rigid docking, simulated annealing, and water refinement, as described in Step 3. The result of the exercise is to produce several models showing how Lewis^X might be bound in the DC-SIGN binding site.

Step 1: Creating the run directory

- The files for your HADDOCK run will be located in: /home/your_user_name/haddock
- Go to <http://www.nmr.chem.uu.nl/haddock/> Select "Project Setup" on the left-hand menu and click "Start a New Project".
- Edit the path for the current HADDOCK program directory to /home/LMorris/haddock2.1
- Edit the path for the current project directory to /home/your_user_name/haddock
- Leave the run number and number of molecules to their default values for this exercise.
- Edit the paths for the pdb files of the protein (1st molecule) and ligand (2nd molecule).
/home/your_user_name/haddock/DC-SIGN.pdb and /home/your_user_name/haddock/LewisX.pdb
- Delete the file list for the second molecule.
- Edit the directory path for the AIR restraint file. /home/your_user_name/haddock/DC-SIGN-Lex-air.tbl
- Edit the directory path for the distance (NOE) constraint file.
/home/your_user_name/haddock/DC-SIGN-Lex-noe.tbl
- Click "Save updated parameters" and then move the saved "new.html" file into the project directory you created
- In a terminal, go to the project directory and execute the command "haddock2.1" to create the run directory.

Step 2: Modifying the run parameters

- In the newly generated "run1" directory locate the "run.cns" file.
- On the same "Project Setup" page as before, now go to the section to "Edit your run.cns file" and click "Browse" to select your run.cns file, then choose "Edit". (We recommend leaving this web browser open once you are finished editing the file, in case you need to go back and make additional changes.)
- Choose a name for your current project and type it into the first box under the "filenames" section.
- Scroll down to "Definition of semi-flexible interface" and for molecule A (the protein) change the number of segments to "2" defined as residues 363-367 and residues 356-359. For molecule B (the ligand) define the number of segments as "1" as residues 3-3.
- Scroll down to "Definition of fully flexible segments" and for molecule A (the protein) change the number of segments to 3 and define them as residues 344-356, 311-315, and 368-374.
- Scroll down to "topology and parameter files" and edit names for the following files:
 - topology file for molecule A - "topallhdg5.4.pro"
 - topology file for molecule B - "carbohydrate.top"
 - linkage file for molecule A - "topallhdg5.4.pep"
 - linkage file for molecule B - "carbohydrate.link"
 - energy parameter file for molecule A - "parallhdg5.4.pro"
 - energy parameter file for molecule B - "carbohydrate.param"
- Scroll down to "Number of structures to dock" and edit the following:
 - number of structures for rigid body docking - "10"
 - number of structures for refinement - "4"
 - number of structures to be analysed - "4"
- Scroll down to "Final explicit solvent refinement" and edit "number of structures for the explicit solvent refinement" to "4"
- Scroll down to "parallel jobs" and edit the path for the cns executable to /home/LMorris/cns_solve_1.3/intel-x86_64bit-linux/bin/cns and the number of jobs to 2

-Click "Save updated file" and copy the new saved run.cns file into your run directory.

Step 3: Running HADDOCK

- Copy the carbohydrate.top, carbohydrate.link, and carbohydrate.param files into the "run1/toppar" directory.

- In the terminal, go to your "run1" directory and type haddock2.1 to run. The progress will update in the terminal as HADDOCK runs. First, it will create the .psf files for each of your molecules and then merge the topologies. This step will fail if you have not copied the carbohydrate topology, linkage, and parameter files into the toppar directory, or not correctly defined them in your run.cns file. Then it will start the initial, rigid body docking (it0); the rigid-body docking is fairly quick. Once it0 is finished, it will move onto the simulated annealing, flexible docking (it1), this step is much slower. At the end of this docking, HADDOCK will perform an analysis on these structures before moving on to the final solvent refinement, this step is also slow. Another analysis will be completed after the solvent refinement. Completing the entire run may take up to 30 min. You can monitor the progress by looking at the run1/structures/it0 and it1 directories to see if pdb files are being generated, and also looking at the ".out" files in the run1 directory, to see if any errors are noted.

Step 4: Looking at your results

HADDOCK outputs several useful files for judging the complexes formed.

- Navigate to the run1/structures/it1/water directory. The "file.list" file will contain a list of the structures from the final refinement and a "score". The list will be sorted from lowest to highest score, where the lower score is better.

- Next go to the analysis directory located within the water directory. Here, you will find the "energies.disp" and "noe.disp" files which contain information about the various energies and distance constraint violations, respectively, for each structure.

- Look at the score, total energy, and distance constraint violations to choose which structure you think is "best".

- Open this structure in chimera and check that the stereochemistry of the sugar has not been changed and that the fucose residue is located closest to the calcium ion.

- Open the pdb file of the crystal structure for the complex in chimera and overlay the two complexes - how well do they agree?