File:hydropro++10_54-PUB-Sept2011.doc

HYDROPRO

Version 10 (compilation 54) , September 2011

Index

1. Introduction to HYDROPRO

HYDROPRO computes the hydrodynamic properties of rigid macromolecules (globular proteins, small nucleic acids, etc) from their atomic-level structure, as specified by the atomic coordinates taken from a PDB file supplied by the user, from which the proper hydrodynamic model is built by the program itself. The HYDROPRO calculation comprises the basic hydrodynamic properties: translational diffusion coefficient, sedimentation coefficient, intrinsic viscosity, and relaxation times, along with the radius of gyration. Optionally, HYDROPRO computes also other solution properties such as the covolume (related to the second virial coefficient) and scattering related properties such as the distribution of intramolecular distances. Beware that we have another similar program – HYDRONMR – for the more specific purpose of calculating quantities relevant in NMR relaxation.

2. Literature

The reference that describes the features of this, latest version of HYDROPRO is:

- A. Ortega, D. Amoros and J. García de la Torre, "Prediction of hydrodynamic and other solution properties of rigid proteins from atomic and residue-level models", *Biophysical Journal* 101, 892-898 (2011).

The primary reference for HYDROPRO, covering the description of the program up to the previous version, 7C, which also describes the fundamentals that are still be basis of the new version, was:

- J. Garcia de la Torre, M.L. Huertas and B. Carrasco, “Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. *Biophys. J.* 78, 719-730 (2000)

If you employ scattering related properties (distribution of distances, longest distance, scattering function), or the covolume, then the you may also cite the reference where these calculations are described:

- J. Garcia de la Torre, B. Carrasco and S. E. Harding, “Calculation of NMR relaxation, covolume and scattering-related properties of bead models using the SOLPRO computer program", *Eur. Biophy. J.*, 28, 119-132" (1999).

Further examples of application of HYDROPRO are available for proteins:

- J. García de la Torre, ”Hydration from hydrodynamics. General considerations and applications of bead modelling to globular proteins” *Biophys. Chem.* 93, 159-170 (2001)

and for small nucleic acids:

- M.X. Fernandes, A. Ortega, M.C. López Martínez y J. García de la Torre. “Calculation of hydrodynamic properties of small nucleic acids from their atomic structures” . *Nucleic Acids Research*, 30, 1782-1788 (2002).

If you wish to cite also the theoretical work on which the bead modeling procedure is based, a proper cite is our 1981 review in *Quarterly Reviews of Biophysics*. In our 1999 paper in *Biophysical Journal*, you can find an update of the theory, and a discussion on bead and shell modeling methodologies:

- J. Garcia de la Torre and V.A. Bloomfield, “Hydrodynamic properties of complex, rigid, biological macromolecules. Theory and applications”. *Q. Rev. Biophy.*, 14, 81-139 (1981)

- B. Carrasco and J. Garcia de la Torre, “Hydrodynamic properties of rigid particles. Comparison of different modeling and computational strategies”. *Biophys. J.* 76, 3044-3057 (1999).

3. Running HYDROPRO. Input data files

As in all the HYDRO family of programs, you will have to supply two input data files: (a) the main input data file, which will specify primary data such as temperature, solvent density, etc.; and (b) a structural data file, which will contain the information about the structure or geometry of the macromolecule or particle that you are considering. In the case of HYDROPRO, the structural file is just the PDB file containing the atomic coordinates. The name of the structural file will be one of the data in the main input

data file.The name of the main input data file for HYDROPRO will be `hydropro.dat.`

An unlimited number of cases can be executed in a single run of HYDROPRO. Then `hydropro.dat` will contain a series of blocks of data, one for each case. Each block will contain the following lines:

3.a. First part of a block of data.

Basically this part provides the information on the structure of the macromolecule. It contains the following lines (the FORTRAN types are specified):

- TITLE(`CHARACTER*20`) Title of the calculation

- FILENAME (`CHARACTER*30`)Name to be used for the various output files corresponding to each subcase in a many-cases execution. The various files produced will have names of the form `filename.xxx`, where xxx is an extension depending on the file type (see section 4 below). This name would eventually include the path for the files.

- INPUT (`CHARACTER*30`) Name of the PDB file with the atomic coordinates, eventually including its path.

- INDMODE : Type of calculation mode. Depends on type of primary model, either atomic-level or residue-level, and in the form of hydrodynamic calculation: shell model with up to 2000 minibeads, or bead model with one bead per residue. Possible values are
  1 -- Atomic-level primary model, shell calculation
  2 -- Residue-level primary model, shell calculation
  4 -- Residue-level primary model, bead calculation

NOTES

(a) INDMODE 1 is the (only) one that was implemented in the previous version of HYDROPRO

(b) INDMODE 3 (Atomic-level primary model, bead calculation) is still not active. Indeed, it is expected to be less advantageous than any of the other options

- AER(`REAL`) Is the value of the hydrodynamic radius, in Å, of the elements in the primary model. Although in this version this is a user-provided value, in our papers we have described the values of AER that must be used in each mode. We strongly urge you to adhere to these values (with the only exception mentioned in Section 5, HYDROPRO with lower resolution, SAXS, etc structures).

  1 -- Atomic-level primary model, shell calculation AER = 2.9 Å

2 -- Residue-level primary model, shell calculation AER = 4.8 Å
4 -- Residue-level primary model, bead calculation AER = 6.1 Å

In INDMODE 1 and 2, HYDROPRO employs the shell-model methodology. The primary hydrodynamic model is not used in the hydrodynamic calculations. Instead, a shell model , composed of 'minibeads' of radius sigma is derived from it, and extrapolation to the shell model limit of sigma=0 is carried out. For this purpose, the following information has to be supplied next in the main input data file

- NSIG (INTEGER) Number of values of the radius of the minibead. It must be greater than 2 (typically 5 to 8). The radius will range from SIGMIN to SIGMAX. There is also the possibility of letting of the program the task of estimating the two extreme values of sigma; this will be indicated giving the value –1 for NSIG, and in this case you will omit the two following lines with the values of SIGMIN and SIGMAX.

.

- SIGMIN (REAL) Lowest value of sigma, the minibead radius

- SIGMAX (REAL) Highest value of sigma, the minibead radius

The smaller the size of the minibeads, the larger the number of them needed to cover the surface of the particle. The present version of the program works with a maximum of 2000 minibeads. If SIGMIN is too small, an error message will be obtained and the program will stop. The value of SIGMAX should be taken such that the number of minibeads is not too small, say in the range 200-400.

In INDMODE 3 there is no shell model calculation, and therefore, no data are needed for NSIG, SIGMIN and SIGMAX; the corresponding lines are absent in the input file.

3.b. Second part of a block of data

This part provides information on some basic properties of macromolecule and solvent. It contains the following lines:

- T (REAL) Temperature, centigrade

- ETA (REAL) Solvent viscosity, poises

- RM (REAL) Molecular weight

- VBAR (REAL) Partial specific volume, cm^3/g

- RHO (REAL) Solution (approx. Solvent) density, g/cm^3

3.C. Third part of a a block of data

This part is intended for the optional calculation of some non-hydrodynamic properties. These properties are the scattering form factor (Debye expression), the distribution of intramolecular distances, and the covolume, and are calculated from the coordinates of the spheres.

The data that you have to supply are:

- NQ(`INTEGER`), the number of values of the scattering (angular) variable, q . If you wish to omit the scattering calculation, the value given here should be 0 (Recommended values are multiples of 5 or 10 plus 1, i.e., 11, 51, etc; value no. 1 will be q=0 with S(q)=1). There is the option of an automatic choice of the scattering data, NQ and QMAX, which is activated giving -1 for NQ.

- QMAX (`REAL`), the largest value of q (cm^{-1}), so that the scattering variable will range from 0 to QMAX, This line will be omited if there is no scattering calculation (NQ =0), of if it is done with automatic values (NQ =-1).

- NS(`INTEGER`), the number of intervals for the distribution of distances. The values of the intramolecular distances will be varied between 0 and the longest distance, which is determined by the program. If you wish to omit the calculation of the distribution of distances, the value given here should be 0. Also, there is the possibility of an automatic choice of data for this calculation, which is indicated giving -1 for NS

- RMAX (`REAL`), the maximum length (in cm) for the calculation of the distribution of distances (in cm). This line will be omitted if there is no distribution of distance calculation (NQ =0), of if it is done with automatic values (NQ =-1).

- NTRIALS (`INTEGER`) is the number of trials or MonteCarlo moves in the calculation of the covolume. Set this value to 0 if you wish to omit the covolume calculation. Recall that this calculation is very time-consuming, and must be restricted to models with not too many beads.

The most important quantities related to translational and rotational diffusion are the translational and rotational diffusion coefficients and the rotational relaxation times. The program will give you these quantities. For some special purposes you may also want the full 6x6 diffusion tensor, which contains the 3x3 translational diffusion tensor, rotational diffusion tensor, and the translation-rotation coupling tensor, as well as the center of diffusion.

- IDIF is a flag that indicates (if IDIF is 1) that you wish a detailed report of the diffusivity of the particle, including the full (anisotropic) translational, rotational and coupling tensor, and the position of the hydrodynamic (diffusion) center.

3.D. End of calculation or next case

- Next or final line: If this case/block is the only or the final one, in the next line you will put an asterisk followed by 19 spaces.

4. Output files

Several files are produced at execution time. There will be a set of files for each case included in a single run. All these files will have a common name, given by the filename specified for each case in the input file, and a different extension. The extensions correspond to:

- `.res` Output file containing the mains results, with the name specified in the input file
- `-pri.bea` is a pdb-formatted file representing the primary hydrodynamic model, that can be viewed with RASMOL
- `-pri.vrml` is a VRML-formatted file containing the coordinates of primary hydrodynamic model, that can be viewed with a VRML viewer, like Cortona. See our separate VisualBeads user Guide.

- `.sol` is an ASCII file containing data needed if you wish to run the separate program SOLPRO

Other files are:

- `summary.txt` is a numeric archive containing a line for each case in the calculation.

(a) If there was no distance distribution calculation and no scattering calculation, this file has 13 columns containing (1) first 10 characters of title; (2) translational diffusion coefficient; (3) radius of gyration; (4) volume; (5-9) the five relaxation times; (10) intrinsic viscosity; (11) sedimentation coefficient; (12) longest distance; (13) covolume. (b) If there was either distance distribution

(b) If there was scattering function calculation, *P(h)*, with a number of values NH, with NH not greater than 100, then the above columns will be followed by NH+2 columns containing NH, HMAX, and the values of the distribution function corresponding to the NS intervals (centered at h_i=[i+0.5]HMAX/NH

(c) If there was distance calculation *p(r)*, with a number of values NS, with NS not greater than 101, then the above columns will be followed by NS+2 columns containing NS, RMAX, and the values of the distribution function corresponding to the NS intervals (centered at r_i=[i+0.5]RMAX/NS

- A file intended for the use of the results in the analysis of experimental data using the HYDFIT program (see the User Guide of that program for more details). This file will be named with the same procedure as the summary file, in this case ending with “-fit.txt” (in the above example, mydata-fit.txt).

5. HYDROPRO with lower-resolution (SAXS, etc) models

Some lower-resolution techniques, like SAXS, SANS, etc produce structures that are represented as array of "virtual", "dummy" atoms by the scattering analysis program. For instance, the widely used programs DAMMIN, GASBOR, etc produce such arrays of dummy atoms/residues (beads) in a PDB-formatted file, so that it is feasible to employ HYDROPRO to predict the hydrodynamic properties of such structures. The question is what should be the mode of calculation and the AER parameter.

We advise to proceed as with an atomic-type/shell-model calculation (INDMODE=1), but obviously with a proper AER value. Probably, this value should be such that the beads overlap to some extent – in order to fill possible gaps. HYDROPRO predicts accurately the volume of the array of overlapping beads. Then, we would make a series of attempts in order to match the value of the volume reported by HYDROPRO with the one that the macromolecule should have. Such volume can be estimated from the molecular weight and partial specific volume, which give the anhydrous volume, adding some amount that would represent hydration. That could be done as classically for ellipsoidal models with the degree of hydration of about 0.3 g/g, or adding at least 1.0 Å to the bead radius that match the anhydrous volume, or by any other mean.

6. Hints and notes.

- Temperature is given in centigrade

- We recall the possibility, indicated in section 3, of running an unlimited cases in a single execution, with a single `hydropro.dat` file containing a series of blocks, one for each case

- Computing time is proportional to the third power of the number of beads in the hydrodynamic model. This number in shell model calculations (mode 1 and 2) is always (regardless of the size of the protein) of up to approximately 2000. Therefore, computing time is the same for any structure. In mode 4 the number of hydrodynamic elements will be equal to the number of residues. Calculations in mode 4 will be more time consuming than in mode 2 if there are more than 2000 residues

- Bead model calculation (mode 4) appears as an advantageous choice over the shell model (modes 1 and 2) for small macromolecules. However, please consider the following hint:

- Thanks to essential improvements in programming and computational efficiency in the new version 10, the shell model, which is in our opinion the theoretically best approach, requires only less than 20 seconds in simple personal computers, and even less in multicore workstations, so that if you don't mind such CPU time per execution, modes 1 or 2 would be the preferred choices.

- The memory requirement grows with the square of the number of elements in the hydrodynamic model. Like for computing time, this is fixed – regardless of the

number of atoms or residues- in shell-model calculations, in modes 1 and 2, and there is practically no limit for the number of atoms or residues. However, growing memory requirements for large macromolecules place a limit in bead model calculations so that in mode 4 the number residues will have a practical limit that would depend on the operating system, memory in the computer, etc. For instance, in MS Windows with 4 GB memory, the limit is about 7000 residues

- In atomic-level models, the program will take into account all non-hydrogen atoms contained in the ATOM records of the PDB file, including non-protein atoms (HETATM records) except water oxigens.

- In residue-level model, the program will recognize nucleotides as residues.

- We strongly recommend the use of the AER values suggested above, with the exception of special cases - like when it is adjusted for calculations from SAXS bead models).

- Now, among the results in the output file, you will find a list of the equivalent radii corresponding to the various properties. For a description of the definition and utility of the equivalent radii, see: A. Ortega, J. García de la Torre, "Equivalent radii and ratios of radii from solution properties as indicators of macromolecular shape, conformation and flexibility" *Biomacromolecules* 8, 2464-2475 (2007)

- **IMPORTANT:** Please read the HydroFAQs, frequently asked questions on the HYDRO programs for further information.

7. Release notes

This is the fourth released version of HYDROPRO (the third one was Version 7c, September 2005 and the second one was Version 5b, October 2003). The main changes in this version are:

- The program has been entirely rewritten in Fortran 90, taking advantage of advanced computational features, like dynamic memory allocation. The program is, furthermore, much faster than version 7c. While the previous version took several minutes, computing times in present multi-core personal computers may be less than 20 seconds, depending on your hardware.

- Temperature is given in centigrade degrees (not in Kelvin as in previous versions)

- Slight change in the parameterization of the atomic/shell-model calculation, AER = 2.9 Å

- New modes for residue-level/shell-model and residue-level/bead-model calculations, with optimum values for their AER (4.8 Å and 6.1 Å, respectively)

- The number of atoms or residues is not limited by the program, but just determined by the operating system and amount of memory in the computer.

- "Automatic" feature for data in scattering and distances calculation

- File `summary.txt` contains the results for the distance distribution and the scattering function.

When running in a Windows PC, the MS DOS/Windows executable can be started from Windows, but we advise to open a MS DOS session in a (console window) for program execution.