File:hydropro7c-pub.doc

HYDROPRO

Version 7c, September 2005

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 was 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 and the distribution of distances. Beware that we have another similar program – HYDRONMR – for the more specific purpose of calculating quantities relevant in NMR relaxation.

2. Literature

The primary reference for HYDROPRO is:

- 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,` and it will contain the following lines:

3.a. First part of the input file.

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.

- Effective radius (angstroms) of the atomic elements ***[a], [b]***

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.

3.b. Second part of the input file.

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

- T `(REAL)` Temperature, Kelvin

- ETA `(REAL)` Solvent viscosity, poises

- RM `(REAL)` Molecular weight . ***[c].***Alternatively you may give the value: `-1.,` and then the program will calculate the molecular weight from the sequence taken from the PDB file.

- VBAR `(REAL)` Partial specific volume, cm^3/g ***[c]*** Alternatively you may give the value `-1.` and then program will calculate the specific volume from the sequence taken from the PDB file.

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

3.C. Third part of the input file.

This part is intended for the calculation of some non-hydrodynamic properties that, in previous versions of our software, were considered separately in the computer program SOLPRO. 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.

For PDB structures, the coordinates are taken from the *.pdb* file, i.e., taken from the primary hydrodynamic model (PHM).

The data that you have to supply are:

- NH`(INTEGER)`, the number of values of the scattering (angular) variable, h. If you wish to omit the scattering calculation, the value given here should be 0, and the following line(HMAX) is omitted.

- HMAX `(REAL)`, the largest value of h (cm^{-1}), so that the scattering variable will range from 0 to HMAX

- 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, and the following line (RMAX) is omitted.

- RMAX `(REAL)`, the maximum length for the calculation of the distribution of distances (in cm). If you give any negative value, it is understood that RMAX will be the longest distance in the model, which is calculated by the program. The value will be given in the same units as the coordinates and dimensions in the structural file

- 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 bead-shell model with no 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 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(for models with only subunits, without additional spheres). 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

5. Hints and notes.

[a] From our experience with a large set of globular proteins, a value of 3.2 Angstroms is recommended. The experience with nucleic acids is not so extensive, and gives a slightly small value, about 2.8 Angstroms. An intermediate value, say 3.0 Angstroms may be recommended for nucleoprotein complexes.

[b] 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.

[c] This possibilities are only valid for protein (not nucleic acid) genuine PDB files. When the molecular weight, M, and/or the partial specific volume, vbar, are not given by the user, the program will read the aminoacid sequence from the SEQRES records in the PDB file, from which it may evaluate one of both of these quantities. If SEQRES records are not available, the program will estimate M=14*(number of non-H atoms) and vbar=0.7 cm^3/g. The calculated sedimentation coefficient is proportional to M(1-rho*vbar), and the calculated intrinsic viscosity is proportional to M. You can use these proportionalities if you later wish to recalculate those properties with other M and vbar. Translational and rotational diffusion quantities and relaxation times do not depend on M and vbar.

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

6. Release notes

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

- The number of atoms in the atomic model may be up to 1 000 000 (it was 300 000) in the previous version
- Calculation of non-hydrodynamic properties: radius of gyration and volume of the model, scattering function, distribution of distances and covolume
- File `summary.txt` contains the results for the distance distribution and the scattering function.
- The method known as “estimated extrapolation” (NSIG=1) that was implemented in the previous version is under reconsideration, and has been removed in this version.
- Visualization of the model in the VRML format

This program has been developed 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 window for program execution, while doing the other tasks (editing, visualization, etc) as usually in Windows. Executables for other platforms will be eventually available from our web site.