Bioinformatics, 31(8), 2015, 1319–1321
doi: 10.1093/bioinformatics/btu824
Advance Access Publication Date: 12 December 2014
Applications Note

OXFORD

Structural bioinformatics

HYCUD: a computational tool for prediction of effective rotational correlation time in flexible proteins

Nasrollah Rezaei-Ghaleh[1,2,*], Frederik Klama[1], Francesca Munari[1,2] and Markus Zweckstetter[1,2,3,*]

[1]Department of NMR-Based Structural Biology, Max Planck Institute for Biophysical Chemistry, 37077 Göttingen, Germany, [2]Structural Biology in Dementia, German Center for Neurodegenerative Diseases (DZNE), 37077 Göttingen, Germany and [3]Center for Nanoscale Microscopy and Molecular Physiology of the Brain, University Medical Center Göttingen (UMG), 37075 Göttingen, Germany

*To whom correspondence should be addressed.
Associate Editor: Anna Tramontano

Received on September 10, 2014; revised on November 12, 2014; accepted on December 8, 2014

Abstract

Motivation: A large fraction of eukaryotic proteins contain unstructured tails or linkers. The presence of flexible regions allows these systems to experience a high level of mobility facilitating their biological function. The complex nature of protein rotation in such flexible modular systems precludes a straightforward application of hydrodynamic methods to calculate their rotational motional properties. We describe the workflow of HYdrodynamic CoUpling of Domains (*HYCUD*), a program for prediction of effective rotational correlation times in multidomain proteins. The usage of *HYCUD* is demonstrated by its application to the ribosomal protein L7/L12. Rotational correlation times predicted by *HYCUD* might be used to detect molecular switch events mediated by disorder–order transitions in interdomain linkers.

Availability and implementation: The source code and documentation are available at www.mpibpc.mpg.de/106144/software.

Contact: mzwecks@gwdg.de or nare@nmr.mpibpc.mpg.de

Supplementary information: Supplementary material is available at *Bioinformatics* online.

1 Introduction

The function of biological macromolecules depends on their internal mobility supported by the presence of intrinsically disordered regions (Huang *et al*., 2014; Tompa, 2012). Reorientational motion of biomolecules can be followed through a variety of experimental techniques (Loman *et al*., 2010; Walker *et al*., 2004). However, for slowly tumbling systems, the use of these experimental techniques is limited. On the other hand, theoretical prediction of rotational motional behavior through standard hydrodynamic methods (Aragon, 2011) is complicated in these systems by the presence of several potentially coupled motions (Wong *et al*., 2009). To overcome these problems, modification of rigid-body hydrodynamics (Bae *et al*., 2009) as well as Brownian dynamics of coarse-grained models with the inclusion of hydrodynamic interactions (Amoros *et al*., 2013) have been employed. Here, we describe the workflow of a recently developed alternative approach (Rezaei-Ghaleh *et al*., 2013) called HYdrodynamic CoUpling of Domains (*HYCUD*) to predict the rotational correlation time (τ_c) of protein domains within flexible modular systems. The details of the *HYCUD* method, together with its application to the flexible two-domain ribosomal protein L7/L12 and protein X of Sendai virus, are reported.

2 Methods

The only input required for *HYCUD* calculations is a structural ensemble of the protein of interest. If high-resolution structures of the

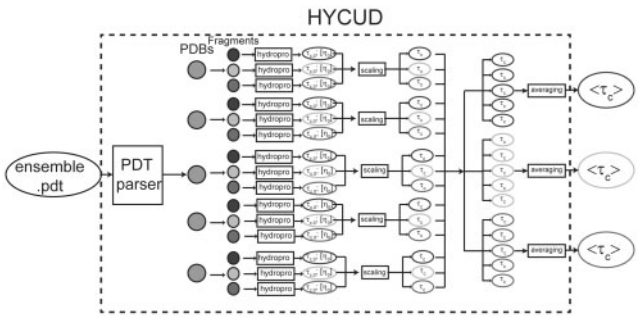

Fig. 1. Workflow of *HYCUD,* starting from a PDT file containing five PDB models each comprised by three fragments and yielding average rotational correlation time ($<\tau_c>$) for each fragment

protein domains are known, ensembles of the whole protein structures can be generated using freely available programs. One such program is ensemble optimization method (EOM) (Bernado *et al.*, 2007). The core of *HYCUD* is formed by a Python script that takes the aforementioned ensemble as a single PDT file (*ensemble.pdt*), and proceeds as follows to calculate the ensemble-average τ_c ($<\tau_c>$) of each protein module (Fig. 1, further details in Supplementary Information):

1. The *ensemble.pdt* input file is parsed, and a temporary directory structure is generated containing one protein databank (PDB) file for each model in the PDT file. If required, a third-party program is called to reconstruct backbone and side-chain coordinates on the basis of Calpha atom traces (Li and Zhang, 2009).
2. According to a user-defined scheme, which determines the boundary of fragments in the aminoacid sequence of the whole protein, each PDB model will be split into some fragments and placed in a temporary subdirectory. Definition of the boundaries of the folded domains is guided by the available experimental knowledge. With regard to the unfolded regions where little experimental evidence exists, fragment lengths are set to 12–14 residues—that is twice the typical persistence length of unfolded polypeptide chains (Schwalbe *et al.*, 1997).
3. For each fragment, the third-party program HYDROPRO (v. 10) (Ortega *et al.*, 2011) is called to calculate the hydrodynamic properties, i.e. rotational correlation time $\tau_{c,0}$ and intrinsic viscosity $[\eta_0]$. A parameter file *hydropro.dat* will guide this step of calculation. To prevent repetition of HYDROPRO calculations for the folded fragments which have the same size/shape across the structural ensemble, their $\tau_{c,0}$ and $[\eta_0]$ are calculated separately and subsequently fixed.
4. After extracting $\tau_{c,0}$ and $[\eta_0]$ of each fragment from HYDROPRO output result files, the script uses the pairwise distance between centers of mass of protein fragments to derive a scaling factor representing the 'effective viscosity' experienced by each fragment within the context of the whole molecule. The $\tau_{c,0}$ of each isolated fragment will then be multiplied by this scaling factor, yielding a new τ_c value, which is then corrected for the local viscosity effects.
5. In the last step, the corrected τ_c of each fragment calculated for any member of the protein ensemble will be averaged over the whole ensemble after elimination of outliers.

With respect to runtime, Steps 1 and 2 scale linearly with the number of PDB models in the ensemble and the total number of fragments per model. Moreover, these two steps are very fast such that they do not pose a significant limit to the speed of *HYCUD* calculations. The most time-consuming step involves the HYDROPRO calculations. As stated earlier, HYDROPRO calculations can be skipped for globular domains by fixing their $\tau_{c,0}$ and $[\eta_0]$. In this way, the computational cost of Step 3 will be mainly determined (linearly scaled) by the number of PDB models in the ensemble, the number of fragments in disordered regions per model and the parameter NSIG in HYDROPRO calculations. The computational costs of Steps 4 and 5 also scale linearly with the number of PDB models in the ensemble and quadratically (linearly for Step 5) with the total number of fragments per model, although they are typically negligible compared to Step 3. With a dedicated 8- or 12-core Linux machine, the HYCUD calculations for an ensemble of 500 models, each containing three globular regions and four disordered fragments (such as the dimeric L7/L12 protein) typically takes ~1–2 h.

3 Results

The 120-residue L7/L12 ribosomal protein contains two globular domains, an N-terminal domain (NTD) responsible for protein dimerization and a C-terminal domain (CTD) (Gudkov, 1997). The NTD and CTD are connected via a long flexible linker (Bocharov *et al.*, 1998). Here, we first generated an ensemble of 5000 random structures for the dimeric L7/L12 protein using the EOM program (Bernado *et al.*, 2007) (Supplementary Fig. S1). Subsequently, *HYCUD* calculations were performed to predict the effective τ_c of the dimerized NTD and each CTD within the full-length protein dimers. Supplementary Figure S2A shows that the *HYCUD* prediction for both NTD and CTD converged within an ensemble size of 500–1000. The τ_c distribution within the ensemble of dimeric L7/L12 is shown in Figure 1.

S2B. Supplementary Table S1 presents the results of *HYCUD* calculations. Close agreement with the experimental results (Bocharov *et al.*, 2004) was obtained considering the uncertainty range of experimental data and ~5% inaccuracy in hydrodynamic calculations (Ortega *et al.*, 2011). Further support for the validity of *HYCUD* is provided by the excellent agreement between predicted and experimental τ_c of the 53-residue domain of protein X from Sendai virus (Houben *et al.*, 2007) (Supplementary Fig. S3 and Supplementary Table S1) as well as several other examples reported in Rezaei-Ghaleh *et al.* (2013) (Supplementary Table S2). In conclusion, we described a computational tool to predict the effective rotational correlation time of domains in flexible multidomain proteins. The presented tool is easy to use and provides users insight into the rotational motion of protein modules within dynamic modular biomolecules.

Funding

This work was supported through the German Science Foundation Collaborative Research Center 860, Project B2 (M.Z.) and the German Science Foundation research unit FOR934 (M.Z.).

Conflict of Interest: none declared.

References

Amoros,D. *et al.* (2013) Prediction of hydrodynamic and other solution properties of partially disordered proteins with a simple, coarse-grained model, *J. Chem. Theory Comput.*, **9**, 1678–1685.

Aragon,S.R. (2011) Recent advances in macromolecular hydrodynamic modeling, *Methods*, **54**, 101–114.

Bae,S.H. *et al.* (2009) Prediction of the rotational tumbling time for proteins with disordered segments, *J. Am. Chem. Soc.*, **131**, 6814–6821.

Bernado,P. *et al.* (2007) Structural characterization of flexible proteins using small-angle X-ray scattering, *J. Am. Chem. Soc.*, **129**, 5656–5664.

Bocharov,E.V. *et al.* (1998) Conformational independence of N- and C-domains in ribosomal protein L7/L12 and in the complex with protein L10, *FEBS Lett.*, **423**, 347–350.

Bocharov,E.V. *et al.* (2004) From structure and dynamics of protein L7/L12 to molecular switching in ribosome, *J. Biol. Chem.*, **279**, 17697–17706.

Gudkov,A.T. (1997) The L7/L12 ribosomal domain of the ribosome: structural and functional studies, *FEBS Lett.*, **407**, 253–256.

Houben,K. *et al.* (2007) Intrinsic dynamics of the partly unstructured PX domain from the Sendai virus RNA polymerase cofactor P, *Biophys. J.*, **93**, 2830–2844.

Huang,J.R. *et al.* (2014) Transient electrostatic interactions dominate the conformational equilibrium sampled by multidomain splicing factor U2AF65: a combined NMR and SAXS study, *J. Am. Chem. Soc.*, **136**, 7068–7076.

Li,Y. and Zhang,Y. (2009) REMO: A new protocol to refine full atomic protein models from C-alpha traces by optimizing hydrogen-bonding networks, *Proteins*, **76**, 665–676.

Loman,A. *et al.* (2010) Measuring rotational diffusion of macromolecules by fluorescence correlation spectroscopy, *Photochem. Photobiol. Sci.*, **9**, 627–636.

Ortega,A. *et al.* (2011) Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models, *Biophys. J.*, **101**, 892–898.

Rezaei-Ghaleh,N. *et al.* (2013) Predicting the rotational tumbling of dynamic multidomain proteins and supramolecular complexes, *Angew. Chem. Int. Ed. Engl.*, **52**, 11410–11414.

Schwalbe,H. *et al.* (1997) Structural and dynamical properties of a denatured protein. Heteronuclear 3D NMR experiments and theoretical simulations of lysozyme in 8 M urea, *Biochemistry*, **36**, 8977–8991.

Tompa,P. (2012) Intrinsically disordered proteins: a 10-year recap, *Trends Biochem. Sci.*, **37**, 509–516.

Walker,O. *et al.* (2004) Efficient and accurate determination of the overall rotational diffusion tensor of a molecule from (15)N relaxation data using computer program ROTDIF, *J. Magn. Reson.*, **168**, 336–345.

Wong,V. *et al.* (2009) Influence of the coupling of interdomain and overall motions on NMR relaxation, *Proc. Natl. Acad. Sci.*, **106**, 11016–11021.