Hydrodynamic Radii of Intrinsically Disordered Proteins

Fast Prediction by Minimum Dissipation Approximation and Experimental Validation

R. Waszkiewicz[1,2], A. Michaś[1], M. K. Białobrzewski[1], B. P. Klepka[1], M. K. Cieplak-Rotowska[1], Z. Staszałek[1], B. Cichocki[2], M. Lisicki[2], P. Szymczak[2], A. Niedzwiecka[1]

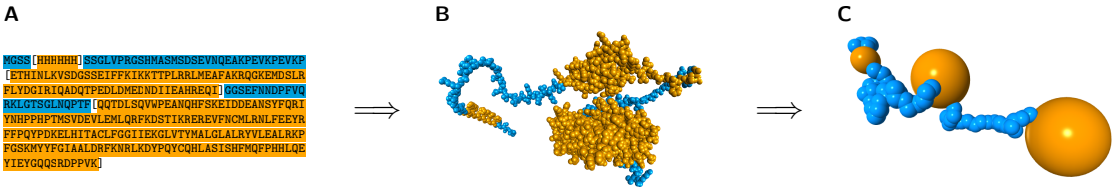
[1] Institute of Physics, Polish Academy of Sciences, Aleja Lotnikow 32/46, PL-02668 Warsaw, Poland

[2] Institute of Theoretical Physics, Faculty of Physics, University of Warsaw, L. Pasteura 5, 02-093 Warsaw, Poland

Abstract

The diffusion coefficients of globular and fully unfolded proteins can be predicted with high accuracy solely from their mass or chain length. However, this approach fails for intrinsically disordered proteins (IDPs) containing structural domains. We propose a rapid predictive methodology for estimating the diffusion coefficients of IDPs. The methodology uses accelerated conformational sampling based on self-avoiding random walks and includes hydrodynamic interactions between coarse-grained protein subunits, modeled using the generalized Rotne-Prager-Yamakawa approximation. To estimate the hydrodynamic radius, we rely on the minimum dissipation approximation recently introduced by Cichocki *et al.* Using a large set of experimentally measured hydrodynamic radii of IDPs over a wide range of chain lengths and domain contributions, we demonstrate that our predictions are more accurate than the Kirkwood approximation and phenomenological approaches. Our technique may prove valuable in predicting the hydrodynamic properties of both fully unstructured and multidomain disordered proteins.

Method



Construction of the coarse-grained globule-linker model (GLM) for an illustratory IDP, H6-SUMO-CNOT1(800-999), containing three ordered domains of different sizes. (A) Sequence with highlighted ordered (orange) and disordered (blue) segments, and domain boundaries marked by square brackets. (B) Full atom AlphaFold2 configuration (for visualization purposes only). Ordered clusters (orange) form dense blobs connected with linkers (blue). (C) Visualization of a representative configuration generated using the GLM method in which beads are displayed with their hydrodynamic radii.

Context

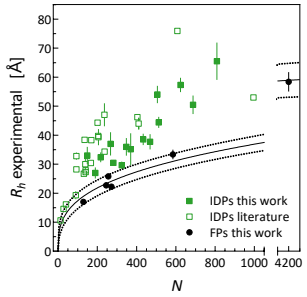


Figure 1. Experimental hydrodynamic size values plotted against number of amino acid residues in the protein chain (*N*), and a power-law curve fitted to folded proteins (FP) together with 95% confidence band.

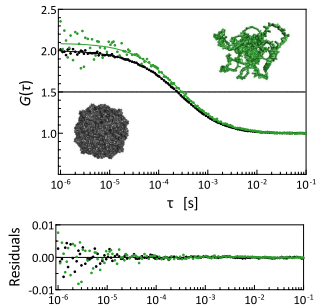


Figure 2. Examples of normalized FCS autocorrelation curves with raw fitting residuals for an intrinsically disordered H_6-SUMO-GW182SD-mCherry (N = 809, R_h = 66±6 Å) (green) in comparison with apoferritin (N = 4200, R_h = 58±3 Å) (black).

Hydrodynamic size from a conformational ensemble

Kirkwood-Riseman (1954)

$$R_h^{\mathrm{KR}} = \left(\frac{1}{N^2} \sum_i^N \sum_{j \neq i}^N \left\langle \frac{1}{r_{ij}} \right\rangle \right)^{-1}$$

Radius computed from diffusion of centre of mass
Equal contribution of all beads

Minimum Dissipation Approximation

$$R_h^{\mathrm{MDA}} = \mathbf{1}^{\mathrm{T}} \left(2\pi\eta \left\langle \mathrm{tr}(\boldsymbol{\mu}(r_{ij})) \right\rangle\right)^{-1} \mathbf{1}$$

Radius computed from centre of diffusion
Optimal contribution of each bead

Results

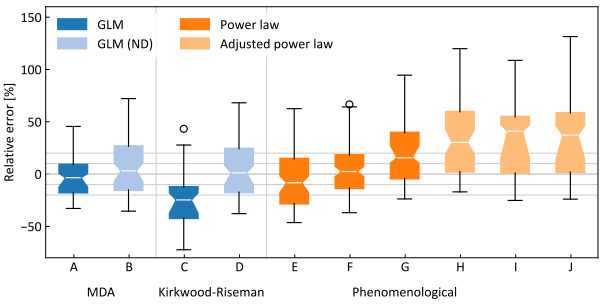


Figure 3. Comparison of different methods of estimation of hydrodynamic size. Boxes show interquartile range with median confidence bands marked by notches. MDA with GLM ensemble generation (**A**) performs best on the IDP benchmark set with standard errors of 18.15 % and 7.09 Å (compared to 24.80 % and 8.46 Å for a simple power law). Methods based on the Kirkwood-Riseman estimation (**C,D**) typically underestimate hydrodynamic size of the molecule. Power law fits with one free parameter (**E**) and two free parameters (**F**) evaluated using leave-one-out cross validation are compared with the formerly reported power law (Marsh and Forman-Kay, 2010) (**G**) and models based on polyproline II structure propensities without (Tomasso et al., 2016) (**H**) and with (English et al., 2017) (**I**) regard to the charge, and a sequence-based model (Marsh and Forman-Kay, 2010) (**J**) which takes into account total charge of the molecule. Theoretical methods with no knowledge about the presence of domains in the IDP (ND; **B,D**) significantly overestimate the hydrodynamic size of the molecule. Domain data can be incorporated into our ensemble generation engine leading to more accurate estimates of hydrodynamic size (**A**). Note that experimental uncertainty also contributes to the errors presented above

Bibliography

Cichocki, B., Rubin, M., Niedzwiecka, A., & Szymczak, P. (2019). Diffusion coefficients of elastic macromolecules. *Journal of Fluid Mechanics, 878*, R3.

Zuk, P. J., Cichocki, B., & Szymczak, P. (2018). GRPY: an accurate bead method for calculation of hydrodynamic properties of rigid biomacromolecules. *Biophysical Journal, 115*(5), 782–800.

English, L. R., Tilton, E. C., Ricard, B. J., & Whitten, S. T. (2017). Intrinsic alpha helix propensities compact hydrodynamic radii in intrinsically disordered proteins. *Proteins: Structure, Function, and Bioinformatics, 85*(2), 296–311.

Marsh, J. A., & Forman-Kay, J. D. (2010). Sequence determinants of compaction in intrinsically disordered proteins. *Biophysical Journal, 98*(10), 2383–2390.

Tomasso, M. E., Tarver, M. J., Devarajan, D., & Whitten, S. T. (2016). Hydrodynamic radii of intrinsically disordered proteins determined from experimental polyproline II propensities. *PLoS Computational Biology, 12*(1), e1004686.

Conclusions

- Our benchmark set is a big step forward – it includes a higher conformational variety, more multidomain proteins, longer chains, and a much wider range of charge states compared to the reference sets used previously.
- Our first-principles model predicts R_h without any fitting parameters and with better precision than published phenomenological models.
- All of the calculations for a given protein can be performed in about a minute on a typical laptop using our Python package or via a Colab notebook.

Check out our paper!

Waszkiewicz, R., Michas, A., Białobrzewski, M.K., Klepka, B.P., Cieplak-Rotowska, M.K., Staszałek, Z., Cichocki, B., Lisicki, M., Szymczak, P. and Niedzwiecka, A., (2024): *Hydrodynamic Radii of Intrinsically Disordered Proteins: Fast Prediction by Minimum Dissipation Approximation and Experimental Validation.* J. Phys. Chem. Lett., 15(19), 5024-5033