Hydrodynamic Radii of Intrinsically Disordered Proteins: Fast Prediction by Minimum Dissipation Approximation and Experimental Validation

Radost Waszkiewicz,[†,§] Agnieszka Michaś,[‡,§] Michał K. Białobrzewski,[‡]

Barbara P. Klepka,[‡] Maja K. Cieplak-Rotowska,[‡,¶] Zuzanna Staszałek,[‡]

Bogdan Cichocki,[†] Maciej Lisicki,[†] Piotr Szymczak,[*,†] and Anna Niedzwiecka[*,‡]

†*Institute of Theoretical Physics, Faculty of Physics, University of Warsaw, L. Pasteura 5, 02-093 Warsaw, Poland*

‡*Institute of Physics, Polish Academy of Sciences, Aleja Lotnikow 32/46, PL-02668 Warsaw, Poland*

¶*present address: IMol Polish Academy of Sciences, Flisa 6, PL-02247 Warsaw, Poland*

§*These authors have contributed equally to this work and share first authorship*

E-mail: piotrek@fuw.edu.pl; annan@ifpan.edu.pl

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.

TOC Graphic
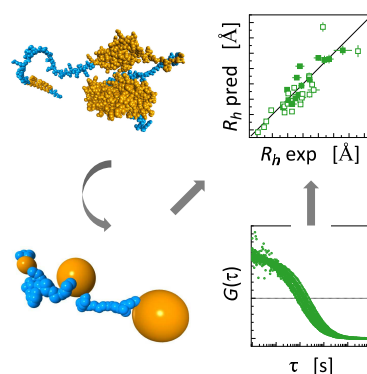
Keywords

intrinsically disordered protein, hydrodynamic radius, diffusion coefficient, coarse-grained model, minimum dissipation approximation

Intrinsically disordered proteins (IDPs) constitute an extensive class of biological macromolecules, and their role in the homeostasis of a living cell has been increasingly recognized in recent decades.[1,2] The frequency of long intrinsically disordered regions (IDRs) in proteins differs significantly between the kingdoms of life, ranging from 2 % in archaea to 33 % in eukaryotes.[3] The IDP molecules display different degrees of structural disorder. Their chains can encompass either several folded globular domains, or supersecondary structures connected by flexible linkers, sparse secondary structural elements, or can be completely natively unstructured. Disordered proteins exhibit a notable characteristic – the absence of a stable, well-defined relative spatial arrangement of their fragments. Instead, their equilibrium properties can be described through a broad set of rapidly inter-converting conformers, posing a challenge for analysis, particularly in the context of long chains.[4]

The average geometric properties of IDPs, including their shape and size, are determined by the equilibrium ensemble of conformational states. This equilibrium state is intricately influenced by environmental conditions,[5] such as temperature,[6] ionic strength[7,8], osmolality,[9] crowding[10], post-translational modifications[11], and the presence of specific molecular binding partners[12]. The formation of transient or more stable non-covalent complexes introduces another non-trivial dependence of the IDP equilibrium geometry on environmental factors.

Because the shape and availability of the binding sites necessary for the interaction of IDP with ligands, other proteins, and nucleic acids are greatly influenced by the environment, IDPs often act as higher-order regulators in key cellular processes such as gene expression[11,13] signaling[2,14], or extracellular biomineralization[15]. It is the different conformations of these flexible proteins, which enable IDPs to perform their multiple functions[1]. In particular, it is worth emphasizing the important roles of IDPs in health and disease, e.g., the role of the p53 protein as a tumour suppressor[16], mutations of which are often responsible for human cancers, the function of 4E-BPs in the inhibition of eukaryotic translation initiation[11,17–19], the significance of GW182 protein in the recruitment of the multi-protein machinery necessary for microRNA-mediated gene

silencing[20–22], or the importance of Tau, FUS, and α-synuclein proteins in neurodegenerative diseases[23,24]. Since the elastic properties of these biomolecules are responsible for the proper functioning of IDPs in the cellular context, *i.e.* for the association of complexes and the formation of biomolecular condensates via liquid-liquid phase separation such as, e.g., RNA-processing membraneless organelles[25,26] much attention is paid to the hydrodynamic properties of IDPs. Experimental techniques, such as analytical ultracentrifugation (AUC), size exclusion chromatography (SEC), pulsed-field gradient nuclear magnetic resonance (PFG–NMR), dynamic light scattering (DLS), and fluorescence correlation spectroscopy (FCS), offer insights into hydrodynamic parameters (as reviewed in[27]). However, due to the distinct limitations of each experimental approach, ongoing research aims to devise phenomenological methods for calculating the hydrodynamic radius (R_h). These methods may involve deriving R_h from the radius of gyration determined by small-angle X-ray scattering (SAXS)[28,29] or exploiting the conformational backbone propensity of IDPs.[30,31]

Simultaneously, significant effort is being invested in developing numerical models that extract the characteristics of IDPs from conformational ensembles obtained using molecular dynamics (MD) simulations or energy minimization algorithms[32–41]. However, the molecular flexibility of IDPs introduces substantial complexities when determining their hydrodynamic properties. Two main issues here are the large number of degrees of freedom and the long timescales of relaxation of the internal coordinates of the molecules. These factors prohibit direct calculation of the experimentally relevant long-time diffusion coefficient from either molecular or Brownian dynamics trajectories. One popular approximation that circumvents this difficulty is to assume that the macromolecule is rigidly frozen in one of a large number of possible conformations. Transport properties are then calculated by treating the molecule as a rigid body, and the results are averaged over an equilibrium ensemble[42–45]. Nevertheless, the validity and accuracy of this approximation remain uncertain. Additionally, the generation of conformational ensembles can be a bottleneck for long chains (beyond approximately 300 amino acid residues) because it requires time-consuming MD simulations and/or the construction of new databases of

short peptide conformations.

There is, therefore, a strong need to develop a numerically efficient solution that would enable reliable calculation of the long-time diffusion coefficient of any long-chain IDP, such as one with 1000 amino acid residues, solely based on its sequence information.

In this study, we introduce a new theoretical approach to both generating conformational ensembles of IDPs and calculating their hydrodynamic properties. This method enables a swift estimation of the diffusion coefficient for long IDPs in a matter of minutes, with superior accuracy compared to existing methods. This assertion is substantiated through rigorous testing of the model on a diverse set of experimental results obtained for 43 proteins. The dataset includes both literature data and R_h values measured for a set of new IDP constructs using FCS under mild conditions (see Supporting Information).

We present our results in terms of the hydrodynamic radius of a molecule, R_h. This radius represents the size of a solid sphere that possesses the same translational diffusion coefficient, D, as the given molecule under identical buffer conditions. Therefore, $R_h = k_B T / 6\pi\eta D$, where T is the temperature and η is the viscosity.

An important observation by Fixman[46,47] is that the diffusion coefficient of a flexible macromolecule is time-dependent, with well-defined short- and long-time limits. The disparity between the two is attributed to the effects associated with relaxation of the internal coordinates of the molecule, as well as rotation of the macromolecule as a whole.[46,48,49] The positivity of the dissipation rate in the system implies that the long-time diffusion coefficient (D_l) is always smaller than the short-time diffusivity (D_s).[47] The focus of theoretical approaches should be the determination of the former quantity, as it is the one measured in experiments utilizing techniques like FCS, AUC, or DLS. Unfortunately, the calculation of D_l is significantly more challenging than that of D_s because it involves the computation of time-dependent quantities, such as the memory function, which describes the relaxation effects. An additional point to keep in mind is that the value of the short-time diffusion coefficient depends on the choice of the point that one tracks.[49–52] In contrast, the long-time diffusivity is independent of the choice of reference point.[53]

The methods for predicting the diffusion coefficient can be broadly split into three

categories: atomistic, phenomenological, and coarse-grained. For small proteins, high-resolution, atomistic MD methods can be used,[54] but they require either simulating the surrounding water molecules explicitly, which is very computationally intensive, or an implicit solvent scheme. In the case of implicit solvent methods, addressing hydrodynamic interactions between distant parts of the molecule[55–58] and thermalization[59] pose significant challenges. Additionally, even for the smallest proteins, it is prohibitively difficult to obtain statistically meaningful data over the 10-100 millisecond scale, which would enable the direct computation of the long-time diffusion coefficient.

The other extreme are phenomenological models that predict R_h from the number of residues N and possibly other parameters, such as total charge or amino acid composition. Theoretical considerations of Rouse, who modelled a protein as a Gaussian chain[60] gave foundation to the power law relationship $R_h \sim N^{1/2}$. The classical Rouse model employs random displacements between the monomers. If we assume complete independence of displacements between each consecutive pair of monomers, the central limit theorem dictates that as N approaches infinity, the squared end-to-end distance should conform to a scaled $\chi^2(3)$ distribution. Consequently, the dimensions of such an idealized chain are expected to scale with $\sqrt{N}$. Later work of Zimm included the effect of excluded volume[61], which resulted in the scaling $R_h \sim N^\gamma$ with $\gamma = 0.588$.

Phenomenological size–length relationships that include other variables involve a number of fitting parameters. As a result, their range of applicability outside of the fitting dataset is difficult to assess. An alternative phenomenological approach proposed by Pesce et al.[29] employs the radius of gyration obtained from SAXS experiments to estimate R_h. This is substantiated by the observation that within the Kirkwood-Riseman approximation[62] R_h and R_g share the same scaling relationship with N as long as the pair-displacement distribution converges under appropriate scaling to a Gaussian for large N.

Finally, coarse-grained models, like our method, employ larger units (typically one or two per amino acid residue) as building blocks for the structure prediction scheme, along with approximate interaction potentials between subunits, to simulate the equilibrium

6

ensemble of configurations for a given molecule. These configurations are then combined with an approximation of the hydrodynamic properties to compute the diffusion coefficient. Essentially, the computation of the latter for elastic macromolecules addresses two interconnected challenges: predicting the conformations of molecules based on available biochemical data and then using these conformations to predict hydrodynamic properties.

The different exponents in the power-law relationships of Rouse[60] and Zimm[61] demonstrate that even the most basic method for approximating configurations must take into account excluded volume interactions.

A software that can accommodate excluded volume interactions for a disordered chain is Flexible Meccano (FM).[34] In addition to volume exclusion, it considers the distribution of Ramachandran angles determined from crystallographic protein structures when sampling conformations. However, FM treats the entire chain as unstructured, so it cannot be used to model proteins that possess both globular and unfolded segments, which are in fact much more common than fully unstructured chains. Unfortunately, FM has a closed license that precludes necessary modifications to accommodate folded regions of proteins.

The complex angle distributions used by FM are crucial when computing NMR parameters that are sensitive to short-range details of the pair-distribution function, such as residual dipolar couplings, paramagnetic relaxation enhancement, or J-coupling. However, upon closer examination, the pair-distance distribution generated by FM and a simpler model presented in this paper, globule-linker model (GLM; described below), become virtually identical for amino acids separated by more than 15 residues along the chain.

The highly localized differences between structures at small sequential distances have a minimal influence on the estimations of R_h. It is important to recall that for amino acid residues separated by a distance r, the dipolar coupling decays as r^{-3}, while the decay rate of hydrodynamic interactions (HI) is only r^{-1}. Therefore, HI are long-range and less sensitive to near-neighbor distributions, with contributions to the diffusion coefficient of near-neighbors and far-neighbors being $O(N)$ and $O(N^{2-\gamma}) = O(N^{1.4})$, respectively.
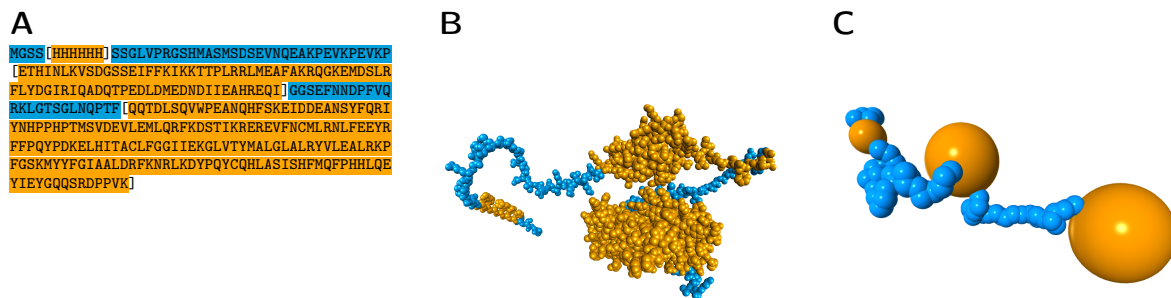
7

Figure 1. Construction of the coarse-grained globule-linker model (GLM) for an illustratory IDP, H_6-SUMO-CNOT1(800-999), containing three ordered domains of different sizes (no. 28 in Table S1). **A)** Sequence with highlighted ordered (orange) and disordered (blue) segments, and domain boundaries marked by square brackets. **B)** Representative full atom conformation generated by AlphaFold2 (for visualisation purposes only,[63,64] beads with van der Waals radii, hydrogen atoms omitted for clarity); ordered clusters (orange) form dense blobs connected with linkers (blue). **C)** Visualisation of a representative configuration generated using the GLM method where beads are displayed with their hydrodynamic radii.

Guided by these considerations, we have implemented the simplest extension of Zimm's chain — the globule-linker model (GLM), designed to comprehensively represent IDPs containing globular domains connected by unstructured fragments. In the model (Figure 1 A-C), we represented the protein as an assembly of spheres of different sizes. Disordered segments of length N were modeled as chains of N identical spheres, each with a diameter equal to the C_α-C_α distance, while structured domains were represented by single spheres, each of them described by their mass (m). The size of a sphere representing a structured domain was computed as $R_h = (3m/4\pi\rho_{\text{globular}})^{1/3} + a_{\text{hydration}}$ with $\rho_{\text{globular}} = 0.52$ Da / $Å^3$,[65] and the single layer hydration shell taken to be $a_{\text{hydration}} = 3$ Å thick. Within this approach, information about domain boundaries in the protein sequence is sufficient to construct an appropriate bead approximation of the IDP. The identification of protein sequence fragments to be treated as ordered regions and mimicked by larger beads in the GLM model was done using Disopred3.[66] The fragment was assumed to be ordered if the disorder probability P was less than 50 % for at least three subsequent amino acid residues, including loops linking such fragments but not exceeding 14 residues.[67]

Using a recursive approach, it is possible to generate GLM conformations with a time complexity of $O(N^{1+\gamma})$, which provides a satisfactory ensemble for the largest of the proteins considered here in under a minute using only a personal computer (a single thread

at 1.8GHz). The speed of the recursive approach should be contrasted with an iterated one where steps are simply added one by one, and intersecting chains are discarded. This easier-to-implement method is characterized by a time complexity of $O(\exp(N))$, which becomes prohibitively slow for chains with $N > 20$.

We have transformed the sampled conformations into a hydrodynamic model by increasing bead sizes in the disordered fragments to $R_{\text{disordered}} = 4.2$ Å, corresponding to the median value for all aminoacids.[68] In the resulting hydrodynamic model of linkers, neighbouring beads show substantial overlaps, requiring a careful treatment of the mobility matrices (see[69] for details). Note that the value of $R_{\text{disordered}}$ has only a minor impact on the final results, since the hydrodynamic radius of long slender filaments depends logarithmically on their thickness.[70–73]

To compute R_h from the estimated ensembles we have implemented two algorithms: the Kirkwood formula, and minimum dissipation approximation (MDA) method of Cichocki *et al.*[53] Within the first approach,[74] the hydrodynamic radius of a macromolecule is approximated by

$$\frac{1}{R_h^{\text{K}}} = \frac{1}{N^2} \sum_{i=1}^{N} \left(\frac{1}{a_i} + \sum_{j=1,j\neq i}^{N} \left\langle \frac{1}{r_{ij}} \right\rangle \right), \tag{1}$$

where N is the total number of beads in the IDP model, a_i is the hydrodynamic radius of bead i, $r_{ij} = |\mathbf{r}_j - \mathbf{r}_i|$ is the distance between beads i and j, and $\langle \cdot \rangle$ denotes the average over the equilibrium ensemble. One can show that this corresponds to the ensemble-averaged short-time diffusion coefficient of the geometric center of the macromolecule, $\mathbf{r}_c = N^{-1} \sum_{i=1}^{N} \mathbf{r}_i$. Note that the geometric center fluctuates as the shape of the molecule evolves and does not correspond to any fixed position within it. A simplified form of the Kirkwood formula is often used[39,75,76]

$$\frac{1}{R_h^{\text{K}}} \approx \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1,j\neq i}^{N} \left\langle \frac{1}{r_{ij}} \right\rangle, \tag{2}$$

where the single-bead terms $1/a_i$ are dropped, as their contribution becomes negligible in the large N limit. This is the form that we will also use in the present work.

A better estimate of R_h, corresponding to the long-time diffusion coefficient, requires

a more in-depth description of the hydrodynamic interactions between the beads. To this end, one introduces the mobility matrix $\boldsymbol{\mu}$,[48] which links the velocities of the beads with the forces acting on them, according to

$$\mathbf{U}_i = \sum_j \boldsymbol{\mu}_{ij} \mathbf{F}_j, \tag{3}$$

where $\mathbf{U}_i$ is the velocity of bead i whereas $\mathbf{F}_j$ is the force with which bead j acts on the fluid. Based on the mobility matrix, one defines a matrix $\mathbf{A}$ indexed by the bead labels (i, j), $A_{ij} = 2\pi\eta \mathrm{Tr}\left\langle \boldsymbol{\mu}_{ij} \right\rangle$ and its inverse $\mathbf{B} = \mathbf{A}^{-1}$. One can then construct the MDA[53] for R_h as

$$R_h^{\mathrm{MDA}} = \sum_{i,j} B_{ij}. \tag{4}$$

Note that the above formula is general, and can be used for different models of hydrodynamic interactions - both simple models (*e.g.* Oseen or Rotne–Prager far-field approximation[77]) or in more sophisticated approaches, like the multipole expansion method[78,79]. In this work, we use the generalized Rotne-Prager approximation for the calculation of the mobility matrix, as described in[80–82]. This approximation is now also available as a Python package, `pygrpy`.[83] For non-overlapping beads, the elements of the matrix $\mathbf{A}$ have then a particularly simple form: $A_{ij} = \left\langle 1/r_{ij} \right\rangle$ for $i \neq j$, and $A_{ii} = 1/a_i$. The formulas for overlapping beads can be found in the Supporting Information.

The MDA corresponds to the calculation of the short-time diffusion coefficient of the diffusion center of a molecule,[52] which is a point inside the molecule where D_s is minimal. The position of the diffusion center is $\mathbf{r}_d = \sum_{i=1}^{N} x_i \mathbf{r}_i$, with the weights given by $x_i = \sum_j B_{ij} / \sum_{k,j} B_{kj}$. Since D_s is always larger than its long-time counterpart, D_l, MDA provides the best estimation for the long-time diffusion coefficient out of all methods that utilize D_s for this purpose. The MDA turns out to be more robust when dealing with large differences in the sizes of beads used to model constituent parts of the macromolecule, because in such cases the equal weights of the geometric center of the macromolecule differ significantly from the optimal weights of the diffusion center.

We combined each method of generating conformers with each method of computing
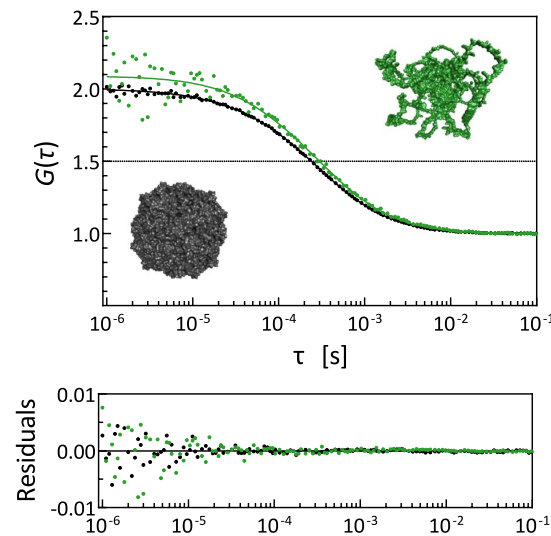
Figure 2. Examples of normalized FCS autocorrelation curves with raw fitting residuals for an intrinsically disordered H_6-SUMO-GW182SD-mCherry ($N = 809$, $R_h = 66 \pm 6$ Å) (green) in comparison with apoferritin ($N = 4200$, $R_h = 58 \pm 3$ Å) (black). Crystal structure of apoferritin (pdb id. code 2w0o[84]) and putative conformation of H_6-SUMO-GW182SD-mCherry predicted by AlphaFold[63] are shown for illustration purposes, preserving relative sizes of solvent accessible surfaces of atoms.

R_h, which resulted in four different theoretical approaches, the predictions of which (Table S2) were then compared with experimental data. For this purpose, we have obtained 15 new IDP constructs covering a wide range of chain lenghts, folded domain contents and charge states, and determined their R_h using FCS (Figure 2 and S2, S3, S4, S5, S6; for further experimental details see Supporting Information).

The experimental benchmark set (Table S1) was thus composed of both the new FCS measurements, and R_h selected from the literature based on the following criteria: the proteins had sequences could be unambiguously identified in the literature or in the UniProtKB database, were measured at well defined, mild conditions (temperature of 20-26 °C, buffer of pH 7-8, ionic strength corresponding to 75-300 mM NaCl), and their hydrodynamic radii were determined directly from appropriate experiments without conversions from other experimental quantities, such as R_g.[29,85–102] This is, to our best knowledge, the largest benchmark set encompassing experimental R_h for 38 IDPs and 6 globular model proteins, measured at comparable conditions.

The results of tests performed for our four theoretical approaches against the bench-
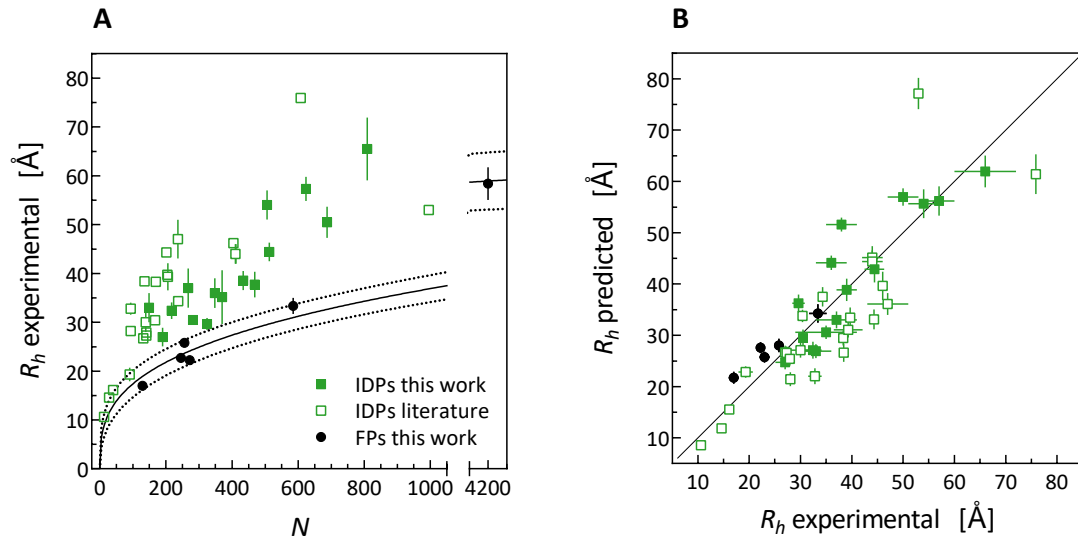
11

Figure 3. **A)** Experimental R_h values plotted against number of amino acid residues in the protein chain, N, and power-law curve fitted to R_h of folded proteins (FP) together with 95% confidence band. **B)** Direct comparison of the predicted vs. measured R_h values for all of the proteins modeled using the MDA+GLM approach.

Table 1. Comparison of error statistics of various models.

model	n_{fp}	RMSD[Å]	RMSRD[%]	R^2	R^2_{adj}	Q_3^{AE}[Å]	Q_3^{RE}[%]
MDA + GLM	0	7.09	18.15	0.71	0.71	6.80	22.51
MDA + GLM(ND)	0	9.48	28.02	0.48	0.48	11.88	29.31
KR + GLM	0	12.82	34.69	0.05	0.05	17.59	42.95
KR + GLM(ND)	0	9.25	27.31	0.50	0.50	11.11	29.44
random coil	1	9.60	27.71	0.47	0.45	10.28	33.69
power law	2	8.46	24.80	0.59	0.56	9.63	26.08
power law (lit.[103])	2	12.01	36.94	0.16	0.12	14.37	39.51
PPII-based (lit.[30])	3	17.25	49.09	-0.72	-0.86	20.62	59.54
sequence-based (lit.[103])	7	22.90	50.78	-2.05	-2.66	19.59	58.32

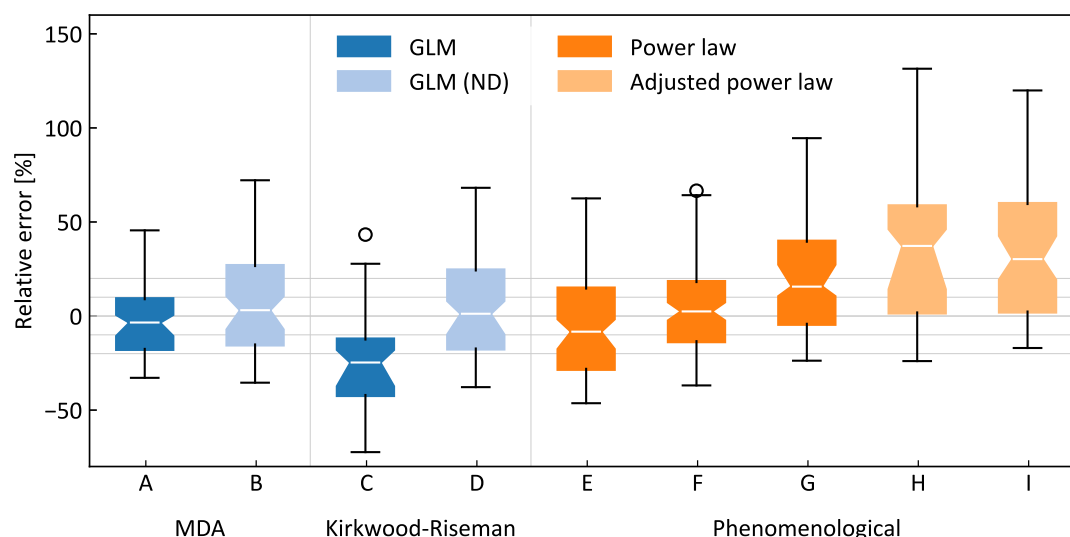n_{fp} – number of fitting parameters, ND – no domain information, Q_3 – 3$^{\text{rd}}$ quartile

12

Figure 4. Comparison of different methods of estimation of R_h. 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 R_h 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[103] (**G**) and a sequence-based model[103] (**H**) which takes into account total charge of the molecule, and a model based on polyproline II structure propensities[30] (**I**). Methods with no knowledge about the presence of domains in the IDP (ND; **B**,**C**) significantly overestimate the hydrodynamic size of the molecule. Domain data can be incorporated into our ensemble generation engine leading to more accurate estimates of R_h (**A**). Note that experimental uncertainty also contributes to the errors presented above and in Table 1.

mark set are gathered in Table 1, and Figure 4 shows a visual comparison of the deviations between theory and experiment. Additionally, we provide power law fits[104,105] and power law fits of Marsh *et al.*[103] for comparison of the prediction accuracy (Table S3).

We compare the accuracy of the previous and new model under six metrics (Table 1): the square root of mean square deviation (RMSD), square root of the mean square relative deviation (RMSRD), Pearson's coefficient (R^2), Pearson's coefficient adjusted for fitting parameters (R^2_{adj}), 3$^{\text{rd}}$ quartile of the absolute error Q_3^{AE}, and 3$^{\text{rd}}$ quartile of the relative error Q_{75}^{RE}. Whenever a fitting procedure is required, we use leave-one-out cross-validation to compute error metrics. We have chosen to test the relative deviations as well in order to reduce the undue weight given to the new, very long sequences in our dataset. Similarly, outlier-robust metrics of the 3$^{\text{rd}}$ quartile were included to reduce the impact of a single sequence misprediction on the final comparisons. In all evaluation metrics, the MDA+GLM approach performs the best. Surprisingly, it is the only model that performs better than the power law baseline in any of the evaluation metrics.

Although explicit intramolecular interactions of the amino acid residues are neglected in MDA+GLM approach, the main cause of discrepancies between the experimental and predicted R_h values (Figure S8) appeared to be the intrinsic properties of individual experimental methods, which suffer from typical errors or limitations and are usually not taken into account when reporting the final experimental results. PGF-NMR measurements are the most unambiguous and accurate, but their effective application is limited to smaller proteins (up to 200-300 amino acid residues long) at high concentrations. It is worth noting that the agreement of values of R_h predicted by MDA+GLM with the PGF-NMR results is excellent (Figure S8 C). FCS is the only method that addresses the self-diffusion of molecules at the limit of low concentrations. Raw FCS measurements can be refined to exclude possible oligomerization or aggregation during the experiment based on the count-rates. However, it is impossible to avoid proteolytic instability of proteins and, consequently, the appearance of impurities with a lower molar mass, which may potentially result in apparently lower values of R_h (Figure S8 B). On the other hand, SEC is the easiest approach to remove lower mass impurities, but it involves diffusion of

14

molecules at higher concentrations through a medium with pores of a specific shape under the influence of pressure. An additional common disadvantage is calibration based on R_h of standard proteins determined in various conditions and the lack of appropriate propagation of the calibration experimental uncertainty. Consequently, SEC measurements can be highly scattered (Figure S8 D). The largest outlier in our analysis concerns R_h determined using SEC for fesselin without providing experimental uncertainty (Id. 43, Tables S1, S2, Figure S7, S8 D). The DLS method is the most prone to overestimating experimental values (Figure S8 E), since the presence of even a small number of aggregates with a larger molar mass generates a huge contribution to the intensity of scattered light. Finally, AUC yields sedimentation coefficients, and their interpretation in terms of exact values of R_h requires some assumptions that are not obvious for IDPs, such as *e.g.* partial specific protein volume.[106] The second largest outlier in our set is the OMM-64 protein (Id. 39, Tables S1, S2, Figure S7, Figure S8 F) with the R_h value determined using AUC, which is very close to the power function curve for completely denatured proteins.[107]

In conclusion, we have presented a simple, first principles model for the prediction of R_h without any fitting parameters and achieved favourable comparison with a large benchmark set. Moreover, due to the relative simplicity of the model, all of the calculations for a given protein can be performed in about a minute on a typical laptop, which is contrasted with MD-based conformer generation methods that require supercomputers and take many days. Furthermore, to our surprise, the GLM-MDA approach demonstrates satisfactory convergence even with ensemble sizes as small as 40 conformers.

Our benchmark set, in which the previously known IDPs were complemented by a set of newly obtained proteins, constitutes a significant step forward in predicting hydrodynamic properties of IDPs. It includes a higher conformational variety, with a stronger emphasis on multidomain proteins, longer chains, and a much wider range of charge states compared to the reference sets used previously.[30,103] This diversity allows for more reliable testing of theoretical models.

Further developments of the MDA+GLM model are needed to take into account the

dependence of R_h on the environmental conditions[6–8] and the formation of complexes. However, our results clearly demonstrate that the relatively simple globule–linker model for conformational ensemble construction, in combination with the minimum dissipation approximation, can serve as the starting point for developing further phenomenological corrections. These improvements could incorporate factors such as amino acid sequence composition, residue charge, and counterion binding. When using MDA+GLM, all excluded volume effects should be correctly accounted for, with any further deviations hinting at interesting physical and chemical properties of the molecules.

Acknowledgement

Supporting Information Available

`Tables S1 S5 Figures S1-S8 Materials and Methods.pdf` containing experimental R_h , Figures S1, S2, S3, S4, S5, S6, S7, S8, and additional methodological details. `Tables S2-S4.xlsx` containing theoretical R_h from GLM, phenomenological R_h, and protein sequences with marked domains.

The programs referenced in this article can be found on GitHub. For reader convenience, an API and a command-line Python utility, `glm_mda_diffusion`, have been provided.[108] This package builds upon previous ones, namely `pychastic` and `sarw_spheres`, both of which are also accessible on GitHub.

References

(1) Oldfield, C. J.; Dunker, A. K. Intrinsically disordered proteins and intrinsically disordered protein regions. *Annual Review of Biochemistry* **2014**, *83*, 553–584.

(2) Wright, P. E.; Dyson, H. J. Intrinsically disordered proteins in cellular signalling and regulation. *Nature Reviews Molecular Cell Biology* **2015**, *16*, 18–29.

(3) Ward, J. J.; Sodhi, J. S.; McGuffin, L. J.; Buxton, B. F.; Jones, D. T. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *Journal of Molecular Biology* **2004**, *337*, 635–645.

(4) Shea, J.-E.; Best, R. B.; Mittal, J. Physics-based computational and theoretical approaches to intrinsically disordered proteins. *Current Opinion in Structural Biology* **2021**, *67*, 219–225.

(5) Uversky, V. N. Intrinsically disordered proteins and their environment: effects of strong denaturants, temperature, pH, counter ions, membranes, binding partners, osmolytes, and macromolecular crowding. *The Protein Journal* **2009**, *28*, 305–325.

(6) Langridge, T. D.; Tarver, M. J.; Whitten, S. T. Temperature effects on the hydrodynamic radius of the intrinsically disordered N-terminal region of the p53 protein. *Proteins: Structure, Function, and Bioinformatics* **2014**, *82*, 668–678.

(7) Müller-Späth, S.; Soranno, A.; Hirschfeld, V.; Hofmann, H.; Rüegger, S.; Reymond, L.; Nettels, D.; Schuler, B. Charge interactions can dominate the dimensions of intrinsically disordered proteins. *Proceedings of the National Academy of Sciences* **2010**, *107*, 14609–14614.

(8) Wohl, S.; Jakubowski, M.; Zheng, W. Salt-dependent conformational changes of intrinsically disordered proteins. *The Journal of Physical Chemistry Letters* **2021**, *12*, 6684–6691.

(9) Moses, D.; Yu, F.; Ginell, G. M.; Shamoon, N. M.; Koenig, P. S.; Holehouse, A. S.; Sukenik, S. Revealing the hidden sensitivity of intrinsically disordered proteins to

their chemical environment. *The Journal of Physical Chemistry Letters* **2020**, *11*, 10131–10136.

(10) Wang, Y.; Benton, L. A.; Singh, V.; Pielak, G. J. Disordered protein diffusion under crowded conditions. *The Journal of Physical Chemistry Letters* **2012**, *3*, 2703–2706.

(11) Bah, A.; Vernon, R. M.; Siddiqui, Z.; Krzeminski, M.; Muhandiram, R.; Zhao, C.; Sonenberg, N.; Kay, L. E.; Forman-Kay, J. D. Folding of an intrinsically disordered protein by phosphorylation as a regulatory switch. *Nature* **2015**, *519*, 106–109.

(12) Vancraenenbroeck, R.; Harel, Y. S.; Zheng, W.; Hofmann, H. Polymer effects modulate binding affinities in disordered proteins. *Proceedings of the National Academy of Sciences* **2019**, *116*, 19506–19512.

(13) Borgia, A.; Borgia, M. B.; Bugge, K.; Kissling, V. M.; Heidarsson, P. O.; Fernandes, C. B.; Sottini, A.; Soranno, A.; Buholzer, K. J.; Nettels, D. et al. Extreme disorder in an ultrahigh-affinity protein complex. *Nature* **2018**, *555*, 61–66.

(14) Seiffert, P.; Bugge, K.; Nygaard, M.; Haxholm, G. W.; Martinsen, J. H.; Pedersen, M. N.; Arleth, L.; Boomsma, W.; Kragelund, B. B. Orchestration of signaling by structural disorder in class 1 cytokine receptors. *Cell Communication and Signaling* **2020**, *18*, 1–30.

(15) Evans, J. S. The biomineralization proteome: protein complexity for a complex bioceramic assembly process. *Proteomics* **2019**, *19*, 1900036.

(16) Krois, A. S.; Dyson, H. J.; Wright, P. E. Long-range regulation of p53 DNA binding by its intrinsically disordered N-terminal transactivation domain. *Proceedings of the National Academy of Sciences* **2018**, *115*, E11302–E11310.

(17) Fletcher, C. M.; McGuire, A. M.; Gingras, A.-C.; Li, H.; Matsuo, H.; Sonenberg, N.; Wagner, G. 4E binding proteins inhibit the translation factor eIF4E without folded structure. *Biochemistry* **1998**, *37*, 9–15.

18

(18) Fletcher, C. M.; Wagner, G. The interaction of eIF4E with 4E-BP1 is an induced fit to a completely disordered protein. *Protein Science* **1998**, *7*, 1639–1642.

(19) Gingras, A.-C.; Raught, B.; Gygi, S. P.; Niedzwiecka, A.; Miron, M.; Burley, S. K.; Polakiewicz, R. D.; Wyslouch-Cieszynska, A.; Aebersold, R.; Sonenberg, N. Hierarchical phosphorylation of the translation inhibitor 4E-BP1. *Genes & Development* **2001**, *15*, 2852–2864.

(20) Sheu-Gruttadauria, J.; MacRae, I. J. Phase transitions in the assembly and function of human miRISC. *Cell* **2018**, *173*, 946–957.

(21) Cieplak-Rotowska, M. K.; Tarnowski, K.; Rubin, M.; Fabian, M. R.; Sonenberg, N.; Dadlez, M.; Niedzwiecka, A. Structural dynamics of the GW182 silencing domain including its RNA recognition motif (RRM) revealed by hydrogen-deuterium exchange mass spectrometry. *Journal of The American Society for Mass Spectrometry* **2018**, *29*, 158–173.

(22) Raisch, T.; Valkov, E. Regulation of the multisubunit CCR4-NOT deadenylase in the initiation of mRNA degradation. *Current Opinion in Structural Biology* **2022**, *77*, 102460.

(23) Louros, N.; Schymkowitz, J.; Rousseau, F. Mechanisms and pathology of protein misfolding and aggregation. *Nature Reviews Molecular Cell Biology* **2023**, 1–22.

(24) Chakraborty, P.; Zweckstetter, M. Role of aberrant phase separation in pathological protein aggregation. *Current Opinion in Structural Biology* **2023**, *82*, 102678.

(25) Banani, S. F.; Rice, A. M.; Peeples, W. B.; Lin, Y.; Jain, S.; Parker, R.; Rosen, M. K. Compositional control of phase-separated cellular bodies. *Cell* **2016**, *166*, 651–663.

(26) Forman-Kay, J. D.; Ditlev, J. A.; Nosella, M. L.; Lee, H. O. What are the distinguishing features and size requirements of biomolecular condensates and their implications for RNA-containing condensates? *RNA* **2022**, *28*, 36–47.

(27) Białobrzewski, M. K.; Klepka, B. P.; Michaś, A.; Cieplak-Rotowska, M. K.; Staszałek, Z.; Niedźwiecka, A. Diversity of hydrodynamic radii of intrinsically disordered proteins. *European Biophysics Journal* **2023**, *52*, 607–618.

(28) Nygaard, M.; Kragelund, B. B.; Papaleo, E.; Lindorff-Larsen, K. An efficient method for estimating the hydrodynamic radius of disordered protein conformations. *Biophysical Journal* **2017**, *113*, 550–557.

(29) Pesce, F.; Newcombe, E. A.; Seiffert, P.; Tranchant, E. E.; Olsen, J. G.; Grace, C. R.; Kragelund, B. B.; Lindorff-Larsen, K. Assessment of models for calculating the hydrodynamic radius of intrinsically disordered proteins. *Biophysical Journal* **2023**, *122*, 310–321.

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

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

(32) Mao, A. H.; Crick, S. L.; Vitalis, A.; Chicoine, C. L.; Pappu, R. V. Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. *Proceedings of the National Academy of Sciences* **2010**, *107*, 8183–8188.

(33) Różycki, B.; Kim, Y. C.; Hummer, G. SAXS ensemble refinement of ESCRT-III CHMP3 conformational transitions. *Structure* **2011**, *19*, 109–116.

(34) Ozenne, V.; Bauer, F.; Salmon, L.; Huang, J.-r.; Jensen, M. R.; Segard, S.; Bernadó, P.; Charavay, C.; Blackledge, M. Flexible-meccano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables. *Bioinformatics* **2012**, *28*, 1463–1470.

(35) Mittal, J.; Yoo, T. H.; Georgiou, G.; Truskett, T. M. Structural ensemble of an intrinsically disordered polypeptide. *The Journal of Physical Chemistry B* **2013**, *117*, 118–124.

(36) Mittal, A.; Holehouse, A. S.; Cohan, M. C.; Pappu, R. V. Sequence-to-conformation relationships of disordered regions tethered to folded domains of proteins. *Journal of Molecular Biology* **2018**, *430*, 2403–2421.

(37) Das, P.; Matysiak, S.; Mittal, J. Looking at the disordered proteins through the computational microscope. *ACS Central Science* **2018**, *4*, 534–542.

(38) Estaña, A.; Sibille, N.; Delaforge, E.; Vaisset, M.; Cortés, J.; Bernadó, P. Realistic ensemble models of intrinsically disordered proteins using a structure-encoding coil database. *Structure* **2019**, *27*, 381–391.

(39) Baul, U.; Chakraborty, D.; Mugnai, M. L.; Straub, J. E.; Thirumalai, D. Sequence effects on size, shape, and structural heterogeneity in intrinsically disordered proteins. *The Journal of Physical Chemistry B* **2019**, *123*, 3462–3474.

(40) de la Torre, J. G.; Cifre, J. H. Hydrodynamic properties of biomacromolecules and macromolecular complexes: concepts and methods. A tutorial mini-review. *Journal of Molecular Biology* **2020**, *432*, 2930–2948.

(41) Gomes, G.-N. W.; Krzeminski, M.; Namini, A.; Martin, E. W.; Mittag, T.; Head-Gordon, T.; Forman-Kay, J. D.; Gradinaru, C. C. Conformational ensembles of an intrinsically disordered protein consistent with NMR, SAXS, and single-molecule FRET. *Journal of the American Chemical Society* **2020**, *142*, 15697–15710.

(42) Zimm, B. H. Chain molecule hydrodynamics by the Monte-Carlo method and the validity of the Kirkwood-Riseman approximation. *Macromolecules* **1980**, *13*, 592–602.

(43) Zimm, B. H. Sedimentation of asymmetric elastic dumbbells and the rigid-body approximation in the hydrodynamics of chains. *Macromolecules* **1982**, *15*, 520–525.

(44) Schmidt, R. R.; Cifre, J. H.; de la Torre, J. G. Translational diffusion coefficients of macromolecules. *Eururopean Physical Journal E* **2012**, *35*, 130.

(45) de la Torre, J. G. *Analytical Ultracentrifugation. Instrumentation, Software, and Applications*; Springer, 2016; pp 195–217.

(46) Fixman, M. Inclusion of hydrodynamic interaction in polymer dynamical simulations. *Macromolecules* **1981**, *14*, 1710–1717.

(47) Fixman, M. Variational bounds for polymer transport coefficients. *Journal of Chemical Physics* **1983**, *78*, 1588–1593.

(48) Happel, J.; Brenner, H. *Low Reynolds Number Hydrodynamics*; Noordhoff, Leyden, 1973.

(49) Cichocki, B.; Ekiel-Jeżewska, M. L.; Wajnryb, E. Communication: Translational Brownian motion for particles of arbitrary shape. *Journal of Chemical Physics* **2012**, *136*.

(50) Wegener, W. A. Bead models of segmentally flexible macromolecules. *Journal of Chemical Physics* **1982**, *76*, 6425–6430.

(51) Harvey, S. C.; Mellado, P.; García de la Torre, J. Hydrodynamic resistance and diffusion coefficients of segmentally flexible macromolecules with two subunits. *Journal of Chemical Physics* **1983**, *78*, 2081–2090.

(52) Wegener, W. A. Center of diffusion of flexible macromolecules. *Macromolecules* **1985**, *18*, 2522–2530.

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

(54) Karplus, M.; Petsko, G. A. Molecular dynamics simulations in biology. *Nature* **1990**, *347*, 631–639.

(55) Doi, M.; Edwards, S. F. *The Theory of Polymer Dynamics*; Oxford University Press, 1988; Vol. 73.

(56) Prakash, J. R. The kinetic theory of dilute solutions of flexible polymers: Hydrodynamic interaction. *Rheology Series* **1999**, *8*, 467–517.

(57) Szymczak, P.; Cieplak, M. Hydrodynamic effects in proteins. *Journal of Physics: Condensed Matter* **2011**, *23*, 033102.

(58) Skolnick, J. Perspective: On the importance of hydrodynamic interactions in the subcellular dynamics of macromolecules. *The Journal of Chemical Physics* **2016**, *145*.

(59) Frenkel, D.; Smit, B. *Understanding molecular simulation: from algorithms to applications*; Elsevier, 2001; Vol. 1.

(60) Rouse, P. E. A Theory of the Linear Viscoelastic Properties of Dilute Solutions of Coiling Polymers. *The Journal of Chemical Physics* **1953**, *21*, 1272–1280.

(61) Zimm, B. H. Dynamics of Polymer Molecules in Dilute Solution: Viscoelasticity, Flow Birefringence and Dielectric Loss. *Journal of Chemical Physics* **1956**, *24*, 269–278.

(62) Kirkwood, J. G.; Riseman, J. The intrinsic viscosities and diffusion constants of flexible macromolecules in solution. *The Journal of Chemical Physics* **1948**, *16*, 565–573.

(63) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589.

(64) Ruff, K. M.; Pappu, R. V. AlphaFold and implications for intrinsically disordered proteins. *Journal of Molecular Biology* **2021**, *433*, 167208.

(65) Murphy, L. R.; Matubayasi, N.; Payne, V. A.; Levy, R. M. Protein hydration and unfolding–insights from experimental partial specific volumes and unfolded protein models. *Folding and Design* **1998**, *3*, 105–118.

(66) Jones, D. T.; Cozzetto, D. DISOPRED3: precise disordered region predictions with annotated protein-binding activity. *Bioinformatics* **2015**, *31*, 857–863.

(67) Choi, Y.; Agarwal, S.; Deane, C. M. How long is a piece of loop? *PeerJ* **2013**, *1*, e1.

(68) Longsworth, L. Diffusion measurements, at 25°C, of aqueous solutions of amino acids, peptides and sugars. *Journal of the American Chemical Society* **1953**, *75*, 5705–5709.

(69) Zuk, P.; Wajnryb, E.; Mizerski, K.; Szymczak, P. Rotne–Prager–Yamakawa approximation for different-sized particles in application to macromolecular bead models. *Journal of Fluid Mechanics* **2014**, *741*, R5.

(70) Cox, R. The motion of long slender bodies in a viscous fluid Part 1. General theory. *Journal of Fluid Mechanics* **1970**, *44*, 791–810.

(71) Johnson, R. E.; Wu, T. Y. Hydromechanics of low-Reynolds-number flow. Part 5. Motion of a slender torus. *Journal of Fluid Mechanics* **1979**, *95*, 263–277.

(72) Majumdar, S. R.; O'Neill, M. E. On axisymmetric stokes flow past a torus. *Zeitschrift für Angewandte Mathematik und Physik ZAMP* **1977**, *28*, 541–550.

(73) Waszkiewicz, R.; Szymczak, P.; Lisicki, M. Stability of sedimenting flexible loops. *Journal of Fluid Mechanics* **2021**, *919*, A14.

(74) Kirkwood, J. G. The general theory of irreversible processes in solutions of macromolecules. *Journal of Polymer Science* **1954**, *12*, 1–14.

(75) Liu, B.; Dünweg, B. Translational diffusion of polymer chains with excluded volume and hydrodynamic interactions by Brownian dynamics simulation. *Journal of Chemical Physics* **2003**, *118*, 8061–8072.

24

(76) Vovk, A.; Zilman, A. Effects of Sequence Composition, Patterning and Hydrodynamics on the Conformation and Dynamics of Intrinsically Disordered Proteins. *International Journal of Molecular Sciences* **2023**, *24*, 1444.

(77) Kim, S.; Karrila, S. J. *Microhydrodynamics: Principles and Selected Applications*; Butterworth-Heinemann: London, 1991.

(78) Mazur, P.; van Saarloos, W. Many-sphere hydrodynamic interactions and mobilities in a suspension. *Physica A* **1982**, *115*, 21–57.

(79) Cichocki, B.; Felderhof, B. U.; Hinsen, K.; Wajnryb, E.; Bławzdziewicz, J. Friction and mobility of many spheres in Stokes flow. *Journal of Chemical Physics* **1994**, *100*, 3780–3790.

(80) Rotne, J.; Prager, S. Variational treatment of hydrodynamic interaction in polymers. *Journal of Chemical Physics* **1969**, *50*, 4831–4837.

(81) Yamakawa, H. Transport properties of polymer chains in dilute solution: Hydrodynamic interaction. *Jorunal of Chemical Physics* **1970**, *53*, 436–443.

(82) Zuk, P.; Cichocki, B.; Szymczak, P. GRPY: An Accurate Bead Method for Calculation of Hydrodynamic Properties of Rigid Biomacromolecules. *Biophysical Journal* **2018**, *115*, 782–800.

(83) Waszkiewicz, R.; Bartczak, M.; Kolasa, K.; Lisicki, M. Pychastic: Precise Brownian dynamics using Taylor-Itō integrators in Python. *SciPost Physics Codebases* **2023**, 11.

(84) de Val, N.; Declercq, J.-P.; Lim, C. K.; Crichton, R. R. Structural analysis of haemin demetallation by L-chain apoferritins. *Journal of Inorganic Biochemistry* **2012**, *112*, 77–84.

(85) McCubbin, W. D.; Kay, C. M.; Lane, B. G. Hydrodynamic and optical properties of the wheat germ Em protein. *Canadian Journal of Biochemistry and Cell Biology* **1985**, *63*, 803–811.

25

(86) Guez, V.; Nair, S.; Chaffotte, A.; Bedouelle, H. The anticodon-binding domain of tyrosyl-tRNA synthetase: state of folding and origin of the crystallographic disorder. *Biochemistry* **2000**, *39*, 1739–1747.

(87) Bouvier, M.; Stafford, W. F. Probing the three-dimensional structure of human calreticulin. *Biochemistry* **2000**, *39*, 14950–14959.

(88) Danielsson, J.; Jarvet, J.; Damberg, P.; Gräslund, A. Translational diffusion measured by PFG-NMR on full length and fragments of the Alzheimer Aβ (1–40) peptide. Determination of hydrodynamic radii of random coil peptides of varying length. *Magnetic Resonance in Chemistry* **2002**, *40*, S89–S97.

(89) Karlin, D.; Longhi, S.; Receveur, V.; Canard, B. The N-terminal domain of the phosphoprotein of morbilliviruses belongs to the natively unfolded class of proteins. *Virology* **2002**, *296*, 251–262.

(90) Longhi, S.; Receveur-Bréchot, V.; Karlin, D.; Johansson, K.; Darbon, H.; Bhella, D.; Yeo, R.; Finet, S.; Canard, B. The C-terminal domain of the measles virus nucleoprotein is intrinsically disordered and folds upon binding to the C-terminal moiety of the phosphoprotein. *Journal of Biological Chemistry* **2003**, *278*, 18638–18648.

(91) Sánchez-Puig, N.; Veprintsev, D. B.; Fersht, A. R. Human full-length Securin is a natively unfolded protein. *Protein Science* **2005**, *14*, 1410–1418.

(92) Sánchez-Puig, N.; Veprintsev, D. B.; Fersht, A. R. Binding of natively unfolded HIF-1α ODD domain to p53. *Molecular Cell* **2005**, *17*, 11–21.

(93) Khaymina, S. S.; Kenney, J. M.; Schroeter, M. M.; Chalovich, J. M. Fesselin is a natively unfolded protein. *Journal of Proteome Research* **2007**, *6*, 3648–3654.

(94) Zhang, X.; Perugini, M. A.; Yao, S.; Adda, C. G.; Murphy, V. J.; Low, A.; Anders, R. F.; Norton, R. S. Solution conformation, backbone dynamics and lipid interactions of the intrinsically unstructured malaria surface protein MSP2. *Journal of Molecular Biology* **2008**, *379*, 105–121.

(95) Mittag, T.; Orlicky, S.; Choy, W.-Y.; Tang, X.; Lin, H.; Sicheri, F.; Kay, L. E.; Tyers, M.; Forman-Kay, J. D. Dynamic equilibrium engagement of a polyvalent ligand with a single-site receptor. *Proceedings of the National Academy of Sciences* **2008**, *105*, 17772–17777.

(96) Paz, A.; Zeev-Ben-Mordehai, T.; Lundqvist, M.; Sherman, E.; Mylonas, E.; Weiner, L.; Haran, G.; Svergun, D. I.; Mulder, F. A.; Sussman, J. L. et al. Biophysical characterization of the unstructured cytoplasmic domain of the human neuronal adhesion protein neuroligin 3. *Biophysical Journal* **2008**, *95*, 1928–1944.

(97) Habchi, J.; Mamelli, L.; Darbon, H.; Longhi, S. Structural disorder within Henipavirus nucleoprotein and phosphoprotein: from predictions to experimental assessment. *PLoS One* **2010**, *5*, e11684.

(98) Choi, U. B.; McCann, J. J.; Weninger, K. R.; Bowen, M. E. Beyond the random coil: stochastic conformational switching in intrinsically disordered proteins. *Structure* **2011**, *19*, 566–576.

(99) Perez, R. B.; Tischer, A.; Auton, M.; Whitten, S. T. Alanine and proline content modulate global sensitivity to discrete perturbations in disordered proteins. *Proteins: Structure, Function, and Bioinformatics* **2014**, *82*, 3373–3384.

(100) Yarawsky, A. E.; English, L. R.; Whitten, S. T.; Herr, A. B. The proline/glycine-rich region of the biofilm adhesion protein Aap forms an extended stalk that resists compaction. *Journal of Molecular Biology* **2017**, *429*, 261–279.

(101) Poznar, M.; Hołubowicz, R.; Wojtas, M.; Gapiński, J.; Banachowicz, E.; Patkowski, A.; Ożyhar, A.; Dobryszycki, P. Structural properties of the intrinsically disordered, multiple calcium ion-binding otolith matrix macromolecule-64 (OMM-64). *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* **2017**, *1865*, 1358–1371.

(102) Więch, A.; Rowińska-Żyrek, M.; Wątły, J.; Czarnota, A.; Hołubowicz, R.; Szewczuk, Z.; Ożyhar, A.; Orłowski, M. The intrinsically disordered C-terminal F

domain of the ecdysteroid receptor from Aedes aegypti exhibits metal ion-binding ability. *The Journal of Steroid Biochemistry and Molecular Biology* **2019**, *186*, 42–55.

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

(104) Le Guillou, J.; Zinn-Justin, J. Critical exponents for the n-vector model in three dimensions from field theory. *Physical Review Letters* **1977**, *39*, 95.

(105) Ding, F.; Jha, R. K.; Dokholyan, N. V. Scaling behavior and structure of denatured proteins. *Structure* **2005**, *13*, 1047–1054.

(106) Philo, J. S. SEDNTERP: a calculation and database utility to aid interpretation of analytical ultracentrifugation and light scattering data. *European Biophysics Journal* **2023**, 1–34.

(107) Wilkins, D. K.; Grimshaw, S. B.; Receveur, V.; Dobson, C. M.; Jones, J. A.; Smith, L. J. Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. *Biochemistry* **1999**, *38*, 16424–16431.

(108) Waszkiewicz, R. GitHub: glm-mda-diffusion. https://github.com/RadostW/glm-mda-diffusion, 2023; Accessed: 2023-12-25.