Minimum dissipation approximation: A fast algorithm for the prediction of diffusive properties of intrinsically disordered proteins

Radost Waszkiewicz,[†,¶] Agnieszka Michaś,[‡,¶] Michał K. Białobrzewski,[‡] Barbara Klepka,[‡] Maja Cieplak-Rotowska,[‡] Zuzanna Staszałek,[‡] Bogdan Cichocki,[†] Maciej Lisicki,[†] Piotr Szymczak,[*,†] and Anna Niedźwiecka[*,‡]

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

‡*Institute of Physics, Polish Academy of Sciences, Al. Lotników 32/46, 02-668 Warsaw, Poland*

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

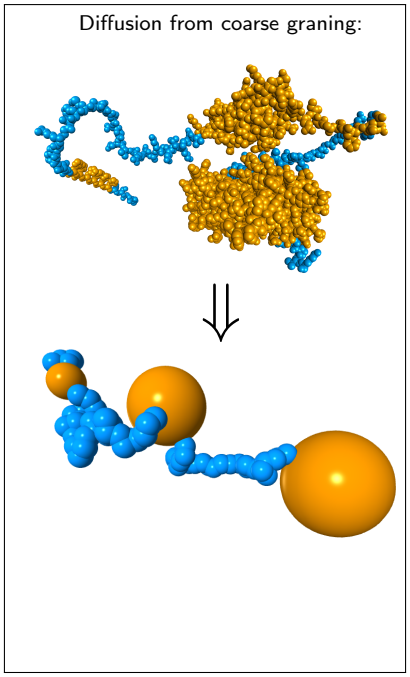
E-mail: piotrek@fuw.edu.pl; annan@ifpan.edu.pl

Abstract

The diffusion coefficients of globular and completely disordered proteins can be predicted with a high degree of precision based solely on the mass of the protein. However, this approach fails when the disordered protein contains structured domains, in which case a more involved method is needed. In this study, we elucidate a rapid predictive methodology for estimating the diffusion coefficients of intrinsically disordered proteins (IDPs), regardless of the presence of structured domains. The methodology takes advantage of expedited conformational sampling based on self-avoiding random walks and includes hydrodynamic interactions between coarse-grained protein subunits, which are modeled using the generalized Rotne-Prager-Yamakawa approximation. To estimate the hydrodynamic radius, we rely on the minimal dissipation algorithm proposed by Cichocki et al.

When contrasting the precision of the proposed technique with the Kirkwood-Riseman approximation and other phenomenological approaches, we demonstrate, using a selection of measured hydrodynamic radii of IDPs, that our methodology provides more accurate predictions without imposing considerable computational burdens. We anticipate that our technique may prove valuable for predicting the hydrodynamic properties of fully disordered and multidomain proteins. Additionally, it can be beneficial for investigating proteins in which experimental evidence indicates a departure from 'random coil' behavior.

TOC Graphic



Keywords

words, words

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 fragments (IDRs) in proteins differs significantly among the kingdoms of life, ranging from 2% in archaea to 33% in eukaryotes.[3] The IDP molecules display different degree of structural disorder. Their chains can encompass either several folded globular domains or super-secondary 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 among their fragments. Instead, their equilibrium properties can be described through a broad family of rapidly inter-converting conformers, posing a challenge for analysis, particularly in the context of long chains.

The average geometric properties of IDPs, including shape and size, are determined by the equilibrium ensemble of conformational states. This equilibrium state is intricately influenced by environmental conditions,[4] such as temperature,[5] ionic strength,[6,7] osmolality,[8] crowding,[9] post-translational modifications,[10] and the presence of specific molecular binding partners.[11] These proteins form transient or more stable non-covalent complexes with these binding partners, introducing a non-trivial dependence on environmental factors.

Because the shape and availability of the binding sites necessary for IDP interaction 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[10,12] signaling,[2,13] or extracellular biomineralization.[14] It is the different conformations of these flexible proteins, which enable IDPs to perform their multiple functions.[1] In particular, it is worth emphasising the important roles of IDPs in health and disease, e.g., the role of the p53 protein as a tumour suppressor,[15] mutations of which are often responsible for human cancers, the function of 4E-BPs in the inhibition of eukaryotic translation initiation,[10,16–18] the significance of GW182 protein in the recruitment of the multi-protein machinery necessary for microRNA-mediated gene silencing,[19–21] or

the importance of Tau, Fus, and α-synuclein proteins in neurodegenerative diseases.[22,23] Since the elastic properties of these biomolecules are responsible for proper functioning of IDPs in the cellular context, i.e. for the association of complexes and formation of biomolecular condensates via liquid-liquid phase separation such as, e.g., RNA-processing membraneless organelles[24,25] 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[26]). However, due to the distinct limitations of each experimental approach, ongoing research aims to devise phenomenological methods for calculating R_h. These methods may involve deriving R_h from the radius of gyration determined through small-angle X-ray scattering (SAXS)[27,28] or leveraging the conformational backbone propensity of the IDPs.[29,30]

Simultaneously, significant effort is being invested in developing numerical models that extract the characteristics of IDPs from conformational ensembles obtained through molecular dynamics simulations or energy minimization algorithms.[31–40] However, the presence of molecular flexibility 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 molecule. These factors prohibit the 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 the macromolecule to be 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 the equilibrium ensemble.[41–44] 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), since 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 the 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 for both generating conformational ensembles of IDPs and calculating their hydrodynamic properties. This method enables a swift estimation of the hydrodynamic radius (R_h) 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 from 43 proteins. The dataset includes both literature data and new R_h values measured for our own IDPs using fluorescence correlation spectroscopy (FCS) under mild conditions (see Supporting Information).

Calculating the diffusion coefficients of IDPs requires going beyond traditional models of protein diffusion that approximate their shape by a prolate or oblate ellipsoid,[45] but we start our reasoning from an even simpler classical result. The diffusion coefficient in a dilute suspension of microscopic hard spheres of radius R suspended in a fluid of temperature T and viscosity η is found from the Stokes-Einstein relationship to be $D = k_BT/6\pi\eta R$. This leads to the definition of a hydrodynamic radius of a molecule, $R_h = k_BT/6\pi\eta D$. Thus, R_h corresponds to the size of a solid sphere of the same translational diffusion coefficient as the given molecule under the same buffer conditions.

An important observation due to Fixman[46,47] is that the diffusion coefficient in a flexible macromolecule is time-dependent, with a well-defined short- and long-time limits. The disparity between the two is attributed to the effects associated with the relaxation of the internal coordinates of the molecule as well as the rotation of the macromolecule as a whole.[46,48,49] Due to the positivity of the dissipation rate in the system, 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 fluorescence correlation spectroscopy, ultracentrifugation, or dynamic light scattering. Conversely, the long-time diffusivity is independent of the choice of the reference point. Additional point to keep

in mind is that the value of the short-time diffusion coefficient depends on the choice of the point that we track[50–53].

The methods of predicting the diffusion coefficient can be broadly split into three categories: atomistic, phenomenological, and coarse-grained. For small proteins, high-resolution (atomistic) methods of molecular dynamics (MD) 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 and thermalization pose significant challenges.[55] 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[56] 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,[57] 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.[28] 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[58] R_h and R_g share the same scaling relationship with N as long as the pair-displacement distribution converges to a Gaussian (under appropriate scaling) for large N.

Finally, coarse-grained models, like our method, employ larger units (typically amino acid residues) as building blocks for the structure prediction scheme, along with approximate interaction potentials between subunits, to simulate the equilibrium 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 involves addressing 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 and Zimm demonstrate that even the most basic method for approximating configurations must take into account excluded volume interactions. One available software that can accommodate excluded volume interactions for a completely disordered chain is Flexible Meccano (FM).[33] In addition to volume exclusion, it also considers the distribution of Ramachandran angles (determined from crystallized proteins) when sampling conformations. However, FM has one important deficiency - it treats the entire chain as disordered, so it cannot be used to model proteins that possess both globular and disordered segments, which are, in fact, much more common than fully disordered chains. Unfortunately, FM is a closed-source software and has a proprietary license that prevents necessary modifications to accommodate ordered 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 minimal influence on the estimations of R_h. It it important to recall that for amino acids separated by a distance r, 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.

Guided by these considerations, we have implemented the simplest extension of Zimm's chain — a globule-linker model (GLM), designed to represent IDPs with both disordered and globular domains. In the model, we represented the protein as an assembly of spheres of variable sizes. Disordered segments of length N were modeled as N identical spheres, each with a diameter equal to the C_α-C_α distance, while structured domains were represented by a single sphere each. The domains we encountered typically have a broadly spheroidal shape and can be sufficiently described with just a single parameter: their mass. The size of a sphere representing a structured domain was then computed from its mass, m, as $R_h = \sqrt[3]{3m/4\pi\rho_{\text{globular}}} + a_{\text{hydration}}$ with $\rho_{globular} = 0.52$ Da / Å^3[59] and hydration taken to be thickness 3 Å (single layer shell). Within this approach, data on the domain boundaries in the protein sequence is sufficient for constructing an appropriate bead approximation of the protein.

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.[60] 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.[61]

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 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 sampled conformations to a hydrodynamic model by increasing bead sizes in the disordered fragments to $R_{disordered} = 4.2$ Å, corresponding to median

value for all aminoacids.[62] This results in neighbouring beads having a significant overlap. Note that the $R_{disordered}$ value has only a minor impact on the final results of the model, as the hydrodynamic radius of long slender filaments depends logarithmically on their thickness.[63–66]

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.[49]

Within the first approach,[67] the hydrodynamic radius of a macromolecule is approximated by

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

where N is the total number of beads in the IDP model, a is the radius of bead i, $r_{ij} = r_j - r_i$ 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, $\mathbf{r}_c = \frac{1}{N}\sum_{i=1}^{N}\mathbf{r}_i$. Note that the geometric center fluctuates as the molecule's shape evolves and does not correspond to any fixed position within it.

A simplified form of the Kirkwood formula is often used[38,68,69]

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

In this form, 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 a long-time diffusion coefficient is provided by the MDA method of Cichocki et al.[49] MDA approximates R_h as

$$R_h^{\mathrm{MDA}} = \frac{1}{6\pi\eta}\sum_{i,j}\left(\mathbf{A}^{-1}\right)_{ij} \tag{3}$$

where $\mathbf{A}$ is a matrix indexed by the bead labels (i, j), $A_{ij} = \frac{1}{3}\mathrm{Tr}\left\langle\boldsymbol{\mu}_{ij}\right\rangle$, the elements of which are expressed through the so-called mobility matrix μ[48] which links the velocities

of the beads with the forces acting on them, according to

$$\mathbf{U}_i = \sum_j \mu_{ij} \mathbf{F}_j, \tag{4}$$

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. For the calculation of the mobility matrix, we use the generalized Rotne-Prager approximation[70]. This approximation is now also available as a Python package `pygrpy`.[71]

The MDA approximation 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 $x_i = \sum_j (\mathbf{A}^{-1})_{ij} \,/\, \sum_{i,j} (\mathbf{A}^{-1})_{ij}$. 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 approximation turns out to be more robust when dealing with large differences in the sizes of beads used to model constituent parts of the macromolecule, as 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 R_h, which resulted in four different theoretical approaches, the predictions of which were then compared with available experimental data. The experimental benchmark set was compiled from the new FCS measurements and the values of R_h selected from literature. The selected proteins had sequences that could be unambiguously identified in the literature or in the UniProtKB database, were measured at well defined, comparable, 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[28,61,72–88]

All theoretical and experimental results are included in Table 1, while figure 4 allows us to visually compare the deviations between theory and experiment. Additionally we

provide power law fits to the presented dataset and power law fits of Marsh et al.[89] as benchmarks for comparison of the prediction accuracy.

We present a comparison of accuracy of predictions of previous and new models in Table 1 under six metrics: square root of mean square deviation (RMSD), square root of mean square relative deviation (RMSRD), Pearson's R^2, Pearson's R^2 adjusted for fitting parameters, 75% quantile of relative error Q_{75}^{RE}, and 75% quantile of absolute error Q_{75}^{AE}. We have chosen to test the relative deviations as well in order to reduce undue weight given to the new, very long sequences in our dataset. Similarly, outlier-robust metrics of the 75% error quantile were included to reduce the impact of a single sequence misprediction on the final comparisons.

In all of these metrics, the MDA+GLM approach performs the best, and notably, it outperforms the power law fit. Furthermore, it is the only model that performs better than the power law baseline in any of the evaluation metrics.

We have presented a simple, first principles model for the prediction of R_h without any fitting parameters and achieved favourable comparison both with previously available data and a set of newly measured proteins. 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 molecular-dynamics-based conformer generation methods which require supercomputers and take many days.

The presented benchmark set of 43 proteins represents 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 compared to, for example, the 32 proteins presented in Marsh & Forman-Kay.[89] This diversity allows for more reliable testing of theoretical models. The appropriate treatment of experimental uncertainties helps establish boundaries for what should be considered a good model ($R^2 \sim 0.6$) or an excellent model (R^2 $0.7 - 0.8$), while values of $R^2 > 0.85$ indicate overfitting, given the current number and precision of measurements.

Further improvements are, of course, possible. In particular, as noted earlier, R_h depends on buffer conditions and temperature,[5–7] which our simplistic model is unable

to capture. However, our results clearly demonstrate that the relatively simple globule-linker model for conformational ensemble construction, in combination with the minimum dissipation approximation, should serve as the starting point for developing further phenomenological corrections. These corrections could incorporate factors such as charge or amino acid residue composition. 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

The authors thank ...

The work of RW and ML was supported by the National Science Centre of Poland grant Sonata to ML no. 2018/31/D/ST3/02408.

Supporting Information Available

Supplementary information

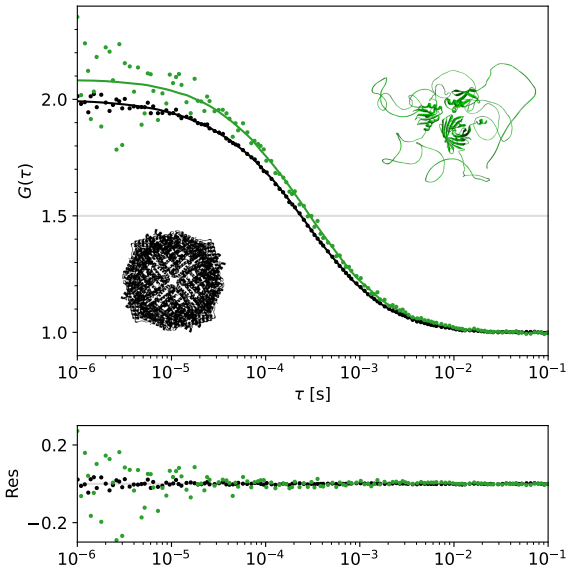


Figure 1: FCS data fitting procedure sample (apoferritin $N = 4200$ – black, His-SUMO-GW182SD-mCherry $N = 809$ – green). Apoferritin conformation from crystalographic data and for His-SUMO-GW182SD-mCherry conformation based on AlphaFold for illustration purposes only (preserving relative size). Measured R_h for appoferritin 65.5 ± 6.5 Å is consistent with theoretical 63 ± 3 Å.

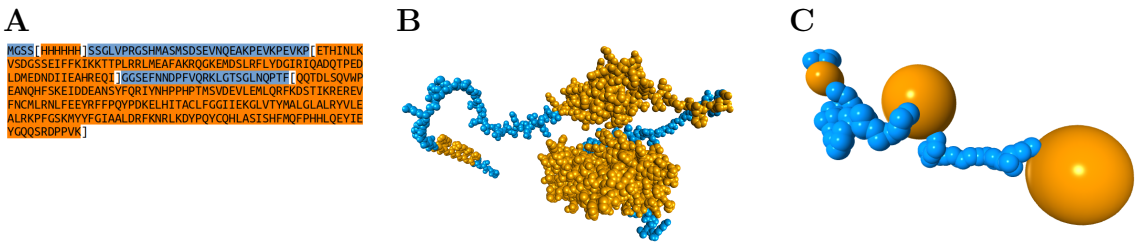


Figure 2: Mental model of coarse graining scheme used. **A)** Sequence of H_6-SUMO-CNOT1(800-999) with highlighted ordered and disordered segments **B)** Representative, full atom (beads with van der Waals radii) configuration generated by AlphaFold (for visualisation purposes only). Ordered clusters (marked orange) form dense blobs connected with linkers (blue). **C)** Visualisation of representative configuration generated using SARW method where beads are displayed with hydrodynamic radii. Notice large overlaps of neighbouring beads in disordered fragments.

Table 1: Comparison of error statistics of various models.

model	FP	RMSD[Å]	RMSRD[%]	R^2	R^2_{adj}	Q^{RE}_{75}[%]	Q^{AE}_{75}[Å]
power law	2	8.46	24.80	0.59	0.56	26.08	9.63
MDA + SARW	0	7.09∗	18.15∗	0.71∗	0.71∗	22.51∗	6.80∗
MDA + SARW(ND)	0	9.48	28.02	0.48	0.48	29.31	11.88
KR + SARW	0	12.82	34.69	0.05	0.05	42.95	17.59
KR + SARW(ND)	0	9.25	27.31	0.50	0.50	29.44	11.11
random coil	1	9.60	27.71	0.47	0.45	33.69	10.28
Marsh *et al.* eq. 5[89]	2	12.01	36.94	0.16	0.12	39.51	14.37
Marsh *et al.* eq. 6[89]	7	22.90	50.78	−2.05	−2.66	58.32	19.59
Tomasso *et al.* eq. 6[29]	3	17.25	49.09	−0.72	−0.86	59.54	20.62

FP – fitted parameters, ND – no domain information,
Q_{75} – 75% quantile, ∗ – better than power law.

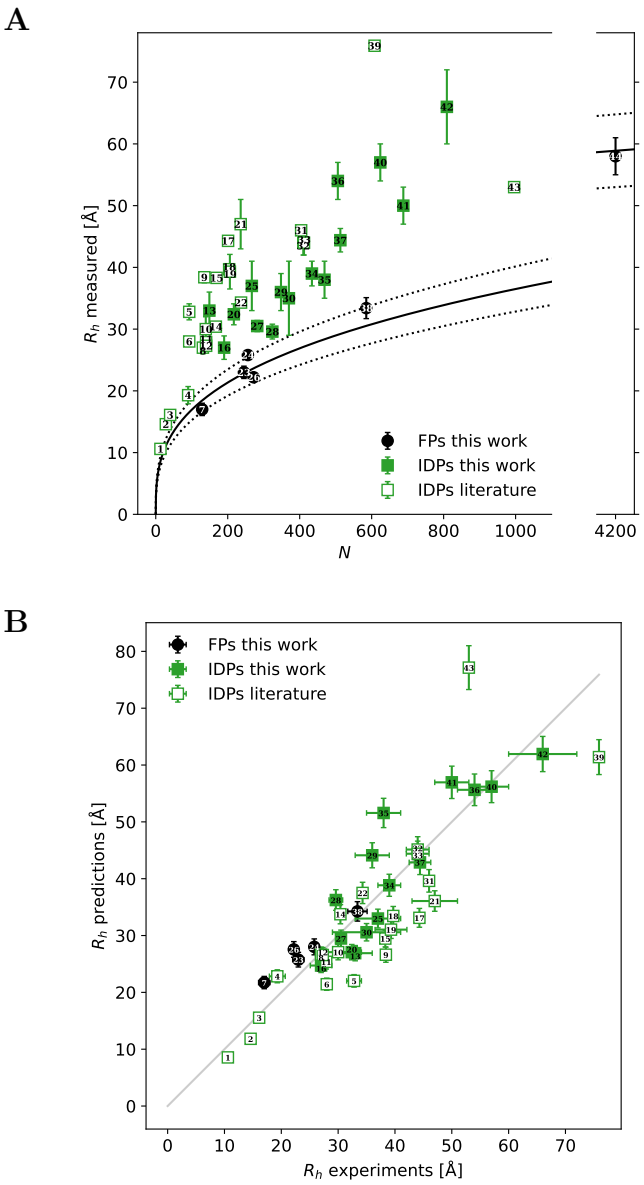


Figure 3: **A)** Raw experimental data plotted against number of aminoacid residues N together with power-law curve for folded proteins. **B)** Direct comparison of the measured and predicted values for all of the proteins modelled using the MDA+SARW technique. Numbers label individual proteins as in the table.

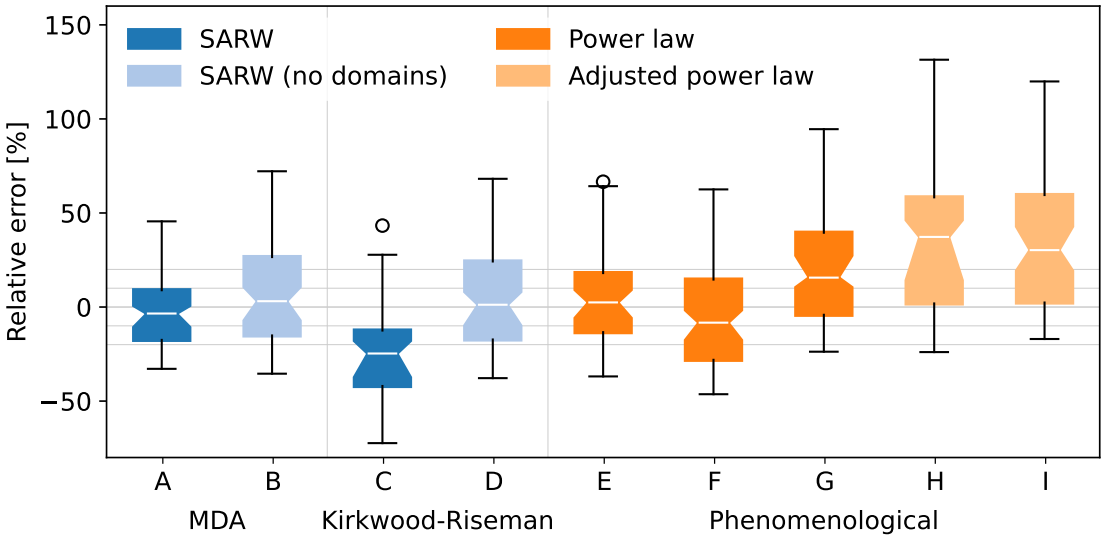


Figure 4: Comparison of different methods of R_h estimation. The minimum dissiaption approximation with SARW ensemble generation (**A**) performs best on the used proteins with standard error of 18.15% or 7.09 Å (compared to 24.80% and 8.46 Å for a simple power law). Methods based on Kirkwood-Riesman R_h esitmation (**C,D**) typically underestimate hydrodynamic size of the moleclue. Power law fits with two free parameters (**E**) or one free parameter (**F**) evaluated using leave-one-out cross validation are compared with power law reported by Marsh and Forman-Kay[89] (**G**) and a more complex model of Marsh and Forman-Kay[89] (**H**) which takes into account total charge of the molecule as well as model of Tomasso et al.[29] (**I**) which takes into account polyproline II propensities. Methods with no knowledge about the presence of the domains in the IDP (**B,C**) significantly overestimate the hydrodynamic size of the molecule. Domain data can be incorporated into our ensemble generation engine (**A,B**) leading to more accurate estimates of R_h. It is worth noting that experimental uncertainty (estimated to be 7% or 2.9 Å) also contributes to the errors presented above.

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) 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.

(5) 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.

(6) 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.

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

(8) 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.

(9) 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.

(10) 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.

(11) 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.

(12) 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.

(13) 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.

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

(15) 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.

(16) 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.

(17) 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.

(18) 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.

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

(20) 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.

(21) 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.

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

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

(24) 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.

(25) 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.

(26) Bialobrzewski, M. K.; Klepka, B. P.; Michas, A.; Cieplak-Rotowska, M. K.; Staszalek, Z.; Niedzwiecka, A. Diversity of hydrodynamic radii of intrinsically disordered proteins. *bioRxiv* **2023**, 2023–06.

(27) 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.

(28) 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.

(29) 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.

(30) 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.

(31) 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.

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

(33) 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.

(34) 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.

(35) 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.

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

(37) 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.

(38) 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.

(39) 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.

(40) 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.

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

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

(43) Schmidt, R. R.; Cifre, J. H.; de la Torre, J. G. Translational diffusion coefficients of macromolecules. *Eur. Phys. J. E* **2012**, *35*, 130.

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

(45) Lebowitz, J.; Lewis, M. S.; Schuck, P. Modern analytical ultracentrifugation in protein science: A tutorial review. *Protein Science* **2002**, *11*, 2067–2079.

(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. *J. Chem. Phys.* **1983**, *78*, 1588–1593.

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

(49) Cichocki, B.; Rubin, M.; Niedzwiecka, A.; Szymczak, P. Diffusion coefficients of elastic macromolecules. *J. Fluid Mech.* **2019**, *878*, R3.

(50) Wegener, W. A. Bead models of segmentally flexible macromolecules. *J. Chem. Phys.* **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. *J. Chem. Phys.* **1983**, *78*, 2081–2090.

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

(53) Cichocki, B.; Ekiel-Jeżewska, M. L.; Wajnryb, E. Communication: Translational Brownian motion for particles of arbitrary shape. *J. Chem. Phys.* **2012**, *136*.

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

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

(56) 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.

(57) Zimm, B. H. Dynamics of Polymer Molecules in Dilute Solution: Viscoelasticity, Flow Birefringence and Dielectric Loss. *J. Chem Phys* **1956**, *24*, 269–278.

(58) 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.

(59) 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.

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

(61) 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.

(62) 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.

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

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

(65) 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.

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

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

(68) Liu, B.; Dünweg, B. Translational diffusion of polymer chains with excluded volume and hydrodynamic interactions by Brownian dynamics simulation. *J. Chem. Phys.* **2003**, *118*, 8061–8072.

(69) 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.

(70) 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.

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

(72) 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.

(73) 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.

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

(75) 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.

(76) 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.

(77) 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.

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

(79) 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.

(80) 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.

(81) 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.

(82) 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.

(83) 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.

(84) 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.

(85) 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.

(86) 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.

(87) 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.

(88) 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.

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