Diversity of hydrodynamic radii of intrinsically disordered proteins

Michał K. Białobrzewski[1] · Barbara P. Klepka[1] · Agnieszka Michaś[1] · Maja K. Cieplak-Rotowska[1,2,3] · Zuzanna Staszałek[1] · Anna Niedźwiecka[1]

Abstract
Intrinsically disordered proteins (IDPs) form an important class of biomolecules regulating biological processes in higher organisms. The lack of a fixed spatial structure facilitates them to perform their regulatory functions and allows the efficiency of biochemical reactions to be controlled by temperature and the cellular environment. From the biophysical point of view, IDPs are biopolymers with a broad configuration state space and their actual conformation depends on non-covalent interactions of its amino acid side chain groups at given temperature and chemical conditions. Thus, the hydrodynamic radius (R_h) of an IDP of a given polymer length (N) is a sequence- and environment-dependent variable. We have reviewed the literature values of hydrodynamic radii of IDPs determined experimentally by SEC, AUC, PFG NMR, DLS, and FCS, and complement them with our FCS results obtained for a series of protein fragments involved in the regulation of human gene expression. The data collected herein show that the values of hydrodynamic radii of IDPs can span the full space between the folded globular and denatured proteins in the $R_h(N)$ diagram.

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

[2] Division of Biophysics, Institute of Experimental Physics, Faculty of Physics, University of Warsaw, Pasteura 5, PL-02093 Warsaw, Poland

[3] Present Address: The International Institute of Molecular Mechanisms and Machines, Polish Academy of Sciences, Flisa 6, PL-02247 Warsaw, Poland

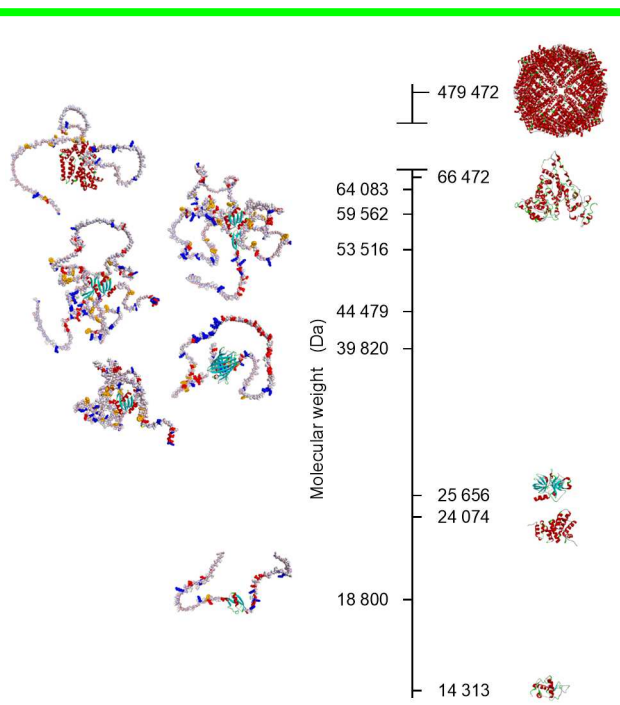

Fig. 1 Example putative conformations of the IDPs studied in this work (left panel) ordered top–down according to their molecular weights, CNOT1 M long, SUMO-GW182 SD, SUMO-GW182 SD10, PARN C-mCherry, GW182 SD10, and SUMO-GW182 SD peptide, predicted by AlphaFold 2 (Jumper et al. 2021; Mirdita et al. 2022) in comparison with the globular proteins (right panel)

R_h of an IDP is a variable and not a constant value describing the hydrodynamic properties of a biomolecule, contrary to, e.g., chemical dyes with a fixed spatial structure. Consequently, the diffusion coefficient (D) of an IDP depends on all the chemical factors that can influence the molecule dimensions (Moses et al. 2020) and on temperature (T) in a more intricate way through R_h

$$D(T,\mu) = \frac{kT}{6\pi\eta R_h(T,\mu)}, \quad (1)$$

where μ is the chemical potential reflecting i.a. the pH value, ionic strength, osmotic stress, and the presence of all putatively interacting small molecules or ions. The conformational heterogeneity of an IDP can thus lead to a range of values of its hydrodynamic parameters at given temperature, instead of a single value. Therefore, the cellular and extracellular milieux may, via conformational changes, determine the kinetics of diffusion-controlled intermolecular reactions involving IDPs, thus regulating the effectiveness of processes occurring at the molecular level, e.g., formation of protein–protein or protein–nucleic acid complexes or the emergence of larger aggregates and microcrystals.

According to the well-established theory of diffusion of polymers (Flory 1949; Le Guillou and Zinn-Justin 1977), R_h can be approximated by a power function of the number of the polymer units (N)

$$R_h(N) = R_0 N^{\nu}, \quad (2)$$

where ν is the critical exponent equal to 1/3 for polymers perfectly folded into spheres and $\nu = 3/5$ for unfolded linear polymers. Within this formalism, N is ascribed to the number of amino acid residues in a protein chain, assuming they are indistinguishable.

To gain a better understanding of the complex nature of the hydrodynamic properties of IDPs, we have gathered an exhaustive set of currently available experimental literature results for IDPs' hydrodynamic radii determined by size-exclusion chromatography (SEC), analytical ultracentrifugation (AUC), pulsed-field gradient NMR (PFG NMR), dynamic light scattering (DLS), and fluorescence correlation spectroscopy (FCS). Against the background of the literature data, we present our FCS results obtained for a series of protein fragments involved in the regulation of human gene expression. The experimental data collected herein show unambiguously that the values of hydrodynamic radii of IDPs span the full space between the folded globular and denatured proteins in the $R_h(N)$ diagram.

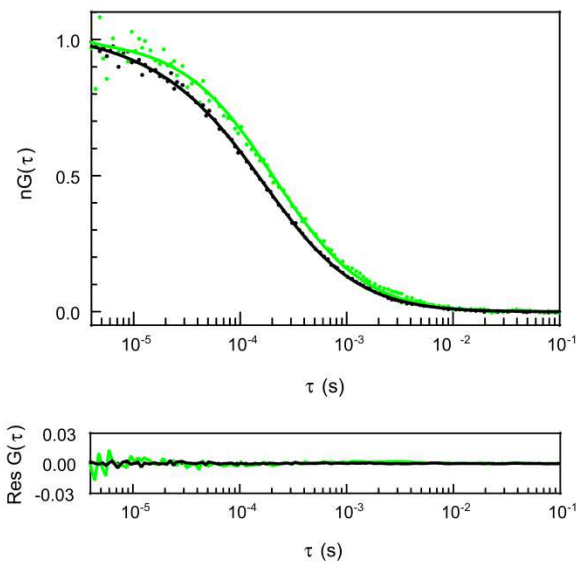

Fig. 2 Example normalized FCS autocorrelation curves with non-normalized fitting residuals for a folded protein, HSA (black), and an IDP, PARN C-mCherry (green), in 50 mM Tris/HCl buffer pH 8.0, 150 mM NaCl, 0.5 mM EDTA, and 1 mM TCEP, at 298 K. Dots, experimental points; lines, curves fitted according to Eq. 3

The FCS raw data were analyzed by Zen2010 (Zeiss) by global fitting of the autocorrelation curve to a set of 10 to 50 single measurements. This was preceded by a detailed inspection of each measurement to exclude possible oligomerization or aggregation of the sample in the confocal volume during the experiment. The autocorrelation function providing for 3D diffusion and photophysical processes was fitted according to the equations (Sahoo and Schwille 2011)

$$G(\tau) = G_T(\tau) \cdot G_D(\tau) \quad (3)$$

$$G_T(\tau) = \left(1 + \frac{P_T}{1-P_T} e^{-\frac{\tau}{\tau_T}}\right) \quad (4)$$

$$G_D(\tau) = \sum_{i=1}^{n} \frac{\Phi_i}{\left(1 + \left(\frac{\tau}{\tau_{d,i}}\right)\right) \cdot \left(1 + \left(\frac{\tau}{\tau_{d,i}}\right) \cdot \frac{1}{s^2}\right)^{1/2}} \quad (5)$$

$$\sum_i \Phi_i = 1, \quad (6)$$

where $G(\tau)$ is the fitted autocorrelation function; $G_T(\tau)$, normalized autocorrelation function for photophysical processes; $G_D(\tau)$, autocorrelation function for the diffusion of n components; P_T, triplet state or blinking fraction; τ_T, lifetime of the photophysical process; $\tau_{d,i}$, diffusion time for the ith component; s, structural parameter of the confocal volume; Φ_i, fraction of the ith component.

The protein R_h was determined from the diffusion time, τ_d, providing for the actual buffer viscosity

$$R_h = \frac{kT \cdot \tau_d}{6\pi\eta_0 \cdot D_{AF488} \cdot \tau_{AF488\_buf}}, \quad (7)$$

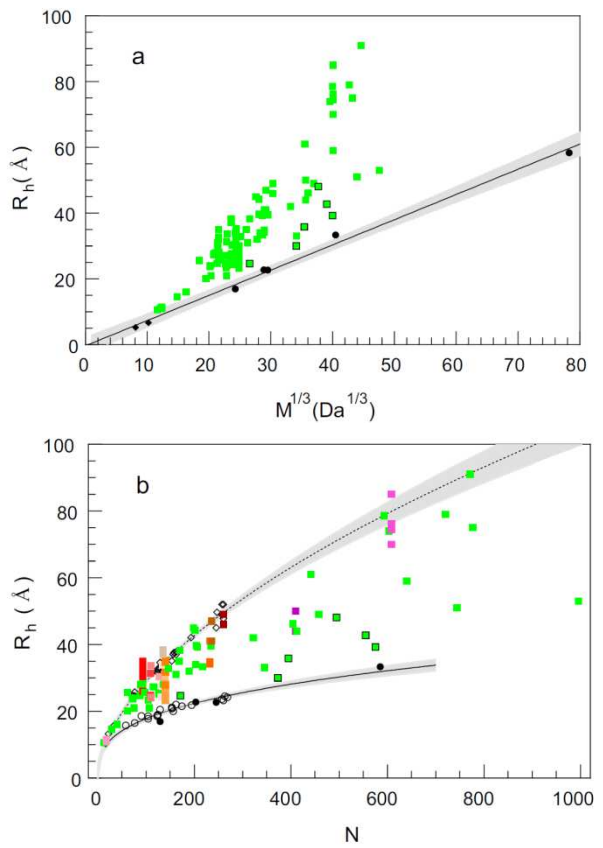

Fig. 3 Hydrodynamic radii, R_h, of IDPs (black-bordered green squares), globular proteins (black circles), and chemical fluorescent probes (AF488 and Alexa Fluor 546, black diamonds) determined by FCS in this study (Table 1) (a) as a function of the molecular mass, M, compared to literature R_h values for IDPs (plain green squares, Table S1); and (b) as a function of the polymer residue number, N, compared to literature data (plain squares, Table S1), where the results for the same protein at different conditions are marked in shades of red–orange–magenta; open circles and diamonds denote folded globular and chemically denatured proteins from literature, respectively. Chemical fluorescent probes and apoferritin are not taken into account in (b). The analytical curve according to Eq. (2) was fitted for folded proteins to the literature data merged with the results of this study, and for denatured proteins to literature data. The gray shaded regions represent the 95% confidence interval bands

Conclusions

The results obtained by FCS for fragments of proteins involved in human gene silencing (GW182, CNOT1, and PARN) determined by FCS, together with the collection of experimental data from literature, show that the hydrodynamic radii of IDPs of a given length can acquire any value in between the power function curves describing folded globular and denatured proteins. This is due to the fact that the R_h of IDPs are both sequence and environment-dependent. The dependence can be significant, especially for highly charged polypeptide chains. New theoretical and semi-empirical models are necessary to enable fast estimation of hydrodynamic properties of large IDPs taking into account their various conformational states that depend on the sequence, post-translational modifications, temperature, and the chemical milieu in an intricate way. The issue of how to combine all these parameters into one time-saving computational model is still a non-trivial task.

Table 1 Hydrodynamic radii, R_h, of protein constructs and chemical dyes determined in this work by FCS, in 50 mM Tris/HCl buffer pH 8.0, 150 mM NaCl, 0.5 mM EDTA, and 1 mM TCEP, at 25 °C

No.	Protein construct	N (res)	M (Da)	R_h (Å)	δR_h (Å)
Intrinsically disordered proteins					
1	CNOT1 M long	575	64,083	39	4
2	SUMO-GW182 SD	555	59,562	43	14
3	SUMO-GW182 SD10	494	53,516	48	5
4	PARN C-mCherry	395	44,479	35.8	1.4
5	GW182 SD10	373	39,820	30.1	1.6
6	SUMO-GW182 SD peptide	171	18,800	24.7	1.3
Globular proteins					
7	Apoferritin	4200	479,472	58	3
8	Human serum albumin	585	66,472	33.4	1.7
9	α-Chymotrypsinogen A	245	25,656	22.7	1.0
10	CNOT1 M short	203	24,074	22.8	1.2
11	Lysozyme	129	14,313	17.0	0.5
Chemical dyes					
12	AF546		1062	6.7	0.8
13	AF488		533	5.3	0.6

N, number of amino acid residues; M, molecular weight of the protein construct, δR_h, experimental uncertainty of R_h

Acknowledgements The work was supported by the National Science Centre under Grant No. UMO-2016/22/E/NZ1/00656. The studies were performed in the NanoFun laboratories co-financed by the European Regional Development Fund within the POIG.02.02.00-00-025/09 Program. The authors would like to thank Prof. Nahum Sonenberg and Dr. Marc Fabian for sharing plasmids for some protein constructs, Dr. Remigiusz Worch for helpful discussion, and Ms. Magdalena Duszka for excellent technical assistance.