RandPunkt
Journal

ISSN 00009172 (Online)
APPLIED MATHEMATICS SERIES

Mathematical Structure of Exceptionally Compact Globular Proteins Containing a Large Internal Functional Cavity

Vladimir N. Uversky[1]

[1] *Department of Molecular Medicine and USF Health Byrd Alzheimer's Research Institute, Morsani College of Medicine, University of South Florida, Tampa, FL 33647, USA*

*Correspondence to: **Vladimir N. Uversky** ▪ vuversky@usf.edu

Volume 1, Issue 1, 2026
Published on: 5 November 2026

Pages: [1–5] ▪ 5 pages
Article DOI: 10.5281/zenodo.24039999

The observation that some globular proteins exhibit hydrodynamic dimensions below the standard mass-dependent scaling law is recast as a *scaling problem*. We show that the total volume of a protein with a functional cavity is governed by the additive decomposition $V'_{\mathrm{tot}}(M) = V_{\mathrm{prot}}(M) + V'_{\mathrm{void}}(M) + V_{\mathrm{cavity}}(M)$, and that the six structural scenarios previously described in qualitative terms correspond to explicit constraints on the scaling exponents of each term. Around this decomposition we identify three mathematical objects: an effective scaling exponent ν', a fractal dimension $D_f = 1/\nu'$, and a packing-density functional $\phi(M)$. Their composition yields a predictive map Ψ that converts measurable hydrodynamic data into a classification of compactness. The note argues that this mathematical profile, rather than the qualitative taxonomy, is what makes the phenomenon of exceptionally compact proteins operational and testable.

Keywords: scaling laws; fractal dimension; packing density; globular proteins; cavity volume; hydrodynamic radius.

1 The mathematical object

The hydrodynamic radius R_S of a globular protein determined by DLS, SEC, or AUC deviates from the ideal spherical scaling $R_S \propto M^{1/3}$ because of asymmetry [1, 2], hydration [3, 4], disorder [5], and aggregation [6]. For a minority of proteins the deviation has the opposite sign: the measured R_S is *smaller* than the spherical baseline. This note formalizes that anomaly [7, 8, 9, 10, 11, 12, 13].

Received: 2 November 2026 ▪ Accepted: 4 November 2026 ▪ Published: 5 November 2026 © 2026 CC by ▪ 4.0v

Let M be the molecular mass. The total volume of a protein with a functional cavity is

$$V'_{\text{tot}}(M) = V_{\text{prot}}(M) + V'_{\text{void}}(M) + V_{\text{cavity}}(M), \tag{1}$$

where $V_{\text{prot}} = v_0 M$ is the van der Waals volume of the polypeptide chain, V'_{void} is the packing-void volume, and V_{cavity} is the functional cavity. For a standard globular protein of the same mass, $V_{\text{tot}}(M) = V_{\text{prot}}(M) + V_{\text{void}}(M)$. The anomaly is the sign of $V'_{\text{tot}} - V_{\text{tot}}$.

Observation 1.1 (Sequential heterogeneity)**.** *The decomposition* (1) *assumes that V'_{void} and V_{cavity} are smooth functions of M. In real proteins, however, the fraction of the chain that contributes to compact packing is not uniform along the sequence. Intrinsic-disorder analyses of proteins with functional cavities—for example, human plakophilin-2 (PKP2, UniProt Q99959), which combines an ordered C-terminal core of ARM repeats with a disordered N-terminal portion—show that a disordered N-terminal portion and an ordered portion coexist within the same chain, and that different disorder predictors assign noticeably different disorder scores to the same chain [14, 15, 16]. In terms of* (1)*, this amounts to replacing $V_{prot} = v_0 M$ by*

$$V_{prot}(M) = v_0\, M\, [1 - \eta(M)] + v_0\, \eta(M)\, M,$$

where $\eta(M) \in [0, 1]$ is the fraction of residues in disordered regions. The first term is compact van der Waals volume; the second is extended-chain volume, which contributes to R_g with a different exponent ($\nu \approx 0.6$ for a self-avoiding chain, versus $\nu \approx 1/3$ for a globule). The mathematical consequence is that ν' in (3) *is no longer a single exponent but an* effective *exponent that averages over the two populations. This does not invalidate the formalism; it makes it more realistic and explains why D_f for real globular proteins falls in* 2.6–2.7 *rather than* 3.

Assume power-law scaling for each term:

$$V'_{\text{void}}(M) = \alpha' M^{\beta'}, \qquad V_{\text{cavity}}(M) = \gamma M^{\delta}, \qquad V_{\text{void}}(M) = \alpha M^{\beta}. \tag{2}$$

The standard globular case has $\beta = 1$, so that $V_{\text{void}}/V_{\text{prot}}$ is constant. The six structural scenarios are then equivalent to the algebraic constraints of Table 1.

Table 1. Correspondence between the six structural scenarios and the scaling exponents of (2).

Case	Verbal condition	Exponent condition
1	$V'_{\text{void}} = V_{\text{void}}$	$\beta' = \beta = 1$
2	$V'_{\text{void}} > V_{\text{void}}$	$\beta' > 1$
3	$V'_{\text{void}} < V_{\text{void}}$	$\beta' < 1$
4	$V'_{\text{void}} \to 0,\ V_{\text{cavity}} = V_{\text{void}}$	$\alpha' \to 0,\ \delta = \beta$
5	$V'_{\text{void}} \to 0,\ V_{\text{cavity}} > V_{\text{void}}$	$\alpha' \to 0,\ \delta < \beta$
6	$V'_{\text{void}} \to 0,\ V_{\text{cavity}} < V_{\text{void}}$	$\alpha' \to 0,\ \delta > \beta$

Result 1.2 (Main predictive function)**.** *The compactness of a globular protein is encoded in the effective scaling exponent*

$$\boxed{\nu' = \frac{\mathrm{d}\ln R_g}{\mathrm{d}\ln M}} \tag{3}$$

and the six cases of Table 1 are recovered as the sign and magnitude of $\nu' - 1/3$. A protein is exceptionally compact when $\nu' < 1/3$.

Equation (3) is the mathematical core of the problem: it converts a measurable quantity (the slope of $\ln R_g$ versus $\ln M$) into a classification of compactness. It is the natural scaling-theoretic analogue of the affine predictor of linear calibration.

2 Three derived objects

The exponent (3) is not yet a usable tool. Three mathematical objects complete it.

First, the *fractal dimension* [17]. Since $R_g \propto M^{\nu'}$, the mass fractal dimension is

$$D_f = \frac{1}{\nu'}. \tag{4}$$

For an ideal sphere $D_f = 3$; for a Gaussian chain $D_f = 2$; for globular proteins $D_f \approx 2.6$–2.7, consistent with the empirical range $\nu \approx 0.37$–0.39 reported in [17]. Exceptionally compact proteins satisfy $D_f > 3 - 1/\nu_{\text{glob}}$.

Second, the *packing-density functional*. The fraction of volume occupied by the chain is

$$\phi(M) = \frac{V_{\text{prot}}(M)}{V'_{\text{tot}}(M)} = \frac{v_0 M}{v_0 M + \alpha' M^{\beta'} + \gamma M^{\delta}}, \tag{5}$$

which generalizes the Kepler bound $\pi/\sqrt{18} \approx 0.7405$ [18, 19]. For $\beta' = \beta = 1$ and $\gamma = 0$, (5) reduces to the constant packing fraction of an ideal close-packed solid; for $\beta' > 1$ it decays with mass, as reported in [20, 21, 22].

Third, the *compaction index*. Define

$$C(M) = D_f(M) - 2 \in (0, 1), \tag{6}$$

so that $C = 0$ corresponds to a Gaussian chain and $C = 1$ to a perfect sphere. Exceptionally compact proteins are those with $C(M) > C_{\text{glob}}$, where $C_{\text{glob}} \approx 0.63$ is the globular baseline.

3 The predictive map

Composing the three objects with (3) gives the predictive map

$$\Psi(R_g, M) = \big(\nu'(M),\ D_f(M),\ C(M)\big), \tag{7}$$

with $\nu'(M) = \mathrm{d}\ln R_g / \mathrm{d}\ln M$. The user enters a set of (R_g, M) pairs, estimates the local slope, and receives a fractal dimension, a compaction index, and a classification into one of the six cases of Table 1. The map Ψ is the mathematical profile of the phenomenon: it is what makes the notion of "exceptionally compact" operational, in the same way that the affine predictor made the nonheme iron assay operational.

The same construction extends to multivariate descriptors. If hydration, asymmetry, and disorder are recorded as additional covariates, (3) becomes a multiple regression in log–log coordinates,

$$\ln R_g = \beta_0 + \nu' \ln M + \beta_1 \ln(1 + \text{hyd}) + \beta_2 \ln(1 + \text{asym}) + \varepsilon, \tag{8}$$

whose matrix solution $\hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$ is the natural bridge to chemometric and structural-bioinformatics practice.

Remark 3.1 (On the disorder covariate)**.** The disorder covariate in (8) is not arbitrary. The variability among disorder predictors—the systematic disagreement between methods documented by the CAID assessment, and the method-dependence of per-residue disorder scores returned by tools such as RIDAO [14, 15, 16]—indicates that $\eta(M)$ is an observable that is *definition-dependent on the method*, not a unique intrinsic property. Consequently, the coefficient β_1 of (8) must be interpreted as a sensitivity to the *disorder estimator*, and its confidence interval must incorporate the inter-predictor dispersion, not only the fit error. This is the mathematical translation of the qualitative observation that distinct algorithms (composition, hydropathy, charge, machine learning) produce contrasting local profiles.

4 Outlook

The phenomenon of exceptionally compact globular proteins admits a compact mathematical profile organized around the effective scaling exponent (3) and the predictive map (7). The disciplines that sustain it are scaling theory and fractal geometry [17], supported by linear algebra and regression, and by the geometry of sphere packing [18, 19]. Framing the phenomenon in these terms does not replace the structural biology [20, 21, 22]; it makes it reproducible and testable, which is precisely what a short note can contribute to a problem that has so far been described only qualitatively.

The case of proteins with an ordered core and disordered N-terminal regions (e.g., PKP2) suggests that the natural extension of (7) is a map $\Psi(R_g, M, \eta)$ that separates the compact from the extended contribution, rather than a two-variable Ψ.

References

[1] Perrin, F., *Mouvement Brownien d'un ellipsoide (II). Rotation libre et dépolarisation des fluorescences. Translation et diffusion de molécules ellipsoidales*, Journal de Physique et le Radium 7.**1** (1936), pp. 1–11.

[2] Torre, J. G. de la and Bloomfield, V. A., *Hydrodynamic properties of complex, rigid, biological macromolecules: theory and applications*, Quarterly Reviews of Biophysics 14.**1** (1981), pp. 81–139.

[3] Kuntz, I. D. and Kauzmann, W., *Hydration of proteins and polypeptides*, Advances in Protein Chemistry 28 (1974), pp. 239–345.

[4] Rupley, J. A. and Careri, G., *Protein hydration and function*, Advances in Protein Chemistry 41 (1991), pp. 37–172.

[5] Marsh, J. A. and Forman-Kay, J. D., *Sequence determinants of compaction in intrinsically disordered proteins*, Biophysical Journal 98.**10** (2010), PMID: 20483348, pp. 2383–2390.

[6] Winzor, D. J., *Analytical gel filtration*, Physical principles and techniques of protein chemistry, ed. by Leach, S. J., Academic Press, 1969, pp. 451–495.

[7] Uversky, V. N., *Use of fast protein size-exclusion liquid chromatography to study the unfolding of proteins which denature through the molten globule*, Biochemistry 32.**48** (1993), pp. 13288–13298.

[8] ———, *Natively unfolded proteins: a point where biology waits for physics*, Protein Science 11.**4** (2002), pp. 739–756.

[9] Wilkins, D. K., Grimshaw, S. B., Receveur, V., Dobson, C. M., Jones, J. A., and Smith, L. J., *Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques*, Biochemistry 38.**50** (1999), PMID: 10600103, pp. 16424–16431.

[10] Garcia De La Torre, J., Huertas, M. L., and Carrasco, B., *Calculation of hydrodynamic properties of globular proteins from their atomic-level structure*, Biophysical Journal 78.**2** (2000), PMID: 10653785, pp. 719–730.

[11] Erickson, H. P., *Size and shape of protein molecules at the nanometer level determined by sedimentation, gel filtration, and electron microscopy*, Biological Procedures Online 11.**1** (2009), p. 32.

[12] O'Fagain, C., Cummins, P. M., and O'Connor, B. F., *Gel-filtration chromatography*, Methods in Molecular Biology 681 (2011), pp. 25–33.

[13] Siegel, L. M. and Monty, K. J., *Determination of molecular weights and frictional ratios of proteins in impure systems by use of gel filtration and density gradient centrifugation. Application to crude preparations of sulfite and hydroxylamine reductases*, Biochimica et Biophysica Acta 112.**2** (1966), PMID: 5329026, pp. 346–362.

[14] Dayhoff Guy W., 2. and Uversky, V. N., *Rapid prediction and analysis of protein intrinsic disorder*, Protein Science 31.**12** (2022), PMID: 36334049; PMCID: PMC9679974, e4496, doi:10.1002/pro.4496.

[15] Necci, M., Piovesan, D., CAID Predictors, DisProt Curators, and Tosatto, S. C. E., *Critical assessment of protein intrinsic disorder prediction*, Nature Methods 18.**5** (2021), PMID: 33875885; PMCID: PMC8105172, pp. 472–481, doi:10.1038/s41592-021-01117-3.

[16] Basu, S., Kihara, D., and Kurgan, L., *Computational prediction of disordered binding regions*, Computational and Structural Biotechnology Journal 21 (2023), PMID: 36851914; PMCID: PMC9957716, pp. 1487–1497, doi:10.1016/j.csbj.2023.02.018.

[17] Enright, M. B. and Leitner, D. M., *Mass fractal dimension and the compactness of proteins*, Physical Review E 71.**1** (2005), PMID: 15697635, p. 011912.

[18] Hales, T. C., *A proof of the Kepler conjecture*, Annals of Mathematics 162.**3** (2005), pp. 1065–1185.

[19] Hales, T., Adams, M., Bauer, G., Dang, T. D., Harrison, J., Hoang, L. T., Kaliszyk, C., Magron, V., McLaughlin, S., and Nguyen, T. T., *A formal proof of the Kepler conjecture*, Forum of Mathematics, Pi 5 (2017), e2.

[20] Tsai, J., Taylor, R., Chothia, C., and Gerstein, M., *The packing density in proteins: standard radii and volumes*, Journal of Molecular Biology 290.**1** (1999), pp. 253–266.

[21] Liang, J. and Dill, K. A., *Are proteins well-packed?*, Biophysical Journal 81.**2** (2001), pp. 751–766.

[22] Zhang, J., Chen, R., Tang, C., and Liang, J., *Origin of scaling behavior of protein packing density: A sequential Monte Carlo study of compact long chain polymers*, The Journal of Chemical Physics 118.**13** (2003), pp. 6102–6109.