

The hydrodynamic and conformational properties of denatured proteins in dilute solutions

Guy C. Berry*

Department of Chemistry, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213

Received 13 August 2009; Accepted 3 November 2009

DOI: 10.1002/pro.286

Published online 13 November 2009 proteinscience.org

Abstract: Published data on the characterization of unfolded proteins in dilute solutions in aqueous guanidine hydrochloride are analyzed to show that the data are not fit by either the random flight or wormlike chain models for linear chains. The analysis includes data on the intrinsic viscosity, root-mean-square radius of gyration, from small-angle X-ray scattering, and hydrodynamic radius, from the translational diffusion coefficient. It is concluded that residual structure consistent with that deduced from nuclear magnetic resonance on these solutions can explain the dilute solution results in a consistent manner through the presence of ring structures, which otherwise have an essentially flexible coil conformation. The ring structures could be in a state of continual flux and rearrangement. Calculation of the radius of gyration for the random-flight model gives a similar reduction of this measure for chains joined at their endpoints, or those containing loop with two dangling ends, each one-fourth the total length of the chain. This relative insensitivity to the details of the ring structure is taken to support the behavior observed across a range of proteins.

Keywords: protein; radius of gyration; intrinsic viscosity; loop formation

Introduction

In a 2005 article devoted to the discussion of “the case for (and against) residual structure in chemically denatured proteins” the authors cited a number of articles to that date suggesting that “nuclear magnetic resonance (NMR) studies suggest that significant secondary structure and long-range hydrophobic clusters persist in unfolded proteins even at high concentrations of urea or guanidine hydrochloride (GuHCl)”, whereas “intrinsic viscosity, hydrodynamic radii, and small-angle scattering experiments, some of which date back almost four decades, have been taken as evidence that the unfolded state is an effectively random coil ensemble.”¹ Since that publication, studies have continued to reveal secondary structure in strongly denatured proteins and polypeptides in dilute solution, mostly based on NMR

analysis, with such studies often motivated by interest in the folding of the denatured chain. The aforementioned use of the random flight model, which has dominated most of the interpretations of hydrodynamic behavior of proteins and polypeptides in dilute solution, starting with the seminal studies of Tanford *et al.*,² led to the conclusion that the chains may be modeled by an ensemble of random flight chains with excluded volume effects.^{1,3} One study treated the same data to conclude that the data on the intrinsic viscosity could be represented by a wormlike chain model with the chain having an appreciable persistence length, but negligible excluded volume.⁴ In this study, we suggest that taken together the experimental data on intrinsic viscosity, hydrodynamic radii, and the radius of gyration suggest a somewhat compact structure of the denatured proteins and polypeptides studied above, possibly owing to intramolecular association to form transient loops among chain elements widely separated along the chain backbone.

Abbreviations: NMR, nuclear magnetic resonance

*Correspondence to: Guy C. Berry, Carnegie Mellon University, Pittsburgh, PA 15213. E-mail: gcberry@andrew.cmu.edu

Results and Discussion

The published data

The addition of data on the root-mean-square radius of gyration R_G for a number of denatured proteins and polypeptides in aqueous (2–6) M GuHCl or (4–8) urea,⁵ motivates this reconsideration of the previously available data on the intrinsic viscosity $[\eta]$, second virial coefficient A_2 , and hydrodynamic radius R_H as functions of the molecular weight M in (5–6) M GuHCl. The data on $R_H = \Xi/6\pi\eta_{\text{solv}}$, with the molecular friction coefficient Ξ determined from the sedimentation coefficient in the original study,² have been augmented in recent study by estimates based on NMR studies on denatured proteins in a gradient field and by dynamic light scattering, discussed in detail in Ref. 6; here, η_{solv} is the solvent viscosity. The experimental data on these parameters are reproduced in Figure 1, where M is replaced by the number N of residues in the backbone, to more nearly represent the behavior in terms of the chain contour length L .

As noted above, prior treatments based on the behavior of $[\eta]$, A_2 , and R_H as functions of M have utilized a random flight flexible model for linear chains or a wormlike chain model in an attempt to elucidate the dependence of R_G on M , and the value $(R_G/\sqrt{M})_\Theta$ of R_G/\sqrt{M} under Flory theta conditions, for which $A_2 = 0$. As the GuHCl solvent system is found to be a thermodynamically good solvent for the proteins studied, with $MA_2 \approx [\eta]$, as often observed for solutions of flexible or wormlike chains,⁷ the evaluation of $(R_G/\sqrt{M})_\Theta$ from the dependence of $[\eta]$ on M requires the use of models and extrapolations to zero M , both of which introduce opportunities for error, especially in view of the relatively narrow range of M for which data are avail-

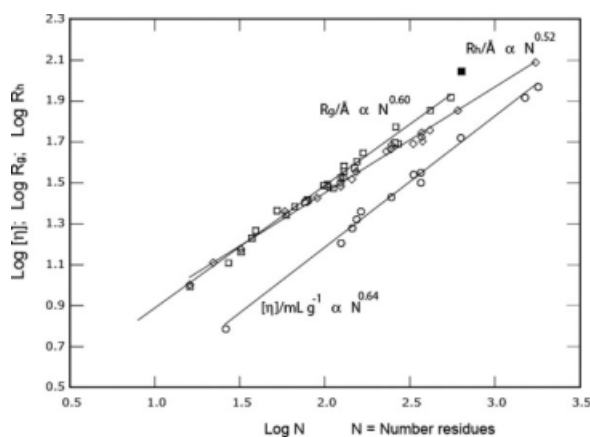


Figure 1. Experimental data reported in the literature on various proteins in strongly denaturing solvents giving the dependence on the number of residues N for the intrinsic viscosity,² $[\eta]$ (circles), the root-mean-square radius of gyration,⁵ R_G (squares) and hydrodynamic radius,⁶ R_H (diamonds).

able. For example, the analyses with the random flight and wormlike chain models, though quite different in conformational interpretations, provide reasonable fits to the data on $[\eta]$ as a function of M in the aqueous GuHCl solutions.^{2,4} However, both treatments provide estimates of $(R_G/\sqrt{M})_\Theta$ that are too large in comparison with the recent experimental data on R_G/\sqrt{M} ,⁵ unavailable to those earlier studies.

As may be seen in Figure 1, although the data are somewhat scattered, it appears that $\partial \ln [\eta]/\partial \ln M = 0.64_0$, $\partial \ln R_G/\partial \ln M = 0.59_7$, and $\partial \ln R_H/\partial \ln M = 0.51_6$. If interpreted alone, the value of $\partial \ln R_G/\partial \ln M$ would indicate a very strong intramolecular excluded volume effect, and that has been the usual interpretation.^{1,5} Similarly, if taken alone the values of $\partial \ln [\eta]/\partial \ln M$ and $\partial \ln R_H/\partial \ln M$ would each represent a more modest excluded volume effect. As mentioned above the data in the intermolecular thermodynamics given by the dependence of A_2 versus M would suggest a so-called good solvent, consistent with a substantial intramolecular excluded volume for a flexible-coil chain, but irrelevant to such behavior for rodlike or even wormlike chains with a large persistence length \hat{a} .⁷ The significance of these possibly disparate tendencies is considered next.

An analysis involving $[\eta]$

In a previous study, theoretical expressions for the parameter K were presented for a number of models, including linear and branched random flight chains, linear rodlike, and wormlike chains and ellipsoids of revolution (including spheres), where⁸

$$K = \frac{M[\eta]}{\pi N_{\text{AV}} R_G^2 R_H}, \quad (1)$$

with N_{AV} the Avogadro number. Of course, K is trivially related to the more familiar Flory-Fox parameter Φ' given by⁹

$$\Phi' = M[\eta]/R_G^3, \quad (2)$$

so that $\Phi' = \pi N_{\text{AV}}(R_H/R_G)K$ includes the dependence of R_H on M encoded into the definition for K such that for linear flexible or wormlike chain models K spans a much smaller range of values and is much more weakly dependent on M than is Φ' . For example, calculations with the random flight or wormlike chain models without intramolecular excluded volume show that $K_0 \leq K \leq K_\infty$, with K_0 and K_∞ the limiting values of K as M tends to zero and infinity, respectively; for example, $K_0 = 1$ and $K_\infty \approx 3$ for linear chains, depending on the details of the model used.⁸ Values of K_∞ for branched chains tend to exceed that for linear molecules, as does K_∞ for the random flight or wormlike chain models for ring-shaped molecules without intramolecular excluded

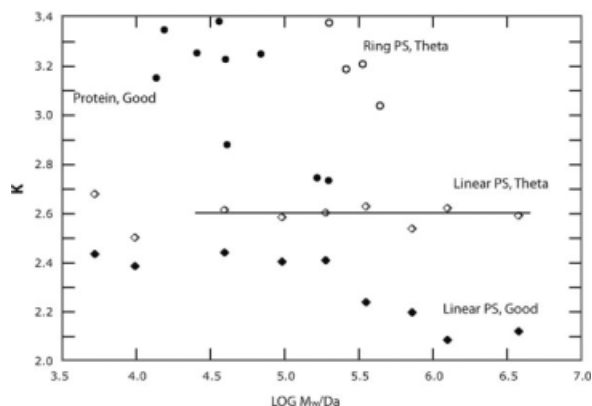


Figure 2. The function $K = [\eta]M/\{\pi N_{AV}R_G^2 R_H\}$ for several systems: unfolded proteins based on the data in Figure 1, filled circles; linear polystyrene under Flory theta conditions and good solvent conditions, unfilled and filled diamonds, respectively,^{14–16} and ring-shaped polystyrene, unfilled circles.^{11–13}

volume, for example, for a ring-shaped molecule, $K_{\infty, \text{RING}} = 1.519K_{\infty, \text{LIN}}$, with $K_{\infty, \text{LIN}}$ the value for the same model for a linear chain.¹⁰

Values of K calculated as a function of M_w for the data on proteins and polypeptides for the data in Figure 1 are shown in Figure 2, along with data on synthetically prepared well-defined model linear and ring-shaped macromolecules. In fact, there are not many such studies, owing principally to the difficulty of preparing well-defined models. Two studies on ring-shaped polymers prepared by the anionic polymerization of styrene, followed by ring closure give comparable results.^{11–13} Thus, for linear polystyrene, the experimental $K \approx K_{\infty}$ is constant for $10^4 < M < 10^6$ ($K \approx 2.6$) under Flory theta conditions, and in a good solvent (toluene) it decreases from this value as $K \propto M^{-0.04}$ with increasing M , reflecting the differing effects of intramolecular excluded volume on $[\eta]$ and $R_G^2 R_H$.^{14–16} Neither of these studies provided the behavior in good solvents needed to compute K . The data on the ring-shaped polystyrene determined under Flory Theta conditions depart from expectations in two ways: (i) K_{RING} is not expected to depend on M under these conditions and (ii) K_{RING} is smaller than the value 3.95 expected by comparison with K_{LIN} for the same polystyrene. The unexpected behavior may reflect the presence of linear chains remaining from the difficult synthesis of the ring-shaped polymers, as observed in some chromatography studies.¹⁷ The data on proteins and polypeptides are seen to be scattered, but with a tendency to decrease from about 3.4 to 2.7 with increasing M over the range studied, with $K \propto 10^{-0.07}$. As the protein and polypeptide samples used in the various studies differ, resort is made to the smoothed correlations shown in Figure 1 to compute K . It is seen that the resulting K for the large M

limit is comparable to K_{∞} for a linear random-flight or wormlike chain for large M under Flory theta conditions, but increases to larger values with decreasing M , so that the overall pattern is similar to that seen in Figure 1 for ring-shaped polystyrenes. The effects on K or related to those note on the Flory-Fox parameter Φ in recent considerations of size exclusion chromatography on denatured proteins.^{18,19} It may also be noted in Figure 1 that $R_H \approx R_G$ for the lower values of M . Although one can conclude that the protein conformation appears to be more compact than a linear random flight or wormlike chain at the lower range of N , and that such behavior is not an indication of free draining hydrodynamics, which would decrease K , it is not possible to interpret these data in terms of a definitive model. Possibilities are discussed in the following.

The effects of association among chain residues

Intermolecular association would lead to branch-shaped aggregates, and as mentioned above, K will tend to be increased above that for the linear counterpart, offering one possible interpretation. However, this possibility is discounted as the existence of such intermolecular association would be expected to have characteristic features in the dependence of the viscosity and light (or X-ray) scattering behavior, not reported for the studies cited.

Studies by NMR and MCD over the past decade have revealed that the so-called denatured state of proteins and polypeptides contain locally ordered sections of a helical nature, even in solvents that strongly suppress hydrogen bonding.^{1,20–22} This includes evidence that these sections may be involved in intramolecular association among units far removed along the chain backbone, forming loops that flicker in and out of existence. As observed above, the result that K for the lower molecular weight proteins in aqueous GuHCl solutions is comparable to the behavior seen for the ring-shaped polystyrene, suggesting the presence of ring-shaped molecules or chains with containing loops, at least in that range of M . As with the data on polystyrene, the modestly decreasing value of K noted with the data on the protein solutions in aqueous GuHCl solution may reflect a modest intramolecular excluded volume effect, with the higher value obtained at low M closer to the limit that would be observed under Flory theta solvent conditions.

The data on $R_G/N^{1/2}$ and $R_H/N^{1/2}$ for the data in Figure 1 are shown in Figure 3, along with data on $(R_G/N^{1/2})_{\text{STAT}}$ computed via statistical coil model for the unfolded state of a range of proteins as a function of N , including excluded volume effects, and taking account of conformational constraints consistent with NMR analysis revealing (non-H-bond stabilized) helical sequences in aqueous GuHCl

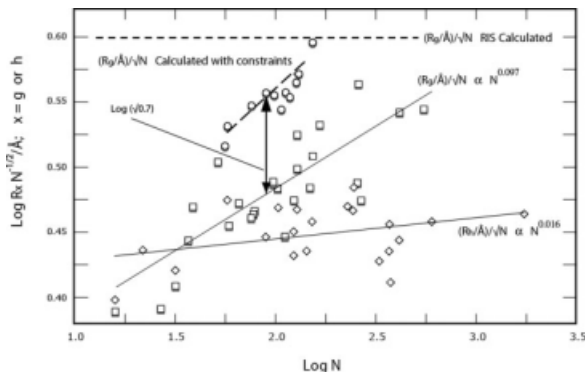


Figure 3. Experimental data on $R_G/N^{1/2}$ and $R_G/N^{1/2}$ versus $\log N$ using the data in Figure 1, along data on $(R_G/N^{1/2})_{\text{STAT}}$ calculated via a statistical model with conformational constraints (aside from intramolecular association)²³ and $(R_G/N^{1/2})_{\text{STAT}}$ versus calculated with rotational-isomeric-state model.²⁴

solution,²³ and a constant value of $(R_G/N^{1/2})_{\text{RIS}}$ calculated for the rotational-isomeric-state model in the limit of large N .²⁴ It may be seen that the experimental values of $R_G/N^{1/2}$ are about equal to $\sqrt{0.7(R_G/N^{1/2})_{\text{STAT}}}$. Thus, if $(R_G/N^{1/2})_{\text{STAT}}$ is presumed to be an accurate measure of $R_G/N^{1/2}$ for the chain without any intramolecular association producing ring structures, $g = R_G^2/(R_G^2)_{\text{LIN}} \approx 0.7$, larger than the value $g = 1/2$ expected for a pure ring-shaped polymer. Although not impossible and perhaps possible for some unfolded proteins,²⁵ it would perhaps be unlikely for transient intramolecular hydrophobic clusters to form rings by association of their endgroups, so it is of interest to consider values of g for a model with a loop formed by association of more generally placed repeating units.

A simplified model with the results shown in Figure 4 comprises a chain with an association of residues each a distance xL from the ends of a chain of length L , resulting in a loop with $(1 - 2x)L$ residues, or length $(1 - 2x)L$, and two “branches,” each with xL residues, emanating from the crosslink locus. Use of the random flight model to compute g including the constraint imposed by the effective intrachain crosslink site according to the usual expression for a chain with n statistical segments of length $b = L/n$.²⁶

$$R_G^2 = \frac{b^2}{2n} \sum \sum (i - j) \quad (3)$$

where the sums are over all possible segments i and j , consistent with the crosslink constraint. The calculation may be conveniently accomplished by summing the contributions segment pairs confined to several components of the total structure: (1) a loop with $n - 2xn$ segments, (2) a linear chain with $2xn$ segments made from the two branches, (3) four lin-

ear chains, each with $n/2$ segments, comprising a $1/2$ loop + one branch, and (4) less the over-count from four linear chains, each with $(n - 2xn)/2$ segments from four $1/2$ loops and four linear chains, each with xn segments from four branches. The calculation gives g as a function of the fraction $1 - 2x$ of the chain in the loop as shown in Figure 4, with

$$g(x) = 4x^3 + 1/2. \quad (4)$$

As required, g is unity or one-half for fractions of residues in the loop equal to zero ($x = 1/2$) or unity ($x = 0$), respectively. Although this model is certainly too simplified to permit a definitive quantitative conclusion, it may be noted that the observed $g \approx 0.7$ would correspond to about a fraction 0.25–0.3 of the residues in the loop. However, even if the conclusion cannot be considered to be quantitatively accurate, the existence of such transient structures with a life-time long enough to be seen in the parameters of interest here would seem to be reasonable in view of the properties found in the NMR studies and could be relevant to the refolding process of the denatured protein.

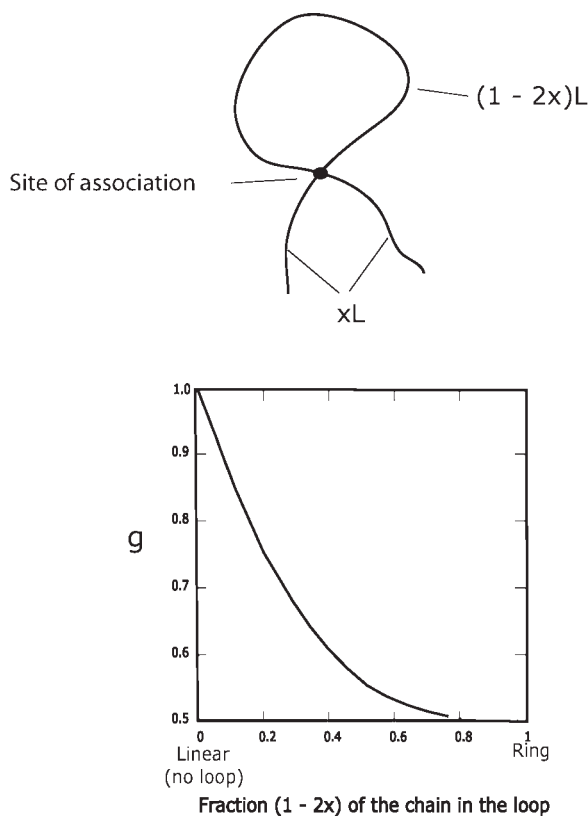


Figure 4. Upper: schematic diagram of a loop of length $(1 - 2x)L$, and two “branches,” each of length xL , emanating from the site of the association creating the structure. Lower: a plot of g versus the fraction $1 - 2x$ of the residues in the loop.

Conclusions

Based on the preceding analysis of published data on the characterization of proteins in dilute solutions in aqueous GuHCl, it is concluded that the data are not fit by either the random flight or worm-like chain models for linear chains, but can be rationalized if it is assumed that transient intramolecular association form temporary ring conformers among the chain units. The analysis includes data on the intrinsic viscosity, root-mean-square radius of gyration, from small-angle X-ray scattering, and hydrodynamic radius, from the translational diffusion coefficient. Such residual ring structures in a state of continual flux and rearrangement in otherwise unfolded proteins are consistent with behavior deduced from NMR on these solutions. Calculation of the radius of gyration for the random flight model gives a similar reduction of this measure for chains joined at their endpoints, or those containing loop with two dangling ends, each one-fourth the total length of the chain. This relative insensitivity to the details of the ring structure is taken to support the behavior observed across a range of proteins.

References

1. McCarney ER, Kohn JE, Plaxco KW (2005) Is there or isn't there? The case for (and against) residual structure in chemically denatured proteins. *Crit Rev Biochem Mol Biol* 40:181–189.
2. Tanford C, Kawahara K, Lapanje S (1967) Proteins as random coils. I. Intrinsic viscosities and sedimentation coefficients in concentrated guanidine hydrochloride. *J Am Chem Soc* 89:729–736.
3. Tanford C (1970) Protein denaturation. C. Theoretical models for the mechanism of denaturation. *Adv Protein Chem* 24:1–95.
4. Bohdanecky M, Petrus V (1991) Analysis of hydrodynamic data for denatured globular proteins in terms of the worm-like cylinder model. *Int J Biol Macromol* 13:231–234.
5. Kohn JE, Millett IS, Jacob J, Zagrovic B, Dillon TM, Cingel N, Dothager RS, Seifert S, Thiyagarajan P, Sosnick TR, Hasan MZ, Pande VS, Ruczinski I, Doniach S, Plaxco KW (2004) Random-coil behavior and the dimensions of chemically unfolded proteins. *Proc Natl Acad Sci USA* 101:12491–12496.
6. Zhou H-X (2002) Dimensions of denatured protein chains from hydrodynamic data. *J Phys Chem B* 106:5769–5775.
7. Casassa EF, Berry GC, Polymer solutions. In: Allen G, Ed. (1988) *Comprehensive polymer science*, Vol. 2. Ch 3; New York: Pergamon Press.
8. Berry GC (1988) Remarks on a relation among the intrinsic viscosity, the radius of gyration, and the translational friction coefficient. *J Polym Sci Part B: Polym Phys* 26:1137–1142.
9. Fox TG, Jr, Flory PJ (1951) Intrinsic viscosity relations for polystyrene. *J Am Chem Soc* 73:1915–1920.
10. Fukatsu M, Kurata M (1966) Hydrodynamic properties of flexible-ring macromolecules. *J Chem Phys* 44:4539–4545.
11. Roovers J (1985) Dilute-solution properties of ring polystyrenes. *J Polym Sci Polym Phys Ed* 23:1117–1126.
12. Hadziioannou G, Cotts PM, ten Brinke G, Han CC, Lutz P, Strazielle C, Rempp P, Kovacs AJ (1987) Thermodynamic and hydrodynamic properties of dilute solutions of cyclic and linear polystyrenes. *Macromolecules* 20:493–497.
13. McKenna GB, Hadziioannou G, Lutz P, Hild G, Strazielle C, Straupe C, Rempp P, Kovacs AJ (1987) Dilute solution characterization of cyclic polystyrene molecules and their zero-shear viscosity in the melt. *Macromolecules* 20:498–512.
14. Einaga Y, Koyama H, Konishi T, Yamakawa H (1989) Intrinsic viscosity of oligo- and polystyrenes. *Macromolecules* 22:3419–3424.
15. Dehara K, Yoshizaki T, Yamakawa H (1993) Translational diffusion coefficient of oligo- and poly(methyl methacrylates) in dilute solutions. *Macromolecules* 26:5137–5142.
16. Abe F, Einaga Y, Yoshizaki T, Yamakawa H (1993) Excluded-volume effects on the mean-square radius of gyration of oligo- and polystyrenes in dilute solutions. *Macromolecules* 26:1884–1890.
17. Cho D, Park S, Kwon K, Chang T, Roovers J (2001) Structural characterization of ring polystyrene by liquid chromatography at the critical condition and MALDI-TOF mass spectrometry. *Macromolecules* 34:7570–7572.
18. Dondos A (2006) Applicability of the modified universal calibration of gel permeation chromatography on proteins. *J Chromatogr* 1127:183–186.
19. Dondos A (2007) Determination of Flory's parameter F for proteins based on the modified universal calibration of the gel permeation chromatography. *Biomacromolecules* 8:2979–2981.
20. Eliezer D (2007) Characterizing residual structure in disordered protein states using nuclear magnetic resonance. *Methods Mol Biol (Clifton, NJ)* 350:49–67.
21. Mittag T, Forman-Kay Julie D (2007) Atomic-level characterization of disordered protein ensembles. *Curr Opin Struct Biol* 17:3–14.
22. Chugh J, Sharma S, Hosur RV (2007) Pockets of short-range transient order and restricted topological heterogeneity in the guanidine-denatured state ensemble of GED of dynamin. *Biochemistry* 46:11819–11832.
23. Jha AK, Colubri A, Freed KF, Sosnick TR (2005) Statistical coil model of the unfolded state: resolving the reconciliation problem. *Proc Natl Acad Sci USA* 102:13099–13104.
24. Miller WG, Brant DA, Flory PJ (1967) Random coil configurations of polypeptide copolymers. *J Mol Biol* 23:67–80.
25. Bernado P, Bertoncini CW, Griesinger C, Zweckstetter M, Blackledge M (2005) Defining long-range order and local disorder in native alpha-synuclein using residual dipolar couplings. *J Am Chem Soc* 127:17968–17969.
26. Zimm BH, Stockmayer WH (1949) The dimensions of chain molecules containing branches and rings. *J Chem Phys* 17:1301–1314.