# Dimensions of Denatured Protein Chains from Hydrodynamic Data

**Huan-Xiang Zhou***

*Department of Physics, Drexel University, Philadelphia, Pennsylvania 19104*

*Received: September 5, 2001; In Final Form: March 7, 2002*

The chain length dependence of the intrinsic viscosity and the Stokes radius demonstrates that denatured proteins are not in the "unperturbed" state but rather have nonlocal excluded-volume interactions between residues. Here, we present hydrodynamic calculations on denatured protein chains with $N = 16$ to 104 residues. Chain conformations were generated by sampling $(\phi, \psi)$ from those collected from loop regions in the protein structure database. Excluded-volume effects were explicitly accounted for by assigning a hard-sphere diameter $d_\alpha$ to the $C_\alpha$ atoms. The value of $d_\alpha$ for a chain at a given length was adjusted so the experimental intrinsic viscosity is reproduced. This procedure allowed a simultaneous reproduction of the experimental Stokes radii for all the chain lengths studied. The distribution of the end-to-end distance $R$ and its mean square $\langle R^2 \rangle$ were obtained from the sampled chain conformations. When the chain length dependence of the intrinsic viscosity and the Stokes radius is fitted to the Zimm model (which predicts $[\eta]M = \Phi \langle R^2 \rangle^{3/2}$ and $R_s = K \langle R^2 \rangle^{1/2}$), the length dependence of the coefficients $\Phi$ and $K$ is obtained. The effective bond length $b_{eff}$ [defined via $\langle R^2 \rangle = (N - 1) b_{eff}^2$] ranges from 5.71 to 8.78 Å, and the distribution of $R$ is nearly Gaussian when $N > 30$. The value of $b_{eff}$ is extrapolated to increase to 10.6 Å at $N = 1000$. For any chain, the distance between two interior residues is characterized by a shorter effective bond length than the end-to-end distance.

## Introduction

As recent experimental[1] and previous and recent theoretical[2−4] studies indicate, the dimension of a denatured protein chain is an important determinant of the folding thermodynamics and kinetics. For denatured proteins, hydrodynamic measurements provide the most direct means to the dimensional information. However, extracting such information from hydrodynamic data has many difficulties. In this paper, we present hydrodynamic calculations on denatured protein chains. On the basis of comparison with experimental data for intrinsic viscosity and Stokes radius, we obtain average dimensions for chains with $N = 16$ to 104 residues.

The conformation of a polypeptide chain is determined by the interactions among the residues and between them and the solvent. Of great historic importance is the "unperturbed" state, in which the nonlocal repulsive or attractive interactions between residues are exactly balanced by residue-solvent interactions so the conformation is solely determined by local interactions. The rotational isomeric state theory developed by Flory and collaborators[5−8] was specifically designed for this state. The unperturbed state provides a convenient reference state but otherwise not particularly relevant for the polypeptide chain under most solvent conditions. The chain length dependence of the intrinsic viscosity and the Stokes radius demonstrates that denatured proteins are not in the "unperturbed" state but rather have nonlocal excluded-volume interactions between residues.

Here, we explicitly account for nonlocal interactions by modeling excluded-volume effects between any two residues in the chain. Each $C_\alpha$ atom was assigned a hard-sphere diameter $d_\alpha$. The value of $d_\alpha$ for a chain at a given length was adjusted so the experimental intrinsic viscosity is reproduced. The distribution of the end-to-end distance $R$ and its mean square $\langle R^2 \rangle$ were obtained from the sampled chain conformations. When the chain length dependence of the intrinsic viscosity and the Stokes radius is fitted to the Zimm model (which predicts $[\eta]M = \Phi \langle R^2 \rangle^{3/2}$ and $R_s = K \langle R^2 \rangle^{1/2}$), the length dependence of the coefficients $\Phi$ and $K$ is obtained. The effective bond length $b_{eff}$ [defined via $\langle R^2 \rangle = (N - 1) b_{eff}^2$] is found to range from 5.71 to 8.78 Å, and the distribution of $R$ is nearly Gaussian when $N > 30$. The value of $b_{eff}$ is expected to increase to 10.6 Å at $N = 1000$. For any chain, the distance between two interior residues is characterized by a shorter effective bond length than the end-to-end distance.

## Methods

**Generation of Peptide Chain Conformations.** Chains were generated from a rigid backbone with standard geometry[9] by varying the $(\phi, \psi)$ angles, as illustrated in Figure 1. The N−$C_\alpha$, $C_\alpha$-C, and C−N bond lengths were 1.458, 1.525, and 1.329 Å, respectively, whereas the N−$C_\alpha$-C, $C_\alpha$-C−N, and C−N−$C_\alpha$ angles were 111.2°, 116.2°, and 121.7°, respectively. The $(\phi, \psi)$ angles were randomly sampled from those adopted by residues located in 4599 loops consisting of 11−30 residues collected from 1889 proteins in the FSSP library.[10] The $(i + 1)$th $C_\alpha$ position was accepted only when its distances to $C_\alpha$'s at positions 1 to $i - 1$ were all beyond a cutoff $d_\alpha$. When $d_\alpha > 5$ Å, the contact distance between $i - 1$ and $i + 1$ was always shortened to 5 Å. The shortening of the cutoff distance for the $i - 1$ to $i + 1$ contact was motivated by the fact that, in protein structures, the distances of this type of contacts ranges from 5 to 7.5 Å. Note that the nearest $C_\alpha$−$C_\alpha$ distance $b$ is fixed at 3.804 Å for the rigid backbone geometry used in this study and contact need not be checked between them.

The $(\phi, \psi)$ map of the 59 064 loop residues is shown in Figure 2. The percentages of the 20 types of amino acids, in descending

* To whom correspondence should be addressed. Address after July 1, 2002: Institute of Molecular Biophysics and Department of Physics, Florida State University, Tallahassee, FL 32306. Phone: (215) 895-2716. Fax: (215) 895-5934. E-mail: zhou@sb.fsu.edu.

10.1021/jp013403+ CCC: $22.00 © 2002 American Chemical Society
Published on Web 05/09/2002

**TABLE 1: Calculated and Experimental Results on Intrinsic Viscosity and Stokes Radius for Seven Native Proteins**

| protein | PDB code | $N$ | $[\eta]$ (cm$^3$/g) all atom | $[\eta]$ (cm$^3$/g) $C_\alpha$ only | $[\eta]$ (cm$^3$/g) exp | $R_S$ (Å) all atom | $R_S$ (Å) $C_\alpha$ only | $R_S$ (Å) exp |
|---|---|---|---|---|---|---|---|---|
| calcitonin | 1bku | 32 | 5.65 | 5.56 | | 14.3 | 14.1 | |
| BPTI | 5pti | 58 | 3.34 | 3.21 | | 15.1 | 14.9 | 15.2 ± 0.2[a,b] |
| horse cytochrome *c* | 1hrc | 104 | 2.90 | 2.73 | | 17.9 | 17.5 | 17.2[b] |
| bovine α-lactalbumin | 1hfz | 123 | 3.17 | 3.07 | | 19.2 | 19.0 | 18.8 ± 0.2[c] |
| bovine ribonuclease A | 7rsa | 124 | 3.24 | 3.31 | 3.30 ± 0.04[d] | 19.2 | 19.4 | 19.0 ± 0.4[e] |
| hen lysozyme | 6lyz | 129 | 3.02 | 3.03 | 2.99 ± 0.01[f,g] | 19.0 | 19.0 | 19.8,[b] 19.0[g] |
| horse myoglobin | 1azi | 153 | 3.11 | 3.15 | 3.15[h] | 20.4 | 20.5 | 20.4[b] |

[a] Reference 17. [b] Reference 18. Note that the Stokes radius determined in this reference by pulse field gradient NMR is subject to an adjustable scaling constant. The authors stated that they determined the scaling constant such that the value for hen lysozyme is 19.75 Å (the same as determined by small-angle X-ray scattering). However, the Stokes radius listed in Table 1 of this reference for hen lysozyme is 20.5 Å. We therefore scaled down all values of Stokes radii in that table by 20.5/19.75 = 1.038. [c] Reference 19. [d] Reference 20. [e] Reference 21. [f] Reference 22. [g] Reference 23. [h] Reference 24.

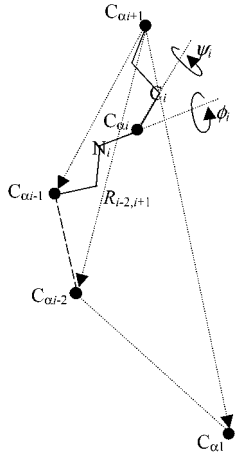


**Figure 1.** Generation of chain conformations. The bond lengths and bond angles are fixed. For residue $i$, trial positions of $C_i$ and $N_{i+1}$ are generated from randomly sampled ($\phi_i$, $\psi_i$) angles. The position of $C_{\alpha i+1}$ is then generated by the requirement that $C_{\alpha i}$, $C_i$, $N_{i+1}$, and $C_{\alpha i+1}$ are on the same plane. The distances of $C_{\alpha i+1}$ from $C_{\alpha 1}$ up to $C_{\alpha i-1}$ are then calculated. If any of these distances is greater than the hard-sphere diameter $d_\alpha$, the trial positions of $C_i$, $N_{i+1}$, and $C_{\alpha i+1}$ are abandoned. New ($\phi_i$, $\psi_i$) angles are sampled until the excluded-volume test is passed.

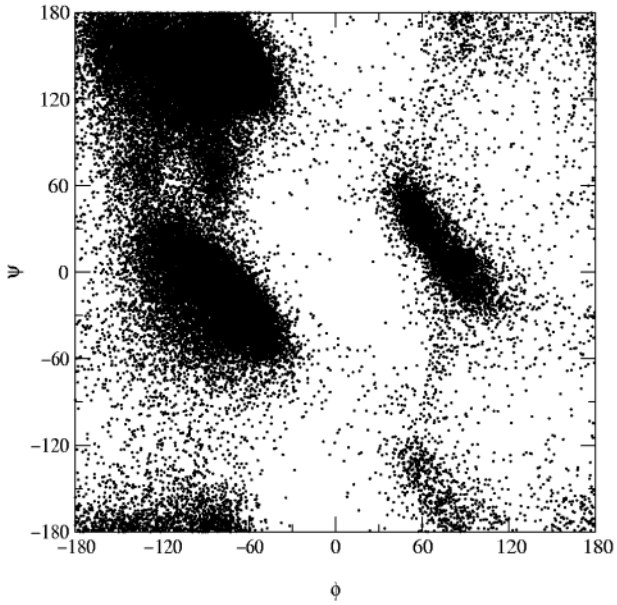


**Figure 2.** ($\phi$, $\psi$) map used to generate chain conformations. It may be divided into three regions: α region with $\phi < 0$ and $-120° < \psi < 60°$, $\beta$ region with $\phi < 0$ and $-180° < \psi < -120°$ or $60° < \psi < 180°$, and the positive $\phi$ region. The fractions of data points in these three regions are 40%, 49%, and 11%, respectively.

order, are 10.2, 8.8, 7.4, 6.9, 6.7, 6.6, 6.0, 5.8, 5.6, 5.5, 4.8, 4.5, 3.7, 3.6, 3.4, 3.4, 2.6, 1.7, 1.6, and 1.5, respectively, for Gly, Pro, Asp, Ser, Ala, Leu, Glu, Thr, Lys, Asn, Val, Arg, Ile, Phe, Gln, Tyr, His, Met, Cys, and Trp.

The above procedure for generating chain conformations differs from the classical approach of Flory and collaborators in two respects. First, local conformations were generated from the ($\phi$, $\psi$) map of residues in the loop regions in the structure database as opposed to from an empirical energy function. Second, nonlocal interactions are included by excluded-volume effects. In both respects, our procedure is similar to one used by Fiebig et al.[11] to predict NMR parameters in the denatured state. Pappu et al.[12] used a related procedure to study the effects of excluded volume on the conformations accessible to short alanine peptides.

**Calculation of Hydrodynamic Properties.** We have developed an efficient method for predicting the intrinsic viscosity $[\eta]$ and Stokes radius $R_S$ of a native protein based on discretizing the surface of the protein structure.[13−15] Here, this method was adopted with two modifications. First, instead of generating the protein surface by using every atom, we used only the $C_\alpha$ atoms. Second, because of the variable nature of the denatured state, we used many conformations to calculate an average $[\eta]$ or $R_S$.

The radius $r_h$ of the $C_\alpha$ atoms for the purpose of generating the hydrodynamic surface of each conformation in the denatured state had to be fixed. To this end, we carried out hydrodynamic calculations on native proteins, either using all atoms or using just $C_\alpha$ atoms with an adjustable radius $r_h$. Table 1 shows that a value of $r_h = 5.1$ Å allows for reproduction of the all-atom results for seven protein chains with 32 to 153 residues and the available experimental results as well.[16]

**Determination of Hard-Sphere Diameter $d_\alpha$.** $d_\alpha$ was determined by trial and error. Initial guesses of $d_\alpha$ were used to generate chain conformations with a given length $N$, and then intrinsic viscosity and Stokes radius were calculated over these conformations. Adjustment of $d_\alpha$ was stopped when the average of the intrinsic viscosity came to agreement with the experimental value (as calculated from eq 1 given below).

The number of conformations generated at the final selected $d_\alpha$ for a given $N$ is given in Table 2. All of the 3500 to 40 000 conformations for different chain lengths were used to calculate the mean square of the end-to-end distance $R$ and obtain the distribution of $R$. However, because of the long CPU times for the hydrodynamic calculations, only ~700 for each chain length were actually used for calculating the intrinsic viscosity $[\eta]$ and the Stokes radius $R_s$. $[\eta]$ and $R_s$ showed much smaller variations relative to $R^2$, so statistically meaningful results for $[\eta]$ and $R_s$ were obtained despite the reduced sampling sizes.

**TABLE 2: Number of Conformations Generated for Calculating Mean Square of End-to-End Distance, for Calculating Intrinsic Viscosity and Stokes Radius, and Fraction of Conformations Surviving Excluded-Volume Test**

| | no. of conformations | | |
|---|---|---|---|
| $N$ | calc of $\langle R^2 \rangle$ | calc of $[\eta]$ and $R_s$ | survival fraction[a] |
| 16 | 39370 | 700 | 1 |
| 22 | 24826 | 700 | $0.731 \pm 0.003$ |
| 27 | 27912 | 700 | $0.486 \pm 0.002$ |
| 32 | 32010 | 700 | $0.270 \pm 0.002$ |
| 37 | 23442 | 700 | $0.119 \pm 0.001$ |
| 42 | 34905 | 700 | $(4.68 \pm 0.04) \times 10^{-2}$ |
| 47 | 30159 | 700 | $(1.85 \pm 0.02) \times 10^{-2}$ |
| 52 | 22633 | 700 | $(6.39 \pm 0.08) \times 10^{-3}$ |
| 58 | 18143 | 700 | $(1.85 \pm 0.03) \times 10^{-3}$ |
| 62 | 15072 | 700 | $(9.36 \pm 0.17) \times 10^{-4}$ |
| 67 | 16151 | 700 | $(5.04 \pm 0.09) \times 10^{-4}$ |
| 72 | 5866 | 715 | $(1.75 \pm 0.06) \times 10^{-4}$ |
| 76 | 7547 | 855 | $(8.58 \pm 0.30) \times 10^{-5}$ |
| 82 | 6781 | 500 | $(4.00 \pm 0.10) \times 10^{-5}$ |
| 90 | 8231 | 776 | $(8.89 \pm 0.30) \times 10^{-6}$ |
| 104 | 3513 | 561 | $(9.71 \pm 0.50) \times 10^{-7}$ |

[a] Because of excluded volume, an overwhelming fraction of conformations generated by sampling the ($\phi$, $\psi$) angles were eliminated. This column lists the fraction of conformations that survive the excluded-volume test. The total number of conformations generated after surviving the excluded-volume test is listed in the second column. Out of these, the number of conformations that were actually used for hydrodynamic calculations is listed under the third column.

Generating the chain conformations itself took long CPU times because of the test for excluded volume. As the chain length increases, the fraction of conformations out of all possible conformations generated by sampling the ($\phi$, $\psi$) angles diminished dramatically. The surviving fractions for different chain lengths are listed in Table 2.

## Results

**Hydrodynamic Properties of Denatured Proteins.** Tanford[25] collected data for the intrinsic viscosity of 16 proteins denatured at GuHCl concentrations from 5 to 7.5 M at 25 °C. The chain length $N$ ranges from 26 to 1739 residues. He found that the data can be well fitted to an expression

$$[\eta]M_0 = 77.3N^{0.666} \tag{1}$$

where $M_0 = 110$ is the average residue molecular weight and $[\eta]$ is in units of cm$^3$/g. We use eq 1 to give the "experimental" result for the intrinsic viscosity at a given $N$.

In the same spirit, we have collected data from the literature for the Stokes radii of 20 denatured polypeptide chains. The chain length ranges from 16 to 1739. The experimental data are listed in Table 3. They can be well fitted to an expression (see Figure 3)

$$R_S = 2.518N^{0.522}\ (\text{Å}) \tag{2}$$

Similar equations have been proposed previously.[18,28] We use eq 2 to give the "experimental" result for the Stokes radius at a given chain length $N$.

The exponent for the $N$ dependence of $[\eta]$ and $R_S$ is exactly 0.5 in the unperturbed state.[30d] Deviations from this value in eqs 1 and 2 demonstrate that denatured proteins are not in the unperturbed state.

**Hard-Sphere Diameter of $C_\alpha$ Atoms.** The results of $d_\alpha$ found for 16 chain lengths ranging from 16 to 104 are listed in Table 4. Contrary to our initial expectation, $d_\alpha$ has a strong dependence on $N$, increasing from 0 at $N = 16$ to 5.68 Å at $N = 104$. However, the increase in $d_\alpha$ slows down significantly. As Figure 4 shows, the $d_\alpha$ versus $N$ curve for $N > 20$ can be fitted to an empirical equation

$$d_\alpha = 5.82\,\frac{N^3 - 794.4N^{0.8}}{N^3 - 19.53^3} \tag{3}$$

For very long chains $d_\alpha$ appears to reach a plateau value of 5.82 Å.

**TABLE 3: Experimental Results Collected from Literature for the Stokes Radii of 20 Denatured Polypeptides**

| protein | $N$ | $R_S$ (Å) | solvent condition | ref |
|---|---|---|---|---|
| Analytical Ultracentrifuge[a] | | | | |
| ribonuclease A | 124 | 30.1 | 6 M GuHCl, 25 °C | 26 |
| hemoglobin | 144 | 32.7 | 6 M GuHCl, 25 °C | 26 |
| myglobin | 153 | 35.5 | 6 M GuHCl, 25 °C | 26 |
| immunoglobulin (fragment I) | 230 | 44.7 | 6 M GuHCl, 25 °C | 26 |
| chymotrypsinogen A | 245 | 45.8 | 6 M GuHCl, 25 °C | 26 |
| G3P dehydrogenase | 331 | 48.7 | 6 M GuHCl, 25 °C | 26 |
| pepsinogen | 376 | 50.0 | 6 M GuHCl, 25 °C | 26 |
| aldolase | 370 | 52.4 | 6 M GuHCl, 25 °C | 26 |
| immuloglobulin | 372 | 55.1 | 6 M GuHCl, 25 °C | 26 |
| serum albumin | 605 | 70.6 | 6 M GuHCl, 25 °C | 26 |
| myosin | 1739 | 121.3 | 6 M GuHCl, 25 °C | 26 |
| Dynamic Light Scattering | | | | |
| horse apocytochrome *c* | 104 | 30.0 | pH 2.3 | 27 |
| ribonuclease A | 124 | 31.4 | 6 M GuHCl | 21 |
| phosphoglycerate kinase | 415 | 56.6 | 6 M GuHCl | 28 |
| Pulse Field Gradient NMR[b] | | | | |
| lysozyme (49−64) | 16 | 10.0 | 8 M urea, pH 2 | 18 |
| fibronectin binding protein D3 (17−38) | 22 | 12.8 | 8 M urea | 18 |
| fibronectin binding protein D3 (7−38) | 32 | 14.9 | 8 M urea | 18 |
| BPTI | 58 | 22.7 | pH 2.5 | 17 |
| SH3 domain of PI3 kinase | 90 | 26.5 | 3.5 M GuHCl | 18 |
| lysozyme | 129 | 33.3 | 8 M urea, pH 2 | 18 |
| triosephosphate isomerase | 247 | 47.9 | 2.5 M GuHCl | 18 |

[a] $R_S$ was obtained from the sedimentation coefficient measured in ref 26 via $s^0 = M(1 - \upsilon\rho)/6\pi\eta_0 N_A R_S$, where $N_A$ is the Avogadro number. Values for the molecular weight $M$ and the partial specific volume $\upsilon$ of the proteins are listed in ref 26. The solution density was taken to be $\rho = 1.14$ g/cm$^3$ and the solvent viscosity was taken to be 1.0.45 cP.[29] An additional data point ($\beta$-lactoglobulin with $N = 162$ and $R_S = 28.7$ Å) falls too far out of line from the rest of the data and is not included. [b] All values from Table 1 of ref 18 were scaled down by 1.038; see footnote *b* of Table 1.

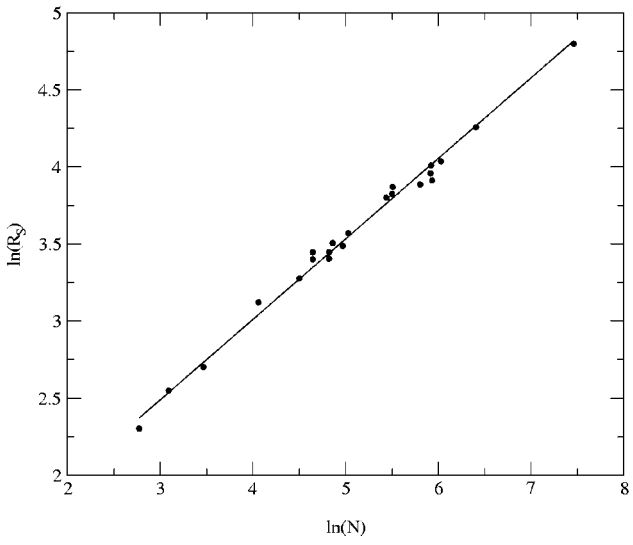


**Figure 3.** Fit of experimental data on Stokes radius to eq 2.

**TABLE 4: Reproduction of Experimental Intrinsic Friction [$\eta$] and Stokes Radius $R_s$ and the Resulting Hard-Sphere Diameter $d_\alpha$, Mean Square $\langle R^2 \rangle$ of the End-to-End Distance, Effective Bond Length $b_{eff}$, and $\Phi$ and $K$ Coefficients**

| | | [$\eta$] (cm$^3$/g) | | $R_S$ (Å) | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $N$ | $d_\alpha$ (Å) | calc[a] | eq 1 | calc[a] | eq 2 | exp[b] | $\langle R^2 \rangle$ (Å$^2$) | $b_{eff}$ (Å) | $\Phi$[c] | $K$[d] |
| 16 | 0 | 4.83 ± 0.04 | 4.38 | 10.85 ± 0.02 | 10.71 | 10.0 | 488.8 | 5.71 | 0.785 | 0.491 |
| 22 | 2.24 | 5.46 ± 0.04 | 5.40 | 12.52 ± 0.03 | 12.64 | 12.8 | 769.2 | 6.05 | 0.619 | 0.451 |
| 27 | 4.12 | 6.25 ± 0.05 | 6.19 | 14.00 ± 0.03 | 14.07 | | 1152.5 | 6.66 | 0.474 | 0.412 |
| 32 | 4.58 | 6.85 ± 0.05 | 6.92 | 15.25 ± 0.03 | 15.37 | 14.9 | 1498.4 | 6.95 | 0.416 | 0.394 |
| 37 | 4.87 | 7.67 ± 0.06 | 7.62 | 16.56 ± 0.04 | 16.58 | | 1879.2 | 7.22 | 0.383 | 0.382 |
| 42 | 5.05 | 8.29 ± 0.07 | 8.28 | 17.70 ± 0.04 | 17.72 | | 2273.5 | 7.45 | 0.353 | 0.371 |
| 47 | 5.20 | 8.91 ± 0.07 | 8.92 | 18.81 ± 0.04 | 18.79 | | 2702.4 | 7.66 | 0.328 | 0.362 |
| 52 | 5.34 | 9.60 ± 0.09 | 9.54 | 19.88 ± 0.04 | 19.81 | | 3162.8 | 7.88 | 0.309 | 0.353 |
| 58 | 5.46 | 10.28 ± 0.09 | 10.25 | 21.05 ± 0.05 | 20.97 | 22.7 | 3699.8 | 8.06 | 0.291 | 0.346 |
| 62 | 5.50 | 10.80 ± 0.10 | 10.71 | 21.85 ± 0.05 | 21.71 | | 4026.7 | 8.12 | 0.288 | 0.344 |
| 67 | 5.50 | 11.44 ± 0.11 | 11.27 | 22.80 ± 0.06 | 22.61 | | 4494.8 | 8.25 | 0.280 | 0.340 |
| 72 | 5.57 | 11.95 ± 0.12 | 11.82 | 23.66 ± 0.06 | 23.47 | | 5047.7 | 8.43 | 0.264 | 0.333 |
| 76 | 5.59 | 12.31 ± 0.12 | 12.25 | 24.33 ± 0.06 | 24.15 | | 5351.5 | 8.45 | 0.263 | 0.333 |
| 82 | 5.59 | 12.89 ± 0.16 | 12.88 | 25.33 ± 0.07 | 25.12 | | 5821.5 | 8.48 | 0.262 | 0.332 |
| 90 | 5.66 | 13.84 ± 0.13 | 13.70 | 26.71 ± 0.06 | 26.37 | 26.5 | 6495.7 | 8.54 | 0.262 | 0.331 |
| 104 | 5.68 | 15.18 ± 0.18 | 15.07 | 28.86 ± 0.08 | 28.44 | 30.0 | 7934.4 | 8.78 | 0.245 | 0.324 |

[a] Calculated from conformations sampled with the hard-sphere diameter given in the second column. [b] Taken from Table 3. [c] Calculated from [$\eta$] and $\langle R^2 \rangle$ according to eq 6. [d] Calculated from $R_s$ and $\langle R^2 \rangle$ according to eq 8.

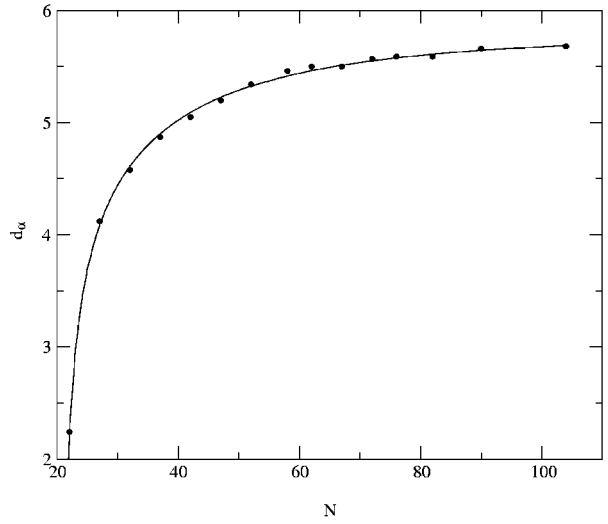


**Figure 4.** Fit of the chain length dependence of the $C_\alpha$ diameter to eq 3.

The zero $d_\alpha$ value for the $N = 16$ chain means that this chain does not experience any nonlocal volume exclusion. That is, it is in the "unperturbed" state. We justify the zero $d_\alpha$ value at $N = 16$ on two considerations. First, as Table 4 and Figure 4 show, there clearly is a need for using smaller $d_\alpha$ values for shorter chain lengths. Second, even with $d_\alpha = 0$, the calculated intrinsic viscosity and Stokes radius are somewhat larger than what are expected from eqs 1 and 2.

The long CPU time it took to generate conformations with the excluded-volume effects explicitly accounted for prevented us from studying longer chains. Out of all possible conformations, the fraction $F_{ex}$ that does not experience volume exclusion decreases rapidly as the chain length $N$ increases (see Table 2). We found $-\ln F_{ex} \approx 0.171N - 3.68$. For $N = 104$, the fraction that survives volume exclusion is less than 1 in a million conformations. The sharp decrease in $F_{ex}$ reflects both the increased chances of residues coming into contact as the chain length increases and the increase in the hard-sphere diameter $d_\alpha$.

**Simultaneous Reproduction of the Stokes Radius.** Are the conformations generated to reproduce [$\eta$] realistic? Fortunately, this question can be addressed to some extent by checking the calculated Stokes radius against eq 2. As can be seen from Table 4, the calculated values of $R_S$ do agree closely with eq 2. The small differences are certainly within the fluctuations of the experimental data from which eq 2 is derived. This simultaneous reproduction of the experimental data on $R_S$ gives credence to the chain conformations we generated.

**Influence of Database Composition on Calculated Hydrodynamic Properties.** The chain conformations generated to calculate the intrinsic viscosity and Stokes radius depend on the distribution of ($\phi$, $\psi$) angles assembled from protein loops. This distribution may, in particular, be biased by the amino acid composition of the assembled loop residues. As the amino acid compositions of proteins vary somewhat, it is of interest to find out how much this composition variation influences the calculated results. This issue was explored on the 58-residue BPTI chain. Sampling ($\phi$, $\psi$) angles with the BPTI composition of amino acids, the intrinsic viscosity and Stokes radius for a 58-residue chain were calculated to be 10.56 ± 0.09 cm$^3$/g and 21.16 ± 0.05 Å, respectively. These are to be compared to the values listed in Table 4, [$\eta$] = 10.28 ± 0.09 cm$^3$/g and $R_S$ = 21.05 ± 0.05 Å. There appears to be a small but detectable influence on the calculated hydrodynamic properties by the composition of amino acids in the sampling of ($\phi$, $\psi$) angles.

We also checked whether the specific sequence of the amino acids (not just the composition) exerts an influence on the calculated [$\eta$] and $R_s$. Using the BPTI sequence, we generated the position for each residue using only ($\phi$, $\psi$) angles assembled for that particular type of residue. The results for the intrinsic viscosity and the Stokes radius calculated from this procedure are 10.61 ± 0.09 cm$^3$/g and 21.23 ± 0.04 Å, respectively. The influence of the sequence (over and above that of the composition) does not appear to be significant.

**Distribution of End-to-End Distance.** The distributions of the end-to-end distance $R$ are found to be nearly Gaussian when $N > 30$. As Figure 5a shows, these distributions can be reasonably fitted to a function

$$p(R) = c_0 R^{c_1} \exp(-R^{c_2}/c_3) \tag{4}$$

At $N = 90$, the exponents are $c_1 = 2.32$ and $c_2 = 2.88$, both somewhat larger than the value 2 of the Gaussian distribution. For $N = 32$, we found $c_1 = 2.55$ and $c_2 = 3.29$.

The average of $R^2$ and the effective bond length $b_{eff} = [\langle R^2 \rangle/(N-1)]^{1/2}$ are shown in Table 4. The effective bond

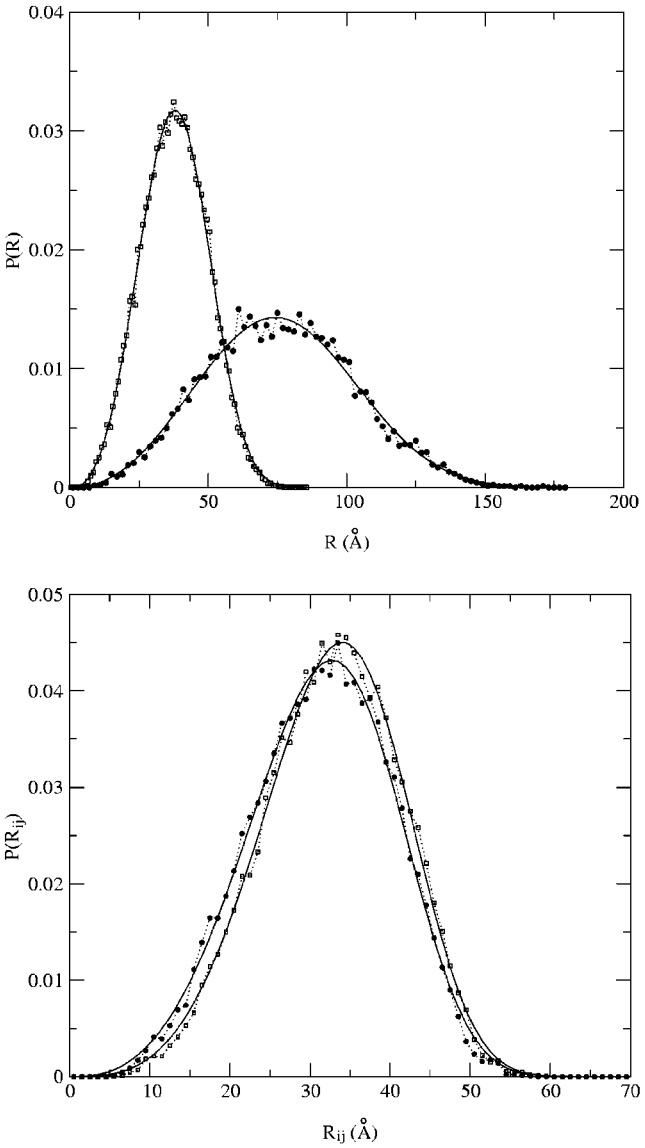


**Figure 5.** Distributions of residue−residue distances. The fit to eq 4 is shown as solid curves. (a) End-to-end distances for the $N = 90$ (filled circles) and $N = 32$ (open squares) chains. (b) Distances between pairs of residues with a sequence separation of 20 in the $N = 90$ chain. Filled circles are for the distances between residues 1 and 21 and between residues 70 and 90, open squares are for distances between residues 35 and 55 and between residues 36 and 56.

length of the "unperturbed" chain, for which there is no volume exclusion (i.e., $d_\alpha = 0$), was found to have a weak $N$ dependence: $b_{\text{eff}}^0 = 6.0[1 - 1.52/(N-1)]^{0.5}$ Å.[31] The expansion factor $\alpha = b_{\text{eff}}/b_{\text{eff}}^0$ ranges from 1 at $N = 16$ to 1.47 at $N = 104$.

Many theories have been developed to calculate the expansion factor.[30] Particularly appealing is the original theory of Flory,[32] which argues that the size of a polymer is determined by the balance of two effects: excluded-volume interactions which tend to swell the polymer, and elastic energies arising from the chain connectivity which tend to shrink the polymer. The result is[33,34]

$$\alpha^5 - \alpha^3 = \frac{N^{1/2}\upsilon}{2(b_{\text{eff}}^0)^3} \tag{5}$$

where $\upsilon = \gamma d_\alpha^3$ is the exclusion volume of each residue ($\gamma$ is an adjustable parameter with a value around 1). Although eq 5 was originally derived for very long chains, we found it reasonably models the expansion factors for the short chains

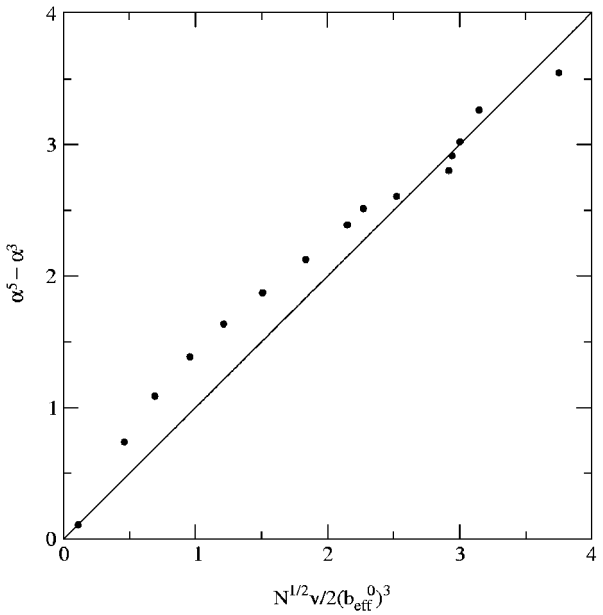


**Figure 6.** Comparison of the two sides of eq 5. The abscissa shows $N^{1/2}\upsilon/2(b_{\text{eff}}^0)^3$, in which $N$ ranges from 16 to 104, $\upsilon = 0.8 d_\alpha^3$ ($d_\alpha$ for a given $N$ is shown in Table 4), and $b_{\text{eff}}^0 = 6.0[1 - 1.52/(N-1)]^{0.5}$ Å. The ordinate shows $\alpha^5 - \alpha^3$, in which $\alpha = b_{\text{eff}}/b_{\text{eff}}^0$ ($b_{\text{eff}}$ for a given $N$ is shown in Table 4). The diagonal line is obtained if eq 5 is exact.

with $N = 16$ to 104 residues studied here. As shown in Figure 6, with $\gamma = 0.8$, eq 5 agrees well with the calculation results for $N$ between 16 and 104. The significant $N$ dependence of the excluded volume ($\sim d_\alpha^3$) (see Figure 4) (along with a milder $N$ dependence of $b_{\text{eff}}^0$) may partly explain the applicability of eq 5 to the short chains.

**Distribution of Distances between Interior Residues.** The distance between two nearest residues is fixed at $b = 3.804$ Å. Thus we expect that the effective bond length characterizing the distance between two residues $i$ and $i + n$ will gradually increase from $b$ to $b_{\text{eff}}$ as $n$ increase from 1 to $N - 1$. This is indeed what was observed in the chain conformations we generated.

Figure 5b shows the distributions of distances between two residues with $n = 20$ in the $N = 90$ chain. The distribution for the distance between residues 1 and 21 is distinctly different from the distribution for the distance between two residues, 35 and 55, which have the same separation along the chain but are more interior. The two effective bond lengths [calculated as $(\langle R_{i,\, i+n}^2\rangle/n]^{1/2}$, are 7.28 and 7.57 Å. The increase in effective bond length as the residue pair is moved to the interior is a manifestation of the fact that the interior pair encounters stronger excluded-volume effects.[34] Moreover, these effective bond lengths for two residues separated by 20 residues in the $N = 90$ chain are substantially larger than the effective bond length for the end-to-end distance of the $N = 21$ chain (see Table 4). In contrast, for chains in the unperturbed state, the effective bond length for a given sequence separation remains the same, regardless of whether the distance is measured between a beginning or an interior pair of residues in a long chain or between the two end residues of a short chain.

**Relation of $R_S$ and $[\eta]$ to $\langle R^2\rangle$.** From a dimensional argument, one may write

$$[\eta]M = \Phi \langle R^2\rangle^{3/2} \tag{6}$$

where $M = NM_0$ is the molecular weight of the whole chain. For a long chain in the unperturbed state, $\langle R^2\rangle \propto N$, and $\Phi$ is a

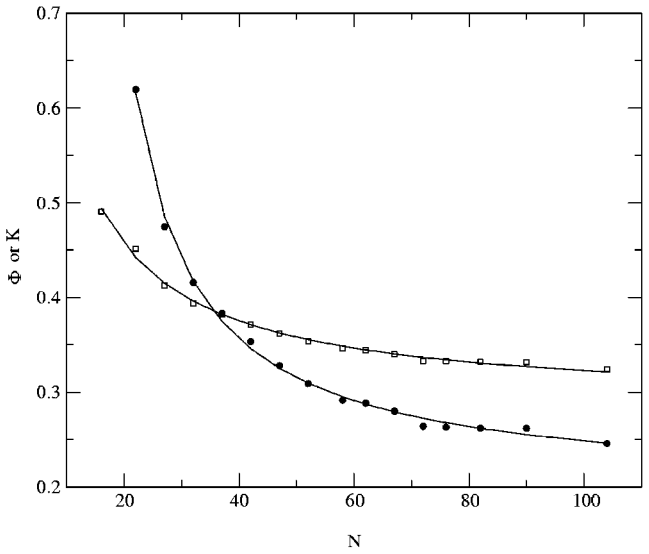


**Figure 7.** Fit of the chain length dependence of Φ and *K* to eqs 7 and 9. Filled circles and open squares represent Φ and *K*, respectively (values are listed in Table 4).

constant. This constant is predicted to be 0.256 by the Zimm model.[30] The *N* dependence of measured $[\eta]$ demonstrates that denatured proteins are not in the unperturbed state. For denatured proteins, Φ is not a constant; rather it shows *N* dependence. The values calculated from $[\eta]$ and $\langle R^2\rangle$ for *N* between 16 and 104 according to eq 6 are shown in Table 4. Φ decreases from 0.785 at $N = 16$ to 0.246 at $N = 104$. The *N* dependence of Φ can be fitted to (see Figure 7)

$$\Phi = 0.1962\,\frac{N + 12.80}{N - 10.92} \tag{7}$$

This indicates $\Phi \rightarrow 0.1962$ as $N \rightarrow \infty$. Similarly, we may write

$$R_S = K\,\langle R^2\rangle^{1/2} \tag{8}$$

For a long chain in the unperturbed state, *K* is a constant. This constant is predicted to be 0.271 by the Zimm model.[30] Again, the *N* dependence of $R_s$ in eq 2 demonstrates the denatured proteins are not in the unperturbed state. The values of *K* for *N* between 16 and 104, calculated from $R_s$ and $\langle R^2\rangle$ according to eq 8, are shown in Table 4. *K* decreases from 0.491 at $N = 16$ to 0.324 at $N = 104$. The *N* dependence can be fitted to (see Figure 7)

$$K = 0.2863\,\frac{N + 14.93}{N + 1.932} \tag{9}$$

indicating $K \rightarrow 0.2863$ as $N \rightarrow \infty$.

## Discussion

We have carried out hydrodynamic calculations on denatured protein chains with excluded volumes explicitly considered. By sampling $(\phi, \psi)$ angles from those collected from loop regions in the protein structure database and adjusting the hard-sphere diameter $d_\alpha$ of the $C_\alpha$ atoms, the experimental intrinsic viscosity and Stokes radius are simultaneously reproduced for chains with 16 to 104 residues. The distribution of the end-to-end distance *R* and its mean square $\langle R^2\rangle$ were obtained from the sampled chain conformations. When the chain length dependence of the intrinsic viscosity and the Stokes radius is fitted to the Zimm model (which predicts $[\eta]M = \Phi\langle R^2\rangle^{3/2}$ and $R_s = K\langle R^2\rangle^{1/2}$), the length dependence of the coefficients Φ and *K* is obtained. The effective bond length $b_{eff}$ is found to range from 5.71 to 8.78 Å, and the end-to-end distances of protein chains are found to have nearly Gaussian distributions when $N > 30$.

Hydrodynamic theories are well developed for polymers in the unperturbed state and with infinitely long chains, and often hydrodynamic data are analyzed according to these theories. Our calculations show that, for most proteins (with tens to a few hundred residues), this procedure is not justified for dealing with the denatured state. In particular, eqs 7 and 9 show that the Φ and *K* coefficients, usually assumed to be constants, still have strong *N* dependence for $N > 100$. Thus, the use of ideal model equations, such as eqs 6 and 8 with constant coefficients, will give misleading information for the dimensions of denatured proteins.

**Extrapolation to Longer Chains.** The calculation results presented in this paper are for chains with 16 to 104 residues. We may extrapolate these results in order to obtain size information for longer chains. There are three routes to this end, using combinations of eqs 5 and 3, of eqs 6, 7, and 1, and of eqs 8, 9, and 2, respectively. The three routes give consistent results for the root-mean-square end-to-end distance for up to $N = 1500$. The consistency suggests that eq 5 may be used for calculating the dimensions of denatured chains at any lengths. For long chains the end-to-end distance is approximately given by $\langle R^2\rangle^{1/2} = 5.31N^{0.6}$. The effective bond length is 10.6 Å at $N = 1000$.

**Physical Meaning of $C_\alpha$ Hard-Sphere Diameter.** The $C_\alpha$ hard-sphere diameter $d_\alpha$ models the balance between residue−residue and residue-solvent interactions. It should thus depend on solvent conditions such as temperature and denaturant concentration. For example, as the denaturant concentration is decreased, residue-solvent interactions become less favorable. This can be modeled by a decreased $d_\alpha$. The dimension of the chain will decrease as a result. Such as decrease is consistent with the expected compaction of the chain as the denaturant concentration is lowered.

The strong dependence of the hard-sphere diameter $d_\alpha$ on chain length found in the range $20 < N < 80$ is somewhat surprising. There are two potential sources for this dependence. First, we assumed that $(\phi, \psi)$ angles of a denatured protein chain sample from those collected from loop regions in the structure database. In reality, the $(\phi, \psi)$ of a denatured protein could well be different from that of loop regions. Second, in the present model, we try to account for volume exclusion by the $C_a-C_a$ distance. In reality, side chains play a major role in excluded-volume interactions of a polypeptide chain. Local excluded-volume interactions are built into the $(\phi, \psi)$ angles. This perhaps explains why for the 16-residue chain no $C_a-C_a$ cutoff distance needs to be invoked. For a longer chain, nonlocal excluded-volume interactions (involving side chains far separated along the sequence) give rise to expansion of the chain that, within the present model, can only be represented by an increase in the hard-sphere diameter.

**Unperturbed State.** In this study, the unperturbed state is determined by the $(\phi, \psi)$ map from which the chain conformations are generated. The main justification is that we can simultaneously reproduce the experimental intrinsic viscosity and Stokes radius from such conformations. The effective bond length $b_{eff}^0$ of the unperturbed state approaches 6 Å for long chains.

This value of $b_{eff}^0$ is considerably smaller than a value of ~11.5 Å determined experimentally[7,35] and calculated within

the rotational isomeric state theory for homo-polypeptides by Flory and co-workers,[6,7] and 18% smaller than a value of 7.1 Å determined experimentally for proteins with $N$ ranging from 124 to 605 by Lapanje and Tanford.[36] Commenting on the difference in $b_{\text{eff}}^0$ between the two groups, Lapanje and Tanford stated:

> Neither the theoretical nor the experimental results of Flory and co-workers have an absolute accuracy superior to the results presented here. Their experimental data, like ours, were obtained under nonideal conditions, and corrected to ideal conditions by use of the second virial coefficient, and this procedure, as we have shown here, involves considerable uncertainty. Their theoretical calculations involve parameters which are not known with great precision and which, to some extent, were adjusted to fit the experimental data.

These statements hold true when comparing the previous two values of $b_{\text{eff}}^0$ with the one determined in the present study.

The much higher value calculated by Flory and co-workers can be traced to the fact that they used an empirical energy function which overwhelmingly favors the $\beta$ region of the ($\phi$, $\psi$) map (with a probability of occupation at 93.4%).[5] It is not surprising that a polypeptide with nearly all ($\phi$, $\psi$) angles in the $\beta$ region has a long end-to-end distance. In contrast, in our ($\phi$, $\psi$) map (see Figure 2), the occupation probabilities of the $\beta$ and $\alpha$ regions are nearly equal (at 49% and 40%, respectively; the rest is for the region with positive $\phi$). By design, our ($\phi$, $\psi$) map mimics the tendency of residues in protein loops. Ten percent of the loop residues which were used to generate our ($\phi$, $\psi$) map are glycine. As shown by Miller et al.,[7] the presence of glycine significantly reduces the effective bond length (their $b_{\text{eff}}$ becomes 9.7 Å at 10% glycine). Further decrease in $b_{\text{eff}}$ may occur in the presence of proline[8] (present in our population of loop residues at 9%).

The problem with the procedure used by Lapanje and Tanford (and by Flory and co-workers as well) for data analysis can be highlighted by the fact that the expansion factor $\alpha$ obtained by them is a constant $\sim$1.34 for proteins with chain lengths varying from 124 to 605 residues. This does not appear to be a self-consistent result. Given that this constant value of $\alpha$ differs from unity by so much, polymer theory would predict a significant increase in $\alpha$ when $N$ increases 5-fold. For example, eq 5 predicts a 15% increase.[37]

The uncertainty in $b_{\text{eff}}^0$ is actually not a cause for great concern. The unperturbed state provides a convenient reference state but otherwise not particularly relevant for the polypeptide chain in most solvent conditions. For proteins chains fully denatured in 6 M GuHCl or under other conditions, the effective bond lengths given in Table 4 should be reliable. Extrapolation to long chains gives $\langle R^2\rangle^{1/2} = 5.31N^{0.6}$.

**Acknowledgment.** This work is supported in part by NIH Grant No. GM58187.

## References and Notes

(1) Hagen, S. J.; Hofrichter, J.; Eaton, W. A. *J. Phys. Chem. B* **1997**, *101*, 2352.
(2) Dill, K. A.; Stigter, D. *Adv. Protein Chem.* **1995**, *46*, 59.
(3) (a) Zhou, H.-X. *J. Phys. Chem. B* **2001**, *105*, 6763. (b) Zhou, H.-X. *J. Am. Chem. Soc.* **2001**, *123*, 6730. (c) Zhou, H.-X. *Biochemistry* **2001**, *40*, 15 069.
(4) (a) Zhou, H.-X.; Dill, K. A. *Biochemistry* **2001**, *40*, 11 289. (b) Zhou, H.-X. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 3569.
(5) Flory, P. J. *Statistical Mechanics of Chain Molecules*; Wiley: New York, 1969.
(6) Brant, D. A.; Flory, P. J. *J. Am. Chm. Soc.* **1965**, *87*, 2791.
(7) Miller, W. G.; Brant, D. A.; Flory, P. J. *J. Mol. Biol.* **1967**, *23*, 67.
(8) Miller, W. G.; Groebel, C. V. *Biochemistry* **1968**, *7*, 3925.
(9) Engh, R.; Huber, R. *Acta Crystallogr. A* **1991**, *47*, 392.
(10) Holm, L.; Sander, C. *Nucl. Acids Res.* **1997**, *25*, 3389.
(11) Fiebig, K. M.; Schwalbe, H.; Buck, M.; Smith, L. J.; Dobson, C. M. *J. Phys. Chem.* **1996**, *100*, 2661.
(12) Pappu, R. V.; Srinivasan, R.; Rose, G. D. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 12 565.
(13) Zhou, H.-X. *Biophys. J.* **1995**, *69*, 2286.
(14) Zhou, H.-X. *Biophys. J.* **1995**, *69*, 2298.
(15) Zhou, H.-X. *Biophys. Chem.* **2001**, *93*, 171.
(16) Note that there is no relation between $d_\alpha$ and $r_h$. The former depends on chain length and solvent condition; the latter is a constant. Included in $r_h$ is a solvation shell.[14]
(17) Pan, H.; Barany, G.; Woodward, C. *Protein Sci.* **1997**, *6*, 1985.
(18) Wilkins, D. K.; Grimshaw, S. B.; Receveur, V.; Dobson, C. M.; Jones, J. A.; Smith, L. J. *Biochemistry* **1999**, *38*, 16 424.
(19) Gast, K.; Zirwer, D.; Muller-Frohne, M.; Damaschun, G. *Protein Sci.* **1998**, *7*, 2004.
(20) Buzzell, J. G.; Tanford, C. *J. Phys. Chem.* **1956**, *60*, 1204.
(21) Noppert, A.; Gast, K.; Muller-Frohne, M.; Zirwer, D.; Damaschun, G. *FEBS Lett.* **1996**, *380*, 179.
(22) Luzzati, V.; Witz, J.; Nicolaieff, A. *J. Mol. Biol.* **1961**, *3*, 367.
(23) Sophianopoulos, A. J.; Rhodes, C. K.; Holcomb, D. N.; van Holde, K. E. *J. Biol. Chem.* **1962**, *237*, 1107.
(24) Wyman, J.; Ingalls, E. N. *J. Biol. Chem.* **1943**, *147*, 297.
(25) Tanford, C. *Adv. Protein Chem.* **1968**, *23*, 121.
(26) Tanford, C.; Kawahara, K.; Lapanje, S. *J. Am. Chem. Soc.* **1967**, *89*, 729.
(27) Damaschun, G.; Damaschun, H.; Gast, K.; Gernat, C.; Zirwer, D. *Biochim. Biophys. Acta* **1991**, *1078*, 289.
(28) Damaschun, G.; Damaschun, H.; Gast, K.; Zirwer, D. *Biochemistry (Moscow)* **1998**, *63*, 259.
(29) Kawahara, K.; Tanford, C. *J. Biol. Chem.* **1966**, *241*, 3228.
(30) (a) Zimm, B. H. *J. Chem. Phys.* **1956**, *24*, 269. (b) Zimm, B. H.; Roe, G. M.; Epstein, L. F. *J. Chem. Phys.* **1956**, *24*, 279. (c) Hearst, H. E. *J. Chem. Phys.* **1962**, *37*, 2547. (d) Doi, M.; Edwards, S. F. *The Theory of Polymer Dynamics*; Clarendon Press: Oxford, 1986.
(31) This result was obtained by fitting data for unperturbed chains with $11 \leq N \leq 151$.
(32) Flory, P. J. *J. Chem. Phys.* **1949**, *17*, 303.
(33) Sanchez, I. C. *Macromolecules* **1979**, *12*, 980.
(34) Chan, H. S.; Dill, K. A. *Annu. Rev. Biophys. Biophys. Chem.* **1991**, *20*, 447.
(35) Brant, D. A.; Flory, P. J. *J. Am. Chm. Soc.* **1965**, *87*, 2788.
(36) Lapanje, S.; Tanford, C. *J. Am. Chem. Soc.* **1967**, *89*, 5030.
(37) An alternative procedure proposed by Lapanje and Tanford[36] is to use an empirical formula for the second virial coefficient: $A_2 = 0.7151(N-1)^{3/2}(b_{\text{eff}}^0)^3M^{-2}(\alpha^2-1)$, which was found by G. C. Berry (*J. Chem. Phys.* **1966**, *44*, 4550) to reproduce experimental data for polystyrene over a wide range of conditions. Using our value for $b_{\text{eff}}^0$ and eq 5 for the expansion factor, we predict the following results for the second virial coefficients of the six proteins studied by Lapanje and Tanford (in units of $10^{-3}$ mol cm$^3$/g$^2$): 1.29, 0.97, 1.24, 1.19, 0.97, and 0.81. These are in reasonable agreement with the experimental data: $1.16 \pm 0.06$, $1.08 \pm 0.09$, $1.03 \pm 0.03$, $1.17 \pm 0.11$, $0.87 \pm 0.07$, and $0.75 \pm 0.09$.