NSRRC - J0 48 74 (2015)

THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

Random-Coil Behavior of Chemically Denatured Topologically Knotted Proteins Revealed by Small-Angle X-ray Scattering

Po-Min Shih,[†,‡] Iren Wang,[†] Yun-Tzai Cloud Lee,[†,§] Shu-Ju Hsieh,[†,‡] Szu-Yu Chen,[†] Liang-Wei Wang,[†,‡,§] Chih-Ting Huang,[†,‡] Chih-Ta Chien,[†,∥] Chia-Yun Chang,[†,§] and Shang-Te Danny Hsu*[,†,‡,§]

[†]Institute of Biological Chemistry, Academia Sinica, 128, Section 2, Academia Road, Taipei 11529, Taiwan

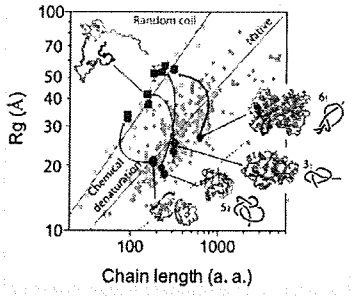
[‡]Institute of Bioinformatics and Structural Biology, National Tsing Hua University, 101, Section 2, Kuang-Fu Road, Hsinchu 30013, Taiwan

[§]Institute of Biochemical Sciences, National Taiwan University, 1, Section 4, Roosevelt Road, Taipei 106, Taiwan

[∥]Department of Chemistry, National Taiwan University, 1, Section 4, Roosevelt Road, Taipei 106, Taiwan

Supporting Information

ABSTRACT: Recent studies on the mechanisms by which topologically knotted proteins attain their natively knotted structures have intrigued theoretical and experimental biophysicists. Of particular interest is the finding that YibK and YbeA, two small trefoil knotted proteins, remain topologically knotted in their chemically denatured states. Using small-angle X-ray scattering (SAXS), we examine whether these chemically denatured knotted proteins are different from typical random coils. By revisiting the scaling law of radius of gyration (R_g) as a function of polypeptide chain length for chemically denatured proteins and natively folded proteins, we find that the chemically denatured knotted proteins in fact follow the same random coil-like behavior, suggesting that the formation of topological protein knots do not necessarily require global compaction while the loosely knotted polypeptide chains are capable of maintaining the correct chirality without defined secondary or tertiary structures.



INTRODUCTION

Systematic surveys of protein structure database have recently identified hundreds of proteins that contain topologically knotted backbone structures.[1−4] These intricate protein topologies range from the simplest trefoil (3_1) knot to more complex figure-eight (4_1) knots, Gordian (5_2) knots, and even a Stevedore's (6_1) knot.[5] Many experimental and theoretical studies have been reported to characterize the folding dynamics and kinetic pathways of a number of knotted proteins. Computational analyses have shown that the knotting events, which may involve the formation of a slipknot, are the rate-limiting step for the folding of knotted proteins into their native conformations and that back tracking of the misfolded slipknot due to incorrect looping chirality could be implicated in the slow folding kinetics and low success rate; only 1 to 2% of the unbiased simulations managed to attain correctly knotted structures.[6−11] In contrast, experimental folding analyses on a number of knotted proteins have shown that many of these topologically knotted proteins can refold spontaneously after chemical denaturation without the aid of auxiliary factors such as molecular chaperones.[12−21] Furthermore, without exception, all of the knotted proteins that have been investigated thus far exhibit kinetic folding intermediates along their folding pathways, which can be parallel, bifurcated, or linear.[13,15]

By engineering two cysteine residues at the N- and C-termini of YibK from *H. influenzae* and YbeA from *E. coli* that contain a deep trefoil knot in their structures and cross-linking them through disulfide bond formation under chemically denaturing conditions, followed by removal of chemical denaturants, this resulted in efficient refolding into their topologically knotted states. Jackson and coworkers provided the strong experimental evidence to suggest that YibK and YbeA remain topologically knotted under chemically denaturing conditions,[12] despite the lack of appreciable secondary and tertiary structures according to far-UV circular dichroism (CD), intrinsic fluorescence, and nuclear magnetic resonance (NMR) spectroscopy.[14,15,22,23] This raises the question as to whether the structural motifs are responsible for maintaining the loosely knotted conformation of these two proteins? Moreover, do these loosely knotted structures adopt any preferential conformations similar to those observed in intrinsically disordered proteins (IDPs)?

To address these questions, we employed synchrotron-based SAXS[24] to investigate the structures of a number of topologically knotted proteins under native and denaturing conditions. SAXS is a robust tool for determining the solution structures of biomolecules.[25−27] It has been used to characterize the dimension of chemically denatured proteins through the determination of their radii of gyration (R_g).[28] The results showed that chemically denatured proteins behave like random coils whose R_g values scale with the chain lengths with a scaling

Received: February 28, 2015
Revised: April 6, 2015
Published: April 7, 2015

ACS Publications © 2015 American Chemical Society

DOI: 10.1021/acs.jpcb.5b01984
J. Phys. Chem. B 2015, 119, 5437−5443

exponent of 3/5.[29] In the case of IDPs, which contain residual structure with varying degrees of compaction, their R_g values are generally smaller than those of chemically denatured proteins of similar chain lengths.[30] We therefore anticipate smaller R_g values of chemically denatured knotted proteins should the presence of loosely knotted elements result in significant degree of structural compaction.

MATERIALS AND METHODS

Compilation of Experimental SAXS Data of Chemically Denatured and Native Proteins for Empirical Scaling Relationship of R_g as a Function of Protein Chain Lengths. Literature search using the keyword combinations of "SAXS protein" or "small angle X-ray scattering protein" was carried out using the database PubMed (http://www.ncbi.nlm.nih.gov/pubmed/). About 700 related reports were identified from the database as of December 2013, of which about 200 entries contain experimentally determined R_g values of different proteins. These reports were examined individually to obtain the protein chain lengths, experimental conditions (native or denaturing, temperature, and pH) and oligomeric states. These manual analyses yielded 37 entries for chemically denatured proteins, including the 26 entries that were compiled by Plaxco and coworkers[28] and 196 entries for native proteins, of which 129 entries correspond to native single-chain proteins and 67 entries correspond to oligomeric proteins (Tables S1−S3 in the SI). We also included 30 R_g values that were deposited in the BIOISIS database (http://www.bioisis.net/welcome) as of December 2013 for native proteins; nevertheless, many of these experimental SAXS data correspond to structural genomics targets whose exact structures and oligomeric states are not available in the literature (Table S4 in the SI). The R_g values of individual groups of proteins, namely, chemically denatured proteins, all native proteins, monomeric and oligomeric native proteins, were subject to nonlinear regression to extract the scaling exponent ν and prefactor R_0 as a function of protein chain length, L (amino acid number, a.a.), according to the Flory's equation[29]

$$R_g = R_0 L^{\nu}$$

The nonlinear regression was carried out using the software package PRISM (Graphpad Software, USA) and outliers were selected using default settings.

Purification of Recombinant Proteins for SAXS Measurements. Recombinant YibK,[31] YbeA,[15] HP0242 from *H. pylori*,[32] human UCH-L1,[17] and UCH-L5 NTD (residues 1−240 in the N-terminal ubiquitin C-terminal hydrolyase domain)[33] and DehI from *P. putida*[34] are overexpressed in the *E. coli* BL21(DE3) strain and purified following the previously described protocols. The plasmids of YibK, YbeA, and UCH-L1 were kind gifts from Dr. Sophie Jackson, University of Cambridge, U.K.; the plasmid of HP0242 was a kind gift from Dr. Ping-Chiang Lyu, National Tsing Hua University, Taiwan; the plasmid of UCH-L5 NTD was a kind gift from Dr. Chittaranjan Das, Purdue University, U.S.A.; the plasmid of DehI was a kind gift from Matthew C. J. Wilce, Monash University, Australia. After purifications, all samples were buffer-exchanged by size exclusion chromatography (SEC) for long-term storage. YibK and YbeA were buffered in 50 mM Tris-HCl (pH 7.5), 200 mM KCl, 10% glycerol. HP0242 and $HP0242_{x2}$ were buffered in 10 mM phosphate (pH 6.5). MJ0366 was buffered in 50 mM potassium phosphate (pH 6.0) and 50 mM sodium chloride. UCH-L1 and UCH-L5 NTD were buffered in 20 mM phosphate (pH 7.6). DehI was buffered in 20 mM potassium phosphate (pH 8.0), 300 mM sodium chloride. All samples contained 1−10 mM DTT to keep the sample in reduced states and to minimize radiation damages.

The DNA sequences that correspond to MJ0366 from *M. jannaschii* and HP0242 tandem repeat ($HP0242_{x2}$) were synthesized and codon-optimized for *E. coli* expression (Genescript, USA) and inserted into the pET-z2 plasmid (EMBL plasmid collection), which contains a z2-fusion tag and a His_6 tag at the N-terminus of the target reading frames. For MJ0366 and $HP0242_{x2}$, the plasmids were transformed into the *E. coli* BL21(DE3) strain. Recombinant proteins were expressed by growing the transformed cells in Luria−Bertani (LB) medium in the presence of kanamycin (30−50 μg/mL) for antibiotics selection. Protein overexpression was induced by the addition of 0.5 mM IPTG when the cell density reached OD600 of 0.6 to 0.8, followed by overnight growth at 16 °C. The cells were harvested by centrifugation using a Beckman J20XP centrifuge with a JLA 8.1K rotor for 30 min with 6000 rpm at 4 °C and resuspended in buffer containing 50 mM sodium phosphate (pH 8.0) and 300 mM NaCl. The harvested cells were disrupted using a sonicator, and cell debris and supernatant were separated by a second centrifugation step at 45 000g for 30 min at 4 °C. The supernatant was loaded onto a prepacked 5 mL His-Trap HP column (GE Healthcare Life Science), followed by extensive wash using buffer containing 15 mM imidazole to remove protein impurities and prevent nonspecific binding. Target fusion protein was eluted using 250 mM imidazole with the same buffer background. The eluted fractions were pooled and incubated with TEV protease in a dialysis bag of 3.5 kDa molecular weight cutoff with a 1:100 dialysis ratio at room temperature overnight. This step can efficiently remove the z2 fusion tag from the target protein and to remove the imidazole from the protein solution. A second His-tag based affinity purification step was carried out on the following day to separate the His-tagged z2 tag from the tag-free $HP0242_{x2}$ or MJ0366. Finally, SEC (HiLoad 26/60 Superdex 75, GE Healthcare Life Sciences) was carried out to remove impurities to yield a purity of higher than 95% based on visual inspection of the coomassie-blue-stained sodium dodecyl sulfate polyacrylamide gel (SDS-PAGE). The protein solution was aliquoted, flash-frozen by liquid nitrogen, and stored at −80 °C until further use.

All SAXS samples were purified again the day before the SAXS measurements using SEC to remove aggregated proteins. Monodispersed elution fractions were collected, pooled, and subsequently concentrated using molecular-weight-cutoff centrifugal concentrators (10 000 MWCO Amicon Ultra 0.5, Millipore, USA). The flow through was taken as the blank for the SAXS measurements. To obtain chemically denatured proteins, we used saturating urea or GdnHCl stock solutions in the same buffers to unfold the native samples with appropriate ratio to yield the correct final denaturant concentrations for equilibrium unfolding measurements.

SAXS Data Collection. All SAXS experiments were carried on the beamline BL23A at the National Synchrotron Radiation Research Center (NSRRC, Hsinchu, Taiwan)[24] using a 2D Pilatus detector. The data collections were made for momentum transfer q range from 0.006 to 0.362 Å^{-1} corresponding to an X-ray wavelength of λ = 1.03 Å, 12 keV. The beam geometry was 0.5 × 0.5 mm^2. The sample temperature was maintained at 15 °C using an oil-bath circulation system. SAXS data collections were performed in batch mode where samples were loaded into a cell

DOI: 10.1021/acs.jpcb.5b01984
J. Phys. Chem. B 2015, 119, 5437−5443

with Kapton (polyimide) windows with a path length of 3 mm. The data were recorded by rocking the sample slightly to minimize radiation damage. Six frames with 50 s of exposure time per frame were recorded for each sample and the corresponding buffers for background subtraction. The sampling frames that displayed significant protein aggregation or radiation damage were discarded prior to data analysis. After the SAXS measurements, all samples were recovered and were subject to SDS-PAGE analysis to ensure that no severe radiation damage occurred after SAXS measurements.

SAXS Data Analysis. All SAXS data collected on the BL23A beamline were processed by in-house written program[24] and output for further analysis. The scattering intensity $I(q)$ and the associated error estimate as a function of momentum transfer q were replotted as Guinier plots ($\ln[I(q)]$ versus q^2), Kratky plot ($I(q)\ q^2$ versus q),[24,25] and a recently proposed presentation by Tainer and coworker ($I(q)\ q$ versus q),[35] hereafter denoted as Tainer plot, to access the quality of the data sets. Additionally, the SAXS data were used to extract the R_g values using the AutoR_g module within the software package PRIMUS using default settings.[36] The fractal dimension, D_m, of chemically denatured knotted proteins was calculated using SAXS data corresponding to the q ranges between 0.08 and 0.2 Å^{-1}, as previously reported for IDPs.[37]

Ensemble Optimization Method Analysis. Ensemble optimization method (EOM) analysis[38] of the SAXS data was made using the online server http://www.embl-hamburg.de/biosaxs/eom.html to extract the distributions of R_g using default settings. Briefly, the symmetry option $P1$ was selected so that no symmetry was imposed for calculation. The angular unit was set to 1/Å, the random-coil model was used for the chain type to generate the Cα-angle distribution, and the number of domains was set to zero. The input SAXS data contain scattering intensities as a function of momentum transfer q, ranging from 0.02 to 0.30 Å^{-1}. The resulting R_g distributions were fit to sums of Gaussian distributions using program PRISM (GraphPad Software, USA), and the widths of the Gaussian distribution are defined as the uncertainties.

RESULTS

We first revisited the scaling law of R_g as a function of polypeptide chain length (L) for chemically denatured proteins by updating the reported R_g values in the literature (filled pink squares in Figure 1 and Table S1 in the SI). The updated data set gave essentially the same scaling exponent ($\nu_{denatured} = 0.58 \pm 0.02$, n = 37) as the previously reported value by Plaxco and coworkers ($\nu_{denatured} = 0.59 \pm 0.02$, $n = 26$);[28] the updated prefactor R_0 of 2.05 ± 0.17 Å is also similar to the previously reported value of 1.95 ± 0.16 Å (Table 1). Note that the scaling exponent of chemically denatured proteins is larger than that for IDPs (ν_{IDP} = 0.52 ± 0.01, n = 50)[30] reflecting the preferential long-range interactions within some of the IDPs, leading to slightly larger chain dimensions (open cyan circles in Figure 1).

We next compiled 226 SAXS-based experimental R_g values of folded proteins under native conditions to obtain a scaling law for native proteins (filled green circles in Figure 1, Table 1 and Tables S2–S4 in the SI). Nonlinear regression of this data set indicated one entry of monomeric proteins and two entries of oligomeric proteins as outliers. These are the peptide subunit 104–363 of V-ATPase (259 residues and R_g = 38.8 Å),[39] TcdB, which is a large multidomain protein that adopts an elongated structure (2366 residues and R_g = 78 Å),[40] and the subunit H of the A1AO ATP synthase, which also adopts an elongated

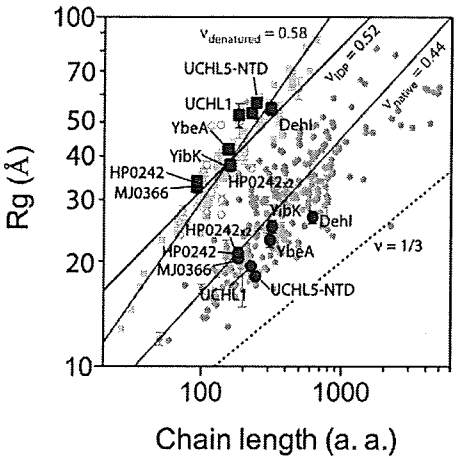


Figure 1. Comparison of R_g values of topologically knotted proteins with nonknotted proteins under native and denaturing conditions. The R_g values of native and chemically denatured knotted proteins as a function of chain length are shown in filled dark green circles and filled red squares, respectively. For comparison, the reported experimental R_g values of native, chemically denatured, and intrinsically disordered proteins are shown in filled light-green circles, filled pink squares, and open cyan circles, respectively. The results of nonlinear regression of individual sets of R_g values are shown in solid lines with the respective scaling exponents indicated. For reference, the theoretical scaling relationship for collapsed chains in poor solvent, that is, the scaling exponent of 1/3, is shown in dashed black line.

structure in solution (208 residues and R_g = 57 Å).[41] Because native proteins do not necessarily fold into compact and spherical conformations, as evidenced by the aforementioned outliers in the most extreme scenario, the distribution of the R_g values as a function of chain length is much more scattered than that of chemically denatured proteins. A scaling exponent of ν_{native} = 0.45 ± 0.02, and the prefactor R_0 of 2.04 ± 0.24 Å was obtained for the complete set of natively folded proteins (Table 1). Limiting the analysis to monomeric protein structures results in the same scaling exponent within error ($\nu_{monomer} = 0.49 \pm 0.08$, n = 129; Table 1). Intriguingly, a separate nonlinear regression for the oligomeric, native proteins yields an even smaller scaling exponent ($\nu_{oligomer} = 0.32 \pm 0.02$, $n = 67$; Table 1). This is much closer to the expectation for collapsed polymer chains in poor solvents. Nevertheless, the much smaller scaling exponent is accompanied by a much larger prefactor $R_{0,oligomer}$ = 4.47 ± 0.47 that is more than double the value of its monomeric counterpart. The reason for such a large difference remains to be established.

Note that the SAXS-based scaling exponent for native, monomeric proteins based on SAXS measurements is very similar to the recent finding based on single-molecule fluorescence measurements, which yielded the scaling exponent of native proteins, ν_{native} = 0.46 ± 0.05 (n = 6).[42] In the same study, the authors also determined the scaling exponent for chemically denatured proteins ($\nu_{denatured}$ = 0.59 to 0.63, n = 6) that is also consistent with our newly updated scaling exponent based on SAXS measurements for chemically denatured proteins albeit the relatively small sampling size in the single-molecule fluorescence measurements.

Having established the scaling relationships of R_g versus L for native and chemically denatured proteins, we set out to use SAXS to obtain the R_g values of eight topologically knotted proteins, including YibK, YbeA, HP0242, and its tandem repeat HP0242$_{x2}$,

DOI: 10.1021/acs.jpcb.5b01984
J. Phys. Chem. B 2015, 119, 5437–5443

Table 1. Nonlinear Regression Results of the Scaling Relationship of Reported R_g Values of Proteins under Different Conditions

state	n	R_0	95% confidence	ν	95% confidence	R^2
chemically denatured	37	2.05 ± 0.17	1.71−2.39	0.58 ± 0.02	0.55−0.62	0.92
native	226	2.08 ± 0.23	1.65−2.56	0.45 ± 0.02	0.41−0.48	0.73
monomeric	129	1.63 ± 0.54	0.45−2.78	0.49 ± 0.08	0.37−0.61	0.57
oligomeric	67	4.42 ± 0.49	3.45−5.39	0.32 ± 0.02	0.29−0.36	0.63
BIOSIS	30					

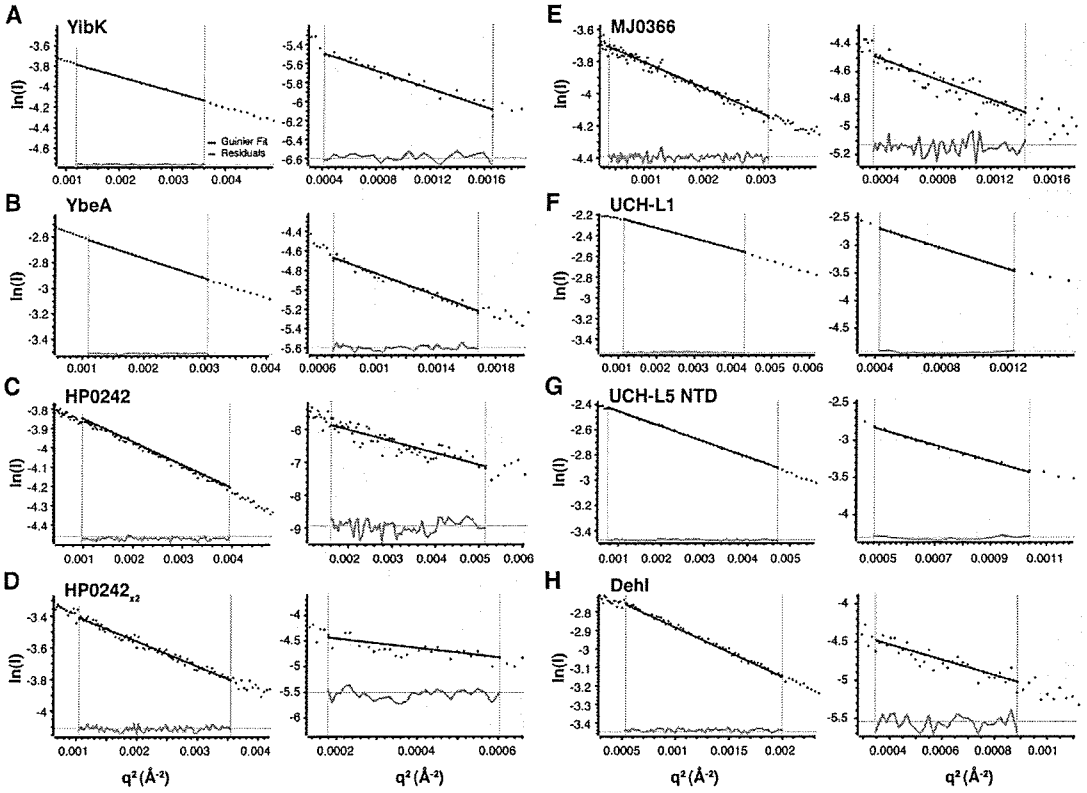


Figure 2. Guinier plot analysis of native and chemically denatured knotted proteins. The Guinier plots of (A) YibK, (B) YbeA, (C) HP0242, (D) HP0242$_{x2}$, (E) MJ0366, (F) UCH-L1, (G) UCH-L5 NTD, and (H) DehI under native and chemically denatured conditions are shown on the left and right panels, respectively. The input data, fitting results, and residuals are shown in black dots, red lines, and green lines, respectively. The AutoR_g module within the PRIMUS package is used for the Guinier plot analyses. The R_g values are tabulated in Table 2.

Table 2. R_g of Topologically Knotted Proteins Determined by SAXS

				R_g (Å)				
protein	knot type	chain length (a.a.)	oligomeric state	native	denatured[a]	r.c.[b]	denatured D_m	denaturing condition
YibK	3_1	160	dimer	20.9 ± 0.2	37.8 ± 1.6	39.1	0.96 ± 0.03	8 M urea
YbeA	3_1	155	dimer	21.7 ± 0.1	42.0 ± 1.2	38.4	0.84 ± 0.03	8 M urea
HP0242		92	dimer	19.0 ± 0.1	32.7 ± 1.3	28.3	1.21 ± 0.10	7 M GdnHCl
HP0242$_{x2}$	3_1	184	monomer	21.8 ± 0.2	52.4 ± 4.1	42.4	0.52 ± 0.10	7 M GdnHCl
MJ0366	3_1	92	dimer	21.6 ± 0.2	33.9 ± 0.9	28.3	1.69 ± 0.05	8 M urea
UCH-L1	5_2	228	monomer	17.3 ± 0.1	53.1 ± 0.6	48.0	1.76 ± 0.02	6.3 M urea
UCH-L5 NTD	5_2	245	monomer	19.0 ± 0.1	56.8 ± 0.4	50.1	1.31 ± 0.04	8 M urea
DehI	6_1	311	dimer	28.4 ± 0.1	54.7 ± 2.4	57.5	0.56 ± 0.08	6 M GdnHCl

[a]When the AutoR_g function of PRIMUS fails to extract denatured R_g values automatically, the data selection is made manually with the aim to fulfill the condition that $R_g < 1.3/q$. The data range that is used for extracting the native R_g is used as the starting point to estimate the corresponding R_g value under chemically denatured states. [b]The random coil (r.c.) R_g values calculated based on the update scaling relationship $R_g = 2.049 \times L^{0.581}$, of which L is the protein chain length (a.a.).

which is the first example of an engineered trefoil-knotted protein, MJ0366, which is the smallest trefoil-knotted protein known to date, human ubiquitin C-terminal hydrolyase, UCH-L1, and -L5 N-terminal domain (UCH-L5 NTD), which are Gordian-knotted protein, and DehI, which contains a Stevedore's knot. These eight topologically knotted proteins

DOI: 10.1021/acs.jpcb.5b01984
J. Phys. Chem. B 2015, 119, 5437−5443

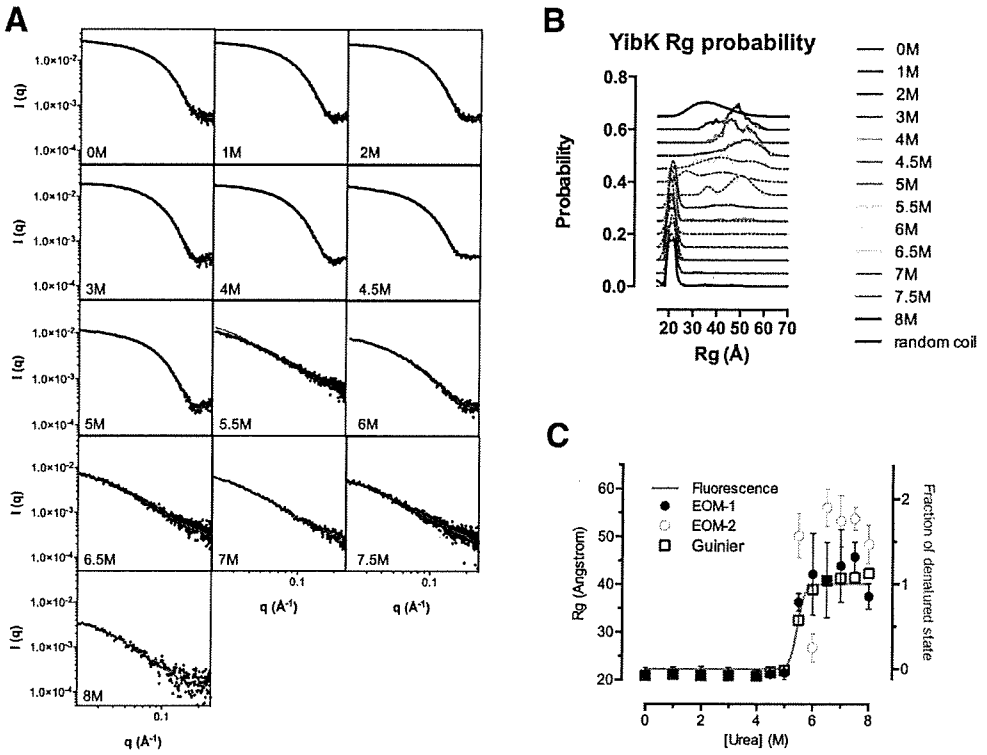


Figure 3. EOM analysis of YibK as a function of urea concentration. (A) Comparison of experimental SAXS profiles (black dots) and the EOM fitting results (red line). The corresponding urea concentrations for individual SAXS profiles are shown at the lower left corner of each panel. (B) EOM-derived R_g distributions. The results are color-ramped from blue to red on increasing urea concentration as indicated. The R_g distributions are fit to Gaussian distributions, and the half widths are defined as the uncertainties. (C) Comparison of the Guinier plot-derived and EOM-derived R_g as a function of urea concentration. The additional R_g distributions, which were attributed to protein aggregation or experimental artifacts, are designated as EOM-2 and are shown in open cyan symbols.

cover most of the protein knot types, except for the figure-eight knot, and sizes, with MJ0366 and DehI being the smallest/simplest and the largest/most complex protein knots, respectively. Characterizations of the chemical denaturation processes of YibK,[14,22] YbeA,[15,23] HP0242,[16] UCH-L1,[17] and MJ0366[21] have been reported in the literature; we have applied similar spectroscopic strategies to characterize the chemical denaturation of UCH-L5 NTD and DehI to ensure that the SAXS data of chemically denatured knotted proteins are collected under fully chemically denatured conditions according to intrinsic fluorescence, far-UV CD, or NMR spectroscopy (Supporting Information Figure S1). Note that the SAXS data of chemically denatured MJ0366 were collected in the presence of 8 M urea. Under this condition, MJ0366 displays no appreciable secondary structure content according to the secondary NMR chemical shifts, although its intrinsic fluorescence suggests the presence of residual structures that may be responsible for the clusters of elevated transverse relaxation of the backbone amide ^{15}N nuclei.[21]

A series of SAXS measurements were carried out on the knotted proteins under native and chemically denatured conditions, and the corresponding R_g values were extracted using standard Guinier plot analysis (Figure 2, Table 2, and Supporting Information Figures S2–S9). The R_g values are plotted as a function of L and compared with the reported R_g values of unknotted proteins (Figure 1). Because many of the knotted proteins are homodimeric, their chemical denaturation processes involve dissociation of dimers into monomers and hence reduce the chain lengths into half of that in their native states, which is taken into consideration in the correlation plot (Figure 1). Indeed, significant increases in R_g values are observed upon chemical denaturation as a result of unfolding. Compared with the experimentally determined R_g values of native unknotted proteins, knotted proteins are generally more compact—their experimental R_g values are comparable to or smaller than the nonlinear regression-derived values according to their chain lengths (Figure 1). In contrast with our initial expectation that the chemically denatured knotted proteins may exhibit residual structures that can result in significant conformational compactions, our results indicate that the chemically denatured knotted proteins are in fact random coil-like according to their R_g values. Some of the denatured states of the knotted proteins are actually larger than the expected values for random-coil proteins. For urea-denatured YibK and YbeA, in particular, which have been shown to maintain their knotted topology in the presence of 8 M urea,[20] their R_g values scale with the same chain length dependency as those of the other chemically denatured proteins which are unknotted; that is, they behave like random-coiled chains without significant degree of conformational compaction while being topologically knotted.

To further analyze the degree of disorder of the chemically denatured knotted proteins, we calculated their fractal dimension, D_m, which was previously used to characterize the bacteriophage λ N protein (Table 2).[37] For well-folded globular proteins, the expected D_m is 4, while the predicated value for well-solvated polymers is 1.7. For an IDP like the bacteriophage λ N

DOI: 10.1021/acs.jpcb.5b01984
J. Phys. Chem. B 2015, 119, 5437–5443

protein, the corresponding D_m value is close to 1.7. This is also the case for chemically denatured MJ0366 (D_m = 1.69 ± 0.05) and UCH-L1 (D_m = 1.76 ± 0.02); however, the fractal dimensions of UCH-L5 NTD and HP0242 are slightly smaller than 1.7, and those of YibK, YbeA, $HP0242_{x2}$, and DehI are significantly smaller than the expected values for well-solved polymers. Part of the reason may be attributed to the low contrasts of the SAXS signals in the mid-q range ($0.08 < q < 0.2$) that was used for the calculations.

To evaluate the contribution of conformational heterogeneity in chemically denatured proteins, an ensemble optimization method (EOM)[38] was employed. We monitored the equilibrium unfolding of YibK over a range of urea concentrations. The resulting R_g distributions of YibK as a function of urea concentrations are in good agreement with those derived from the Guinier plot, although EOM analysis also identifies an additional population whose R_g values are significantly larger than expected R_g values for random coils, which we attribute to potentially aggregated populations (Figure 3). Importantly, both Guinier and EOM analyses yield an apparent two-state-like unfolding transition that is consistent with previously reported equilibrium unfolding data based on intrinsic fluorescence and far-UV CD measurements.[14,19] Similar urea-titration and EOM analysis could not be carried out for YbeA due to its tendency to aggregate under intermediate urea concentrations at the protein concentrations required for SAXS measurements (data not shown).

DISCUSSION

Recent years have witnessed a growing body of experimental studies on topologically knotted proteins.[12−20] One of the most remarkable findings is the fact that YibK and YbeA can retain their topologically knotted conformations under chemically denaturing conditions without appreciable amount of secondary or tertiary structures.[20] Our current SAXS analysis suggests that 8 M urea-denatured YibK and YbeA are, in fact, random coil-like (Figure 1). Such an observation is at variance with the computational studies of the folding pathways of a number of topologically knotted proteins. One of the earliest simulation studies on knotted proteins reported that the introduction of non-native interactions was necessary for successful folding and knotting of YibK using a simplified coarse-grained simulation model.[43] A slip-knotting mechanism was subsequently proposed for both YibK and YbeA in which a compact and slip-knotted folding intermediate is needed to overcome the topological bottleneck.[8] While it is unclear whether chemically denatured tandem repeat of HP0242, that is, $HP0242_{x2}$, remains topologically knotted, two recent simulation studies suggested that knot formation of $HP0242_{x2}$ takes place after a significant number of native contacts are formed.[7,10] Similar findings were reported for the smallest knotted protein, MJ0366.[9,44] While further experimental data are needed to establish when and how the knots are formed along the folding pathways of these topologically knotted proteins and whether knot formation indeed requires relatively compact chain dimensions, our SAXS data suggest that, at least in the case of YibK and YbeA, being topologically knotted does not necessarily require significant compaction in their global polypeptide chain dimensions. This is consistent with a previous computational study, which showed that the presence of random-walk polymers exhibits similar R_g values regardless of being unknotted or trefoil-knotted, as long as the knotted element is localized.[45] Further experimental data on chemically denatured knotted proteins with different knot types are required to test the theoretical predictions, which suggested that polymers with more complex knots are more spherical than those with less complex knots.[46]

In summary, we have revisited the scaling relationships of R_g in relationship with chain lengths for proteins under native and chemically denatured conditions based on reported experimental SAXS R_g values. Our own SAXS data on a collection of topologically knotted proteins show that YibK and YbeA are not significantly more compact than random coils in terms of their R_g values despite being loosely topologically knotted in the presence of 8 M urea.[12] Likewise, the R_g values of other chemically denatured knotted proteins also fall within the expected values for random coils, suggesting that flexible polypeptide chains can adopt specific folding chiralities and topologies, that is, being topologically knotted, while their chain dimensions remain random -coil-like.

ASSOCIATED CONTENT

Supporting Information

Compilation of reported R_g values of native and chemically denatured proteins, SAXS sample preparation, data collection, and analysis. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: sthsu@gate.sinica.edu.tw.

Notes

The authors declare no competing financial interests.

ACKNOWLEDGMENTS

S.-T.D.H. is a recipient of the Career Development Award (CDA-00025/2010-C) from the International Human Frontier Science Program and is supported by funding from the National Science Council (100-2113-M-001-031-MY2 and 101-2627-M-001-004), National Synchrotron Radiation Research Center, and Academia Sinica, Taiwan. We thank the staff of BL23A of National Synchrotron Radiation Research Center for their assistance in SAXS data collection.

REFERENCES

(1) Taylor, W. R. A Deeply Knotted Protein Structure and How It Might Fold. *Nature* **2000**, *406*, 916−919.

(2) Lai, Y. L.; Chen, C. C.; Hwang, J. K. pKNOT v.2: The Protein KNOT Web Server. *Nucleic Acids Res.* **2012**, *40*, W228−231.

(3) Virnau, P.; Mirny, L. A.; Kardar, M. Intricate Knots in Proteins: Function and Evolution. *PLoS Comput. Biol.* **2006**, *2*, e122.

(4) Millett, K. C.; Rawdon, E. J.; Stasiak, A.; Sulkowska, J. I. Identifying Knots in Proteins. *Biochem. Soc. Trans.* **2013**, *41*, 533−537.

(5) Virnau, P.; Mallam, A.; Jackson, S. Structures and Folding Pathways of Topologically Knotted Proteins. *J. Phys.: Condens. Matter* **2011**, *23*, 033101.

(6) Noel, J. K.; Sulkowska, J. I.; Onuchic, J. N. Slipknotting Upon Native-Like Loop Formation in a Trefoil Knot Protein. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 15403−15408.

(7) Sulkowska, J. I.; Noel, J. K.; Onuchic, J. N. Energy Landscape of Knotted Protein Folding. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 17783−17788.

(8) Sulkowska, J. I.; Sulkowski, P.; Onuchic, J. Dodging the Crisis of Folding Proteins with Knots. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 3119−3124.

(9) Noel, J. K.; Onuchic, J. N.; Sulkowska, J. I. Knotting a Protein in Explicit Solvent. *J. Phys. Chem. Lett.* **2013**, *4*, 3570−3573.

DOI: 10.1021/acs.jpcb.5b01984
J. Phys. Chem. B 2015, 119, 5437−5443

(10) Li, W.; Terakawa, T.; Wang, W.; Takada, S. Energy Landscape and Multiroute Folding of Topologically Complex Proteins Adenylate Kinase and 2ouf-Knot. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 17789–17794.

(11) Sulkowska, J. I.; Noel, J. K.; Ramirez-Sarmiento, C. A.; Rawdon, E. J.; Millett, K. C.; Onuchic, J. N. Knotting Pathways in Proteins. *Biochem. Soc. Trans.* **2013**, *41*, S23–S27.

(12) Mallam, A. L.; Morris, E. R.; Jackson, S. E. Exploring Knotting Mechanisms in Protein Folding. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 18740–18745.

(13) Andersson, F. I.; Pina, D. G.; Mallam, A. L.; Blaser, G.; Jackson, S. E. Untangling the Folding Mechanism of the 5_2-Knotted Protein UCH-L3. *FEBS J.* **2009**, *276*, 2625–2635.

(14) Mallam, A. L.; Jackson, S. E. Folding Studies on a Knotted Protein. *J. Mol. Biol.* **2005**, *346*, 1409–1421.

(15) Mallam, A. L.; Jackson, S. E. A Comparison of the Folding of Two Knotted Proteins: Ybea and Yibk. *J. Mol. Biol.* **2007**, *366*, 650–665.

(16) King, N. P.; Jacobitz, A. W.; Sawaya, M. R.; Goldschmidt, L.; Yeates, T. O. Structure and Folding of a Designed Knotted Protein. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 20732–20737.

(17) Andersson, F. I.; Werrell, E. F.; McMorran, L.; Crone, W. J.; Das, C.; Hsu, S. T. D.; Jackson, S. E. The Effect of Parkinson's-Disease-Associated Mutations on the Deubiquitinating Enzyme UCH-L1. *J. Mol. Biol.* **2011**, *407*, 261–272.

(18) Mallam, A. L.; Jackson, S. E. Knot Formation in Newly Translated Proteins Is Spontaneous and Accelerated by Chaperonins. *Nat. Chem. Biol.* **2012**, *8*, 147–153.

(19) Mallam, A. L.; Jackson, S. E. Probing Nature's Knots: The Folding Pathway of a Knotted Homodimeric Protein. *J. Mol. Biol.* **2006**, *359*, 1420–1436.

(20) Mallam, A. L.; Rogers, J. M.; Jackson, S. E. Experimental Detection of Knotted Conformations in Denatured Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 8189–8194.

(21) Wang, I.; Chen, S. Y.; Hsu, S.-T. D. Unraveling the Folding Mechanism of the Smallest Knotted Protein, MJ0366. *J. Phys. Chem. B* **2015**, *119*, 4359–4370.

(22) Hsieh, S. J.; Mallam, A. L.; Jackson, S. E.; Hsu, S. T. D. Backbone NMR Assignments of a Topologically Knotted Protein in Urea-Denatured State. *Biomol. NMR Assignments* **2014**, *8*, 439–442.

(23) Hsieh, S. J.; Mallam, A. L.; Jackson, S. E.; Hsu, S. T. Backbone NMR Assignments of a Topologically Knotted Protein in Urea-Denatured State. *Biomol NMR Assignments* **2014**, *8*, 283–285.

(24) Jeng, U. S.; Su, C. H.; Su, C. J.; Liao, K. F.; Chuang, W. T.; Lai, Y. H.; Chang, J. W.; Chen, Y. J.; Huang, Y. S.; Lee, M. T.; et al. A Small/Wide-Angle X-Ray Scattering Instrument for Structural Characterization of Air–Liquid Interfaces, Thin Films and Bulk Specimens. *J. Appl. Crystallogr.* **2010**, *43*, 100–121.

(25) Putnam, C. D.; Hammel, M.; Hura, G. L.; Tainer, J. A. X-Ray Solution Scattering (SAXS) Combined with Crystallography and Computation: Defining Accurate Macromolecular Structures, Conformations and Assemblies in Solution. *Q. Rev. Biophys.* **2007**, *40*, 191–285.

(26) Hura, G. L.; Menon, A. L.; Hammel, M.; Rambo, R. P.; Poole, F. L., 2nd; Tsutakawa, S. E.; Jenney, F. E., Jr.; Classen, S.; Frankel, K. A.; Hopkins, R. C.; et al. Robust, High-Throughput Solution Structural Analyses by Small Angle X-Ray Scattering (SAXS). *Nat. Methods* **2009**, *6*, 606–612.

(27) Blanchet, C. E.; Svergun, D. I. Small-Angle X-Ray Scattering on Biological Macromolecules and Nanocomposites in Solution. *Annu. Rev. Phys. Chem.* **2013**, *64*, 37–54.

(28) Kohn, J. E.; Millett, I. S.; Jacob, J.; Zagrovic, B.; Dillon, T. M.; Cingel, N.; Dothager, R. S.; Seifert, S.; Thiyagarajan, P.; Sosnick, T. R.; et al. Random-Coil Behavior and the Dimensions of Chemically Unfolded Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 12491–12496.

(29) Flory, P. J. *Principles of Polymer Chemistry*; Cornell Univ. Press: Ithaca, NY, 1953.

(30) Bernado, P.; Svergun, D. I. Structural Analysis of Intrinsically Disordered Proteins by Small-Angle X-Ray Scattering. *Mol. BioSyst.* **2012**, *8*, 151–167.

(31) Mallam, A. L.; Jackson, S. E. Folding Studies on a Knotted Protein. *J. Mol. Biol.* **2005**, *346*, 1409–1421.

(32) Tsai, J. Y.; Chen, B. T.; Cheng, H. C.; Chen, H. Y.; Hsaio, N. W.; Lyu, P. C.; Sun, Y. J. Crystal Structure of Hp0242, a Hypothetical Protein from *Helicobacter pylori* with a Novel Fold. *Proteins* **2006**, *62*, 1138–1143.

(33) Morrow, M. E.; Kim, M. I.; Ronau, J. A.; Sheedlo, M. J.; White, R. R.; Chaney, J.; Paul, L. N.; Lill, M. A.; Artavanis-Tsakonas, K.; Das, C. Stabilization of an Unusual Salt Bridge in Ubiquitin by the Extra C-Terminal Domain of the Proteasome-Associated Deubiquitinase UCH37 as a Mechanism of Its Exo Specificity. *Biochemistry* **2013**, *52*, 3564–3578.

(34) Schmidberger, J. W.; Wilce, J. A.; Weightman, A. J.; Whisstock, J. C.; Wilce, M. C. The Crystal Structure of DehI Reveals a New α-Haloacid Dehalogenase Fold and Active-Site Mechanism. *J. Mol. Biol.* **2008**, *378*, 284–294.

(35) Rambo, R. P.; Tainer, J. A. Accurate Assessment of Mass, Models and Resolution by Small-Angle Scattering. *Nature* **2013**, *496*, 477–481.

(36) Konarev, P. V.; Volkov, V. V.; Sokolova, A. V.; Koch, M. H. J.; Svergun, D. I. Primus: A Windows PC-Based System for Small-Angle Scattering Data Analysis. *J. Appl. Cryst.* **2003**, *36*, 1277–1282.

(37) Johansen, D.; Trewhella, J.; Goldenberg, D. P. Fractal Dimension of an Intrinsically Disordered Protein: Small-Angle X-Ray Scattering and Computational Study of the Bacteriophage λ N Protein. *Protein Sci.* **2011**, *20*, 1955–1970.

(38) Bernado, P.; Mylonas, E.; Petoukhov, M. V.; Blackledge, M.; Svergun, D. I. Structural Characterization of Flexible Proteins Using Small-Angle X-Ray Scattering. *J. Am. Chem. Soc.* **2007**, *129*, 5656–5664.

(39) Dip, P. V.; Saw, W. G.; Roessle, M.; Marshansky, V.; Gruber, G. Solution Structure of Subunit A, a104–363, of the *Saccharomyces cerevisiae* V-ATPase and the Importance of Its C-Terminus in Structure Formation. *J. Bioenerg. Biomembr.* **2012**, *44*, 341–350.

(40) Albesa-Jove, D.; Bertrand, T.; Carpenter, E. P.; Swain, G. V.; Lim, J.; Zhang, J.; Haire, L. F.; Vasisht, N.; Braun, V.; Lange, A.; von Eichel-Streiber, C.; Svergun, D. I.; Fairweather, N. F.; Brown, K. A. Four Distinct Structural Domains in *Clostridium difficile* Toxin B Visualized Using SAXS. *J. Mol. Biol.* **2010**, *396*, 1260–1270.

(41) Biukovic, G.; Rossle, M.; Gayen, S.; Mu, Y.; Gruber, G. Small-Angle X-Ray Scattering Reveals the Solution Structure of the Peripheral Stalk Subunit H of the A1AO ATP Synthase from *Methanocaldococcus jannaschii* and Its Binding to the Catalytic a Subunit. *Biochemistry* **2007**, *46*, 2070–2078.

(42) Hofmann, H.; Soranno, A.; Borgia, A.; Gast, K.; Nettels, D.; Schuler, B. Polymer Scaling Laws of Unfolded and Intrinsically Disordered Proteins Quantified with Single-Molecule Spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 16155–16160.

(43) Wallin, S.; Zeldovich, K. B.; Shakhnovich, E. I. The Folding Mechanics of a Knotted Protein. *J. Mol. Biol.* **2007**, *368*, 884–893.

(44) a Beccara, S.; Skrbic, T.; Covino, R.; Micheletti, C.; Faccioli, P. Folding Pathways of a Knotted Protein with a Realistic Atomistic Force Field. *PLoS Comput. Biol.* **2013**, *9*, e1003002.

(45) Millett, K.; Dobay, A.; Stasiak, A. Linear Random Knots and Their Scaling Behavior. *Macromolecules* **2005**, *38*, 601–606.

(46) Rawdon, E. J.; Kern, J. C.; Piatek, M.; Plunkett, P.; Stasiak, A. J.; Millett, K. C. Effect of Knotting on the Shape of Polymers. *Macromolecules* **2008**, *41*, 8281–8287.

DOI: 10.1021/acs.jpcb.5b01984
J. Phys. Chem. B 2015, 119, 5437–5443