

Small Angle Neutron Scattering Studies of R67 Dihydrofolate Reductase, a Tetrameric Protein with Intrinsically Disordered N-Termini

Purva P. Bhojane,[†] Michael R. Duff, Jr.,[†] Khushboo Bafna,[‡] Pratul Agarwal,[§] Christopher Stanley,^{||} and Elizabeth E. Howell^{*,†,‡,§}

[†]Department of Biochemistry and Cellular and Molecular Biology, University of Tennessee, Knoxville, Tennessee 37996-0840, United States

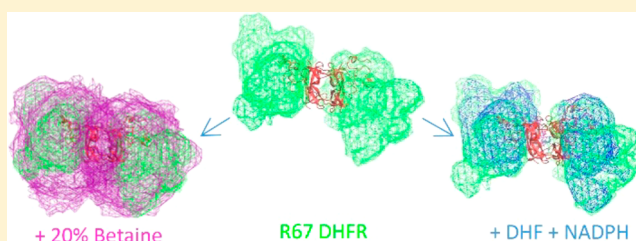
[‡]Genome Science and Technology Program, University of Tennessee, Knoxville, Tennessee 37996-0840, United States

[§]Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, United States

^{||}Biology and Soft Matter Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, United States

Supporting Information

ABSTRACT: R67 dihydrofolate reductase (DHFR) is a homotetramer with a single active site pore and no sequence or structural homology with chromosomal DHFRs. The R67 enzyme provides resistance to trimethoprim, an active site-directed inhibitor of *Escherichia coli* DHFR. Sixteen to twenty N-terminal amino acids are intrinsically disordered in the R67 dimer crystal structure. Chymotrypsin cleavage of 16 N-terminal residues results in an active enzyme with a decreased stability. The space sampled by the disordered N-termini of R67 DHFR was investigated using small angle neutron scattering. From a combined analysis using molecular dynamics and the program SASSIE (http://www.smallangles.net/sassie/SASSIE_HOME.html), the apoenzyme displays a radius of gyration (R_g) of 21.46 ± 0.50 Å. Addition of glycine betaine, an osmolyte, does not result in folding of the termini as the R_g increases slightly to 22.78 ± 0.87 Å. SASSIE fits of the latter SANS data indicate that the disordered N-termini sample larger regions of space and remain disordered, suggesting they might function as entropic bristles. Pressure perturbation calorimetry also indicated that the volume of R67 DHFR increases upon addition of 10% betaine and decreased at 20% betaine because of the dehydration of the protein. Studies of the hydration of full-length R67 DHFR in the presence of the osmolytes betaine and dimethyl sulfoxide find around 1250 water molecules hydrating the protein. Similar studies with truncated R67 DHFR yield around 400 water molecules hydrating the protein in the presence of betaine. The difference of ~ 900 waters indicates the N-termini are well-hydrated.



Dihydrofolate reductase (DHFR) catalyzes the reduction of dihydrofolate (DHF) to tetrahydrofolate (THF) using NADPH as a cofactor. THF and its derivatives serve as cellular cofactors for one-carbon transfer reactions involved in the synthesis of nucleotides such as purines and thymidine, amino acids such as methionine and glycine, and various other metabolites. Trimethoprim is a potent inhibitor of *Escherichia coli* chromosomal DHFR (EcDHFR) and has been widely used as an antibacterial drug. The gene encoding R67 DHFR, carried by an R-plasmid, confers resistance against trimethoprim. Recent clinical isolates of *E. coli* causing urinary tract infections have the gene encoded in class I integrons flanked by other drug resistance genes.¹ There are no antibiotics that target R67 DHFR, though promising leads have recently been discovered.^{2,3} This type II DHFR (78 amino acids long) is genetically and structurally unrelated to EcDHFR. R67 DHFR is a homotetramer, and each monomer has five antiparallel β -strands that assemble into a dimer with a six-stranded β -barrel at the subunit interface. Using loop–loop interactions, two

dimers assemble into a tetrameric “doughnut” with a single active site pore.⁴

Numerous experiments indicate the 16–20 N-terminal residues of R67 DHFR are disordered and can tolerate various sequences. For example, several disorder predictors indicate the N-terminal sequence is intrinsically disordered.⁵ Also, the first 17 amino acids for each monomer do not appear in the dimer crystal structure.⁶ The N-termini can be cleaved after F16 by chymotrypsin treatment, and the truncated protein is almost fully active, although 2.6 kcal/mol less stable.⁷ The truncated tetrameric protein was crystallized, and the structure was first determined at a resolution of 1.7 Å⁴ and later at 0.96–1.26 Å.^{8,9} High thermal factors in the latter structure suggest the stretch of residues 17–21 is also disordered. In addition, electron

Received: August 23, 2017

Revised: October 2, 2017

Published: October 11, 2017

densities for residues 21–23 were diffuse, indicating high mobility.⁹

Other type II DHFR variants (e.g., R388 and R751) show different N-terminal sequences, but the same core sequence contributes to the β -barrel structure.^{5,10,11} This can also be seen from a sequence alignment of the type II DHFR variants¹² showing non-identity in the first 21 residues. His tags can also be added to the N-termini.^{13–15} In addition, a tandem array of four R67 DHFR gene copies encodes a protein in which the C- and N-termini of the first and second monomers are fused as well as the second and third monomers and the third and fourth monomers. The resulting Quad1 protein possessing 4 times the molecular mass of the R67 DHFR monomer is stable as well as functional.¹⁶ Asymmetric mutations in the core of R67 DHFR that favor one topology^{17,18} are used in these Quad constructs. Similarly, the N-terminal sequences from R388 and R751 can be used as the linker domains to give a functional monomeric Quad4 protein.⁵ These various experiments and constructs indicate the N-termini can be fused without a loss of function. Pelletier and co-workers have made similar dimeric fused constructs of R67 DHFR.^{19,20}

To gain information about the conformational space occupied by the disordered N-terminal sequences in R67 DHFR, we used small angle neutron scattering (SANS) experiments. Because of the inherent contrast between hydrogen and deuterium atoms, SANS data of the hydrogenated protein in D₂O buffer allow modeling of the ensemble of conformations sampled by the disordered tails.

As disordered sequences often undergo coupled binding and folding, we monitored any potential changes in the conformational sampling of the disordered tails upon formation of a binary complex (R67 DHFR–NADP⁺) or a ternary complex (R67 DHFR–NADP⁺–DHF). Also, osmolytes have been shown to exert protein-stabilizing forces via a preferential exclusion mechanism.^{21,22} To determine whether addition of an osmolyte leads to folding of the termini, we added deuterated betaine to examine any changes in the R67 DHFR shape using SANS.

Also of interest, the water associated with the protein surface comprises the hydration layer, which can be differentiated from the bulk solvent. The first hydration shell can contain tightly bound water as well as water that can freely exchange. These differences are due to the varied environments associated with the protein surface, which can display different clefts and bumps as well as different atom types.²³ Computational studies have shown that water molecules hydrating the disordered chains exhibit properties different from the properties of those surrounding globular domains, in terms of both the number of waters and the structural order of the water molecules in the hydration layer.^{24,25}

To monitor the preferential hydration of full-length and truncated R67 DHFR, we used hydrogenated osmolytes in a D₂O buffer solution in additional SANS experiments. This is analogous to a H₂O/D₂O contrast variation approach; i.e., the contrast created by addition of a hydrogenated osmolyte allows measurement of the hydration shell associated with R67 DHFR. The contrast created by osmolytes differentiates between the hydration layer and the bulk solvent. The information obtained from the scattering contrast can be used to obtain the number of water molecules in the hydration layer that are responsible for exclusion of the added osmolyte from the protein surface.

METHODS

Protein Expression and Purification. Full-length R67 DHFR (MIRSSNEVSN PVAGNFVFPV NATFGMGDRV RKKSGAAWQG QVGVWYCTNL TPEGYAVESE AHPGSV-QIYP VAALERIN) was expressed and purified as described by Reece et al.⁷ Briefly, cell lysates were subjected to ammonium sulfate precipitation and ion-exchange column chromatography to purify the protein to homogeneity. Purified samples were dialyzed against distilled, deionized H₂O and lyophilized.

Chymotrypsin-truncated R67 DHFR was obtained as described by Reece et al.,⁷ starting from full-length His-tagged R67 DHFR. The His-tagged construct has the synthetic R67 DHFR gene⁷ cloned into the pRSETB vector from Invitrogen.²⁶ Purification was performed with a nickel-nitrilotriacetic acid (Ni-NTA) column (Qiagen), followed by elution from a DEAE Fractogel column. The resulting protein was incubated with immobilized chymotrypsin (Sigma-Aldrich) in 10 mM Tris/1 mM EDTA, pH 8.0 buffer overnight at 4 °C and later at room temperature for \leq 24 h. Chymotrypsin cleaves after F16 in the R67 DHFR sequence (or after F47 in the His-tagged sequence). The progress of the reaction was monitored by sodium dodecyl sulfate electrophoresis (see Figure S1). Immobilized chymotrypsin was removed, and the truncated tetramer was separated from peptide fragments by gel filtration at pH 8 using G75 Sephadex. A Ni-NTA column further separated the cleaved N-terminus from the tetrameric core of the protein. The purified truncated R67 DHFR was dialyzed against water using a 7 kDa cutoff membrane and then lyophilized. Protein concentrations were determined by measuring the absorbance at 280 nm of the solution using an extinction coefficient determined with a bicinchoninic acid (Pierce) assay.

Small Angle Neutron Scattering (SANS). The sample of lyophilized, full-length, apo R67 DHFR was reconstituted in 20 mM deuterated Tris buffer in D₂O (pD 7.5). Experiments were also performed to study any changes in the ordering of the N-termini upon binding of NADP⁺ to apo R67 DHFR (binary complex formation) and upon binding of dihydrofolate (DHF) to the R67 DHFR–NADP⁺ complex (ternary complex formation) under saturating ligand concentrations (3 mM NADP⁺ for binary samples and 3 mM NADP⁺ and 2 mM DHF for ternary samples). To study the effect of betaine on the disordered N-termini of R67 DHFR, the change in overall shape and compaction of apoprotein in the presence of 20% deuterated betaine was explored. Additionally, samples of full-length apo R67 DHFR with no osmolyte and with the osmolytes betaine and dimethyl sulfoxide (DMSO) were prepared to investigate protein hydration. The osmolytes were hydrogenated to create a contrast with the deuterated buffer conditions, allowing measurement of changes in preferential hydration of apo R67 DHFR.²⁷ The concentrations of osmolytes ranged from 2.5 to 20% (w/v) for betaine and from 2.5 to 17.5% (v/v) for DMSO. The protein concentrations ranged from 4.5 to 7.5 mg/mL. Similar sample sets using the truncated R67 DHFR protein with 0 to 20% (w/v) betaine were prepared. The concentration of the truncated protein was 2.4–2.6 mg/mL. Buffer controls were run to detect the background scattering. All samples were prepared, centrifuged, and loaded into banjo-shaped quartz cuvettes (Hellma USA, Plainville, NY) with a path length of 2 mm.

Experiments were performed on the EQ-SANS instrument at the Spallation Neutron Source at the Oak Ridge National

Laboratory. In 60 Hz operation mode, a 4 m sample–detector distance with a 2.5–6.1 Å wavelength band was used. Neutron exposure times were approximately 1 h, and the scattered neutrons were detected on a 1 m × 1 m two-dimensional detector at 25 °C.

The data collected for all experiments were reduced using MANTID Plot,²⁸ and the total two-dimensional scattering was corrected by the scattering from the empty quartz cell. Then, the scattering was normalized by the incident beam flux and radially averaged to obtain the absolute scale intensity, $I(q)$, versus scattering angle, q . The background scattering for the respective buffers was subtracted from the total scattering. Guinier analysis with a linear plot of $\ln I(q)$ versus q^2 for low- q data gave a slope of $-(R_g^2)/3$, where R_g is the radius of gyration and the intercept on the Y -axis gave the $I(0)$ value. Estimates of R_g and zero-angle scattering intensity $I(0)$ were obtained using eq 1:²⁷

$$I(q) = I(0)e^{-q^2 R_g^2 / 3} \quad (1)$$

where $I(q)$ is the scattering intensity at small angles (q).

The data were also analyzed using GNOM in the ATSAS package.²⁹ GNOM reads the scattering profile and evaluates the particle distance distribution function, $P(R)$, in a defined range of distances and yields the apparent radius of gyration (R_g) and zero-angle scattering intensity $I(0)$. Data for each sample were fit using Guinier analysis and GNOM (see Table S1).

The R_g values and zero-angle scattering intensities, $I(0)$, of R67 DHFR in the presence of varying concentrations of osmolytes (betaine and DMSO) were normalized by the protein concentration of each sample. To obtain information about the preferential hydration of R67 DHFR and the effect of osmolytes on hydration, the change in $I(0)$ with an increasing concentration of osmolytes obtained from the GNOM analysis was fit to eq 2 from ref 27

$$I_s(0)/I(0) = \left[1 + f_v \left(\frac{\rho_w - \rho_s}{\rho_p - \rho_w} \right) \left(\frac{V_p + V_w}{V_p} \right) \right]^2 \quad (2)$$

where $I_s(0)$ and $I(0)$ are the zero-angle scattering intensities in the presence and absence of an osmolyte, respectively, f_v or fractional volume, is the concentration of osmolyte added (w/v for betaine and v/v for DMSO), ρ_w , ρ_s , and ρ_p are the scattering-length densities of water, solute (=osmolyte), and protein, respectively, and V_p and V_w are the volumes of protein and protein-associated water, respectively. The scattering-length densities of the full-length ($3.248 \times 10^{10} \text{ cm}^{-2}$) and truncated ($3.228 \times 10^{10} \text{ cm}^{-2}$) proteins, betaine ($0.817 \times 10^{10} \text{ cm}^{-2}$), and DMSO ($-0.051 \times 10^{10} \text{ cm}^{-2}$) and the protein volumes were calculated using the online tool MULCh.³⁰ The volume of protein-associated water gives the number of water molecules in the hydration layer of R67 DHFR upon addition of the osmolyte. Note that eq 2 assumes V_p and V_w remain constant across the range of osmolyte concentrations used. In other words, the measured V_w is an average across the betaine range used for the preferential hydration SANS experiments.

Analysis Using MD and SASSIE. Our next step was to analyze the SANS data using models generated via MD and SASSIE (see http://www.smallangles.net/sassie/SASSIE_HOME.html).³¹ The latter creates atomistic models of the protein using Monte Carlo simulations, calculates theoretical scattering data for these models using SasCalc or Xtal2SAS tools, and compares the theoretical data to the experimental

data. The experimental SANS data were interpolated into SASSIE in a defined q range using the data interpolation module.

As both MD and SASSIE require a model of the full-length protein sequence to generate structures for fitting, sixteen residues (MIRSSNEVSNPVAGNF-) were added to each of the four N-termini of the truncated R67 DHFR structure (PDB entry 2RH2)⁸ using Modeller (version 9.15).³² A total of 100 models were generated, and 10 structures with the lowest discrete optimized potential energy score were used further. The selected models were minimized under vacuum for 10000 steps. The minimized structures were then solvated in a SPC/E water box³³ and equilibrated using the protocol as described by Ramanathan et al.³⁴ The protein and ligand parameters were generated using AMBER force field *ff14SB*. Extensive MD simulations were performed for the apoenzyme, the binary complex with NADPH, and the ternary complex with NADPH and DHF using the AMBER 14 simulation package.³⁵ Initially, 10 models of the apoenzyme were simulated for 100 ns each, and selected conformations from these runs that gave good fits to the SANS data were further studied for four apo simulations of 1 μs each. Therefore, the total aggregate sampling for apo systems was 5 μs . Similarly, nine MD simulations for the binary complex and five for the ternary complex were performed for 1 μs each, providing aggregate 9 μs and 5 μs MD sampling for binary and ternary complexes, respectively.

Additionally, two more models were built. The first model had two pairs of N-termini interacting with each other on both sides of the pore, and a second model moved all the four termini to block access to the active site pore. This approach allowed the construction of numerous structures in which the N-termini sample a large area of conformational space.

Frames from the MD trajectories were analyzed in SASSIE using the SasCalc module that generated theoretical SANS profiles, which were then compared to the experimental SANS data using the χ^2 analysis module. Those structures with a low χ^2 value (<10) were chosen as good fits. Complex Monte Carlo simulations also generated additional conformers for fitting. In this process, the core of the protein remained constant, and only alternate conformations of the 21 N-terminal amino acids were generated. Acceptable frames avoided atom overlaps. In addition, on the basis of the average R_g obtained (for example, $\sim 21.5 \text{ \AA}$ for apo R67 DHFR), directed Monte Carlo sampling additionally generated >100000 structures with R_g values limited to a range from 20.5 to 22.5 Å. These structures were subjected to a 500-step minimization using NAMD. Again, the theoretical SANS profiles were calculated using the SasCalc module in SASSIE, followed by a χ^2 analysis. We chose to use a strategy by which we analyzed single structures, as opposed to an ensemble structure method, because our relatively exhaustive analysis in SASSIE using our MD and Monte Carlo structures was found to be adequate to fit our SANS data.

Both MD and SASSIE analyses were performed to fit the experimental SANS profiles for the ligand-bound complexes (binary and ternary) as well as for apo R67 DHFR in 20% deuterated betaine. To generate sufficient conformers with low χ^2 values, the apo, binary, and ternary structures were interconverted by a Python script by adding or removing the coordinates of the NADP⁺ and DHF ligands in the active site pore of the MD and Monte Carlo-generated structures. The ligand coordinates were obtained from PDB entries 2RK1 and 2RK2.⁸ For the analysis of the binary data, initially one NADP⁺ was positioned in the active site pore as per the 2RK2 crystal

structure;⁸ however, only 92 good fits were obtained. To gain more fits, we considered the possibility that a second cofactor could bind as the concentration of NADP⁺ in the SANS sample was high (3 mM) and the active site pore can accommodate two homoligands.³⁶ There are four sets of coordinates for NADP⁺ in the 2RK2 crystal structure because of the symmetry of the active site. Therefore, we positioned a second set of NADP⁺ coordinates in a symmetry-related position using the 2RK2 structure and continued with the analysis. A workflow and summary of the various steps in our analyses for apo, binary, and ternary complexes are provided in Figure S2.

Snapshots from MD simulations were also used to compute the number of water molecules present in the first solvation shell of the protein. AMBER's *ptraj* module and command *watershell*, with the default cutoff of 3.4 Å, were used.

Data Mining. The structures that best fit the SANS profiles were data mined to find the most frequent interactions between the N-terminus and all the other residues in the protein. To analyze the structures, the *cpptraj* program in Amber14³⁵ was used to calculate the distances between the center of mass (COM) of each residue and all the other individual residues in the R67 DHFR tetramer. A python script was then used to determine the minimum distance between the COM for each pair of residues for all structures that fit the SANS data. Additionally, the number of times the center of masses for each pair of residues was within 5 Å was calculated. Heat maps of the inter-residue interactions were created from the matrix of residue pair interactions using Matlab (version r2017a).

Differential Scanning Calorimetry (DSC). Thermal unfolding of full-length and truncated R67 DHFRs was monitored between 25 and 95 °C using a Microcal VP differential scanning microcalorimeter. The concentration of full-length R67 DHFR was 150–160 μM in MTA buffer (100 mM MES, 50 mM Tris, and 50 mM acetic acid) at pH 8. Samples were also prepared in MTA buffer with 20% betaine or 15% DMSO. Scans were repeated two times with scan rates of 1 °C/min. For truncated R67 DHFR, concentrations of 50–100 μM were used with a scan rate of 1 °C/min. The data obtained were analyzed using Origin version 7.0 supplied by the manufacturer and the melting temperatures obtained.

Pressure Perturbation Calorimetry (PPC). Effects of betaine on the volume, or hydration, of R67 DHFR were estimated using the change in the thermal expansion coefficients (α) of the protein in the absence and presence of betaine. Pressure perturbation calorimetry (PPC) can be used to determine the α_0 value in buffer containing osmolytes using the following equation (eq 3):^{37,38}

$$\alpha_s = \alpha_0 - \frac{\Delta Q}{T m_s V_s \Delta p} \quad (3)$$

where α_s and α_0 are the thermal expansion coefficients for the solute and solvent, respectively, ΔQ is the heat released or absorbed after each application, or release, of pressure, T is the temperature, m_s is the mass of the solute in the solution, V_s is the specific volume of the solute, and Δp is the change in pressure applied above the solution. A VP-DSC instrument from MicroCal (Malvern) outfitted with a PPC appendage was used to calculate the α_s for R67 DHFR. Pressure pulses of 60 psi of nitrogen were applied above the sample, using buffer alone as a reference. The thermal expansion coefficients were determined between 10 and 95 °C in 2.5 °C increments. Samples of 3–5 mg/mL R67 DHFR (these concentrations are

equivalent to 90–150 μM for full-length and 99–150 μM for truncated R67 DHFRs) were prepared in 45 mM Na₂HPO₄, pH 8.0 buffer containing 0, 10, or 20% (w/w) betaine. Control experiments using buffer versus buffer, buffer versus water, and water versus water were used to correct the α_s of R67 DHFR for the thermal expansion of the buffer and water components of the sample (which are contained in α_0). The raw data from the PPC were manually curated so that they could be integrated using NITPIC.³⁹ Files with the heat obtained from NITPIC were used to analyze the PPC data in the Origin 7.0 software package provided by MicroCal. The mass of the solute in the solution was determined spectroscopically, and the specific volume of 0.716 mL/g for R67 DHFR was previously obtained.⁴⁰

RESULTS

SANS of Apo R67 DHFR. Representative SANS profiles for full-length and truncated apo R67 DHFR are shown in Figure 1A and Figure S3A, respectively (see the Supporting

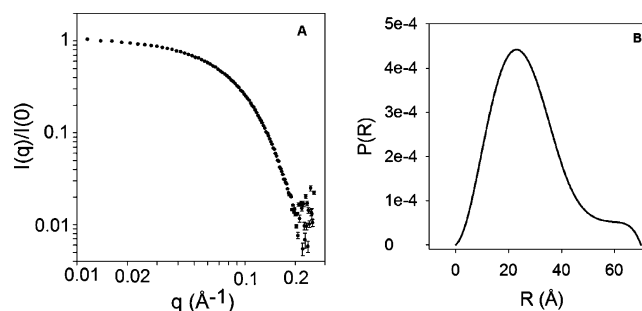


Figure 1. SANS profile and GNOM analysis for apo, full-length R67 DHFR. (A) Normalized scattering intensity of the protein, $I(q)/I(0)$, with an increase in q . The SANS profile was obtained by subtracting the scattering contribution from the buffer and normalizing the scattering intensity by $I(0)$. (B) A GNOM fit of the profile gave the pairwise distance distribution and an R_g value of 21.89 ± 0.12 Å.

Information). A dimensionless Kratky plot⁴¹ of the full-length and truncated proteins indicates that both are globular (Figure S3C). Analysis using GNOM yields an R_g value of 21.89 ± 0.12 Å for the full-length protein (see Figure 1B) and 17.86 ± 0.14 Å for truncated R67 DHFR (Figure S3B). The latter is comparable to the R_g values of 17.1 and 17.5 Å for the 2RH2⁸ and 2GQV⁹ crystal structures of truncated R67 DHFR, respectively, calculated using CRYSON.⁴² The molecular weight of R67 DHFR was calculated from the $I(0)$ and R_g of the SANS profile, using a model that is independent of protein concentration.⁴³ A value of 36470 g/mol matches well with the expected value of 33720 g/mol for full-length R67 DHFR, indicating that the sample is not aggregating under our conditions (Table 1).

As described in Methods, we used MD and the NIST program SASSIE to gain information about the space sampled by the N-termini. We generated 19 μs of MD trajectories and ~307000 structures from nondirected as well as directed Monte Carlo analysis in SASSIE. A large number of structures were used to analyze our SANS data, and the volumes sampled by the N-termini are shown in Figure S4, differentiated by the method by which they were generated (e.g., MD, Monte Carlo, or hand built). The directed Monte Carlo analysis restricted structures to an R_g range of 20.5–22.5 Å, substantially helping us find conformers that fit the SANS data. Figure 2A shows a χ^2

Table 1. Data Analysis of the SANS Profiles and Comparison of R_g Values for R67 DHFR Obtained Using GNOM and SASSIE Analyses

protein samples	$I(0)$	theoretical MW (Da)	calculated MW (Da)	GNOM R_g (Å)	no. of good fits in SASSIE ($\chi^2 < 10$)		
					no.	R_g range (Å)	mean R_g (Å)
truncated R67 DHFR	0.0227	26906	23390	17.86 ± 0.14	—	—	—
apo R67 DHFR	0.1711	33720	36470	21.89 ± 0.12	7936	20.84–23.53	21.46 ± 0.50
R67 DHFR–2NADP ⁺	0.1633	35218	33630	21.45 ± 0.14	758	20.67–22.77	21.56 ± 0.39
R67 DHFR–NADP ⁺ –DHF	0.1605	34912	35065	21.45 ± 0.18	15551	20.14–22.74	20.64 ± 0.27
apo R67 DHFR in 20% deuterated betaine	0.1148	33720	33620	23.08 ± 0.12	58277	21.04–25.94	22.78 ± 0.87

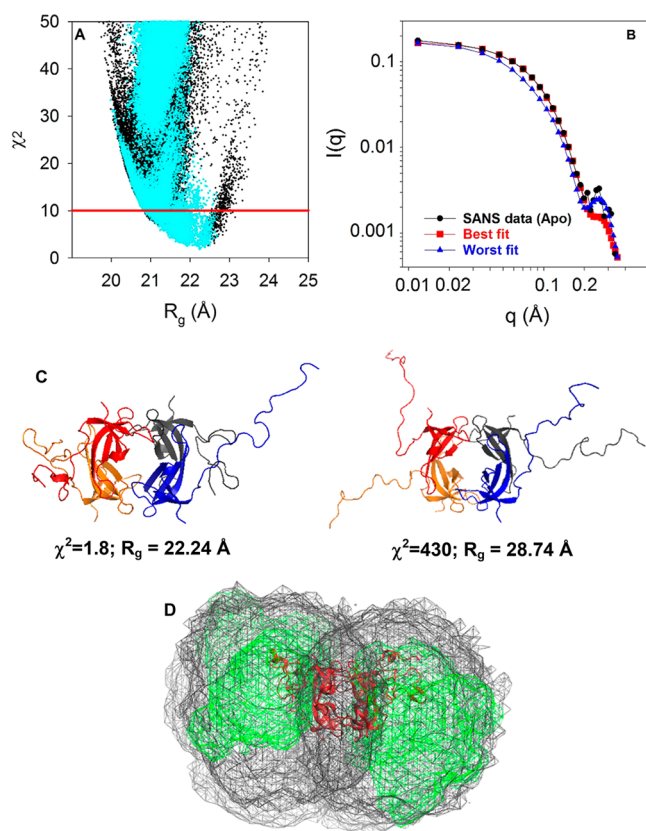


Figure 2. SASSIE analysis of apo R67 DHFR suggests compaction of the N-termini. (A) Analysis of the frames generated by directed Monte Carlo sampling (117000, black filled circles) and frames from the MD runs (154000, cyan filled circles). The red line indicates a $\chi^2 = 10$ cutoff. Data with $\chi^2 > 50$ are not shown for the sake of clarity. (B) Overlays of the theoretical SANS profiles for the best [$\chi^2 = 1.8$ (red squares)] and worst [$\chi^2 = 430$ (blue triangles)] fits compared to the experimental SANS data (black circles). (C) Corresponding best and worst structures. (D) Density plot describing all the space sampled by MD and Monte Carlo structures (dark gray mesh) representing most of the available “structural space”. The structures identified by SASSIE as providing good fits to the experimental SANS data are shown by green mesh. One of the good models describing full-length homotetrameric R67 DHFR is shown in the center of the mesh.

versus R_g plot for the 117000 frames from the directed Monte Carlo simulations and the apo frames obtained from MD runs for binary and ternary complexes upon removal of the ligands using a Python script. The χ^2 versus R_g plot shows a “U” shape, indicating neither very compact nor very extended states fit the data well. Instead, more intermediate structures fit the data. We identified 7936 structures that fit the apoenzyme SANS data with a χ^2 value of <10 . The lowest χ^2 value obtained was 1.8,

and the R_g of the corresponding frame was 22.24 Å. An average R_g value of 21.46 ± 0.50 Å was obtained (see Table 1), which is similar to the R_g value obtained from the GNOM fitting.

SASSIE also generates mesh plots that show the space sampled by the 21 N-terminal residues. All ~ 461000 structures (307000 from Monte Carlo and 154000 from MD) sample the area shown by the dark gray mesh in Figure 2D. Best fits identified by SASSIE show a more restricted area explored by the N-termini (see the green mesh in Figure 2D), indicating compaction of the N-termini as compared to full extension. This trend is also identified in a plot of the center of mass for the N-terminal methionines (see Figure S5 for those structures for which $\chi^2 < 10$). Here the tendency of the methionines to sample space mostly near the sides of the protein core can be seen. However, other methionine positions fit the data, indicating other successful sampling positions. The range of 20.84–23.53 Å (see Table 1) for the best fit R_g values generated by SASSIE also indicates sampling of compact and slightly extended conformations of the N-termini. Any asymmetry in the mesh and sampling positions likely arises from non-convergence of the MD trajectories, even though a total of 19 μ s was used. In contrast, the protein termini have millisecond to second sampling times available.

Effect of Ligand Binding on the Disordered Termini in R67 DHFR. SANS data were also collected to monitor if there were any changes in the disordered N-termini of R67 DHFR upon ligand binding. Data collected for binary (R67 DHFR–NADP⁺) and ternary (R67 DHFR–NADP⁺–DHF) complexes were analyzed using GNOM. A comparison of the pairwise distribution plots for the apo, binary, and ternary complexes is shown in Figure 3A. The R_g values for the apoprotein, NADP⁺ binary complex, and NADP⁺–DHF ternary complex are 21.89 ± 0.12 , 21.45 ± 0.14 , and 21.45 ± 0.18 Å, respectively.

To gain deeper insights into the disordered tail conformations, SANS data for the R67 DHFR–NADP⁺ binary and R67 DHFR–NADP⁺–DHF ternary complexes were further analyzed using both MD and SASSIE. SASSIE analyses used the same set of ~ 307000 frames described above for the apoenzyme, with ligands added by a Python script with 154000 frames from MD simulations. Figure S2 indicates the various steps used to generate best fit conformers. The plots shown in Figures S6 and S7 are for the analyses of binary and ternary data, respectively. Fitting our data to structures lacking the ligands did not yield any conformers with good χ^2 values. Therefore, we repeated our analyses with ligands in the protein structures. Our first fits to a single bound NADP⁺ yielded only 92 conformers with good χ^2 values, so we docked in another cofactor as two homoligands can bind in the active site pore.³⁶ This analysis yielded a total of 758 frames with χ^2 values of <10 for the binary data. The number of frames that fit the binary data is low, suggesting (1) a mixed population of species may

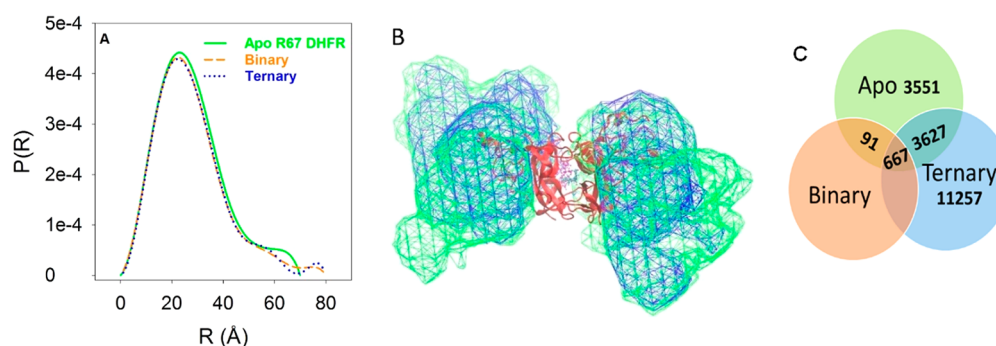


Figure 3. SASSIE analysis of ligand-bound R67 DHFR complexes indicates compaction of the N-termini. (A) GNOM fits of the pairwise distribution plots of R_g for apo R67 DHFR (green line), R67 DHFR–NADP⁺ binary (dashed orange line), and R67 DHFR–NADP⁺–DHF ternary (dotted blue line) complexes. SANS data were collected using 6.05 mg/mL DHFR in 20 mM deuterated Tris buffer in D₂O (pD 7.0) with no osmolyte. Binary or ternary complexes were formed by adding 3 mM NADP⁺ or NADP⁺ with 2 mM DHF, respectively. The R_g values for the apoprotein, binary complex, and ternary complex are 21.89 ± 0.12 , 21.45 ± 0.14 , and 21.45 ± 0.18 Å, respectively. While it may seem that the maximal diameter of the protein (D_{\max}) varies, it is more that there is not a well-defined D_{\max} value for proteins with flexible regions. Rather, a small range of D_{\max} values all appear satisfactory. Within this “optimized” range, R_g and $I(0)$ values are not changing significantly, allowing reliable parameters to be gained from the analysis. (B) Comparison of the mesh plots obtained from SASSIE for the apo (green) and NADP⁺–DHF ternary (blue) complexes. Bound DHF (cyan) and NADPH (magenta) are shown as ball-and-stick models in the center of the active site pore. (C) Venn diagram comparing the overlaps associated with the number of apo, binary, and ternary best fits. Figure S6 shows similar figures for the binary complex.

be present (i.e., both singly and doubly bound NADP⁺) and (2) our model of the 2NADP⁺ complex may only approximate this species. Fitting of the ternary complex data was more successful, yielding 15551 frames with a χ^2 of <10. The best χ^2 values for the binary and ternary complexes were 5.2 and 4.8, respectively. The mean R_g value obtained for the 758 good binary structures was 21.56 ± 0.39 Å, which is within the error of the R_g obtained by GNOM analysis. However, the mean R_g value for 15551 structures for the ternary complex was 20.64 ± 0.27 Å, which is different from the R_g value determined by GNOM analysis. The range of the R_g values in the acceptable SASSIE fits for the ternary complex was 20.14–22.74 Å, which is slightly lower than the range obtained for the apoprotein.

A comparison of density plots (or space sampled) in Figure 3B shows the N-termini of the apoprotein and ternary complex sample space at the monomer–monomer interfaces at the sides of the protein. In addition, 28% of the good fits for the ternary data fit to the apo data, indicating overlap in the conformations sampled by the N-termini in the apo and ternary complex. This also can be noted from the Venn diagram shown in Figure 3C. Also, all the 758 frames that fit to the binary data fit the apo data, suggesting similar conformational sampling of the N-termini under both conditions.

A COM for the N-terminal methionine residues in the best fit frames of the ternary complex data was again calculated. These values, represented in Figure S7E, depict sampling of a restricted space near the sides of the protein (i.e., monomer–monomer interface). A comparison of the mesh plots for the apo form and binary complex is additionally shown in Figure S6D, and a COM representation for the 758 frames for the binary analysis is shown in Figure S6F.

Effect of Betaine on the R67 DHFR Structure. As addition of osmolytes can lead to protein folding,²¹ we added 20% deuterated betaine to R67 DHFR to determine if osmolytes can provide order to the N-termini. SANS data were analyzed using GNOM; Figure 4A shows the pairwise distribution plot. The R_g was 22.84 ± 0.31 Å, which is larger than the value for R67 DHFR in the absence of betaine (e.g., 21.89 ± 0.12 Å). These results indicate a more swollen state in

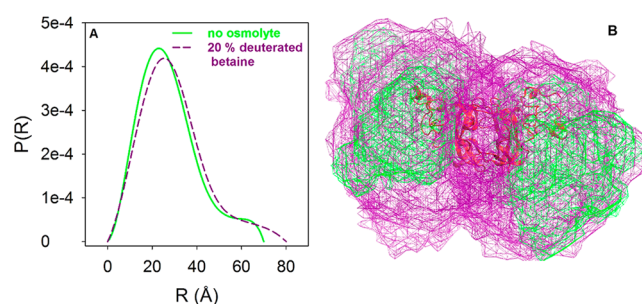


Figure 4. SASSIE analysis of apo R67 DHFR in the presence of 20% deuterated betaine indicates compact as well as partially extended N-terminal conformations. (A) GNOM fits for the pairwise distribution of R_g for apo R67 DHFR with and without 20% deuterated betaine. SANS data were collected for R67 DHFR at 6.5 mg/mL in 20 mM deuterated Tris buffer in D₂O (pD 7.0) with 20% deuterated betaine (dashed purple line). A wider distribution in the presence of betaine and an R_g value of 23.0 ± 0.30 Å indicate an increased number of slightly more extended conformations for the termini of R67 DHFR. (B) Overlay of the density plots for the best frames with χ^2 values of <10 for the deuterated betaine protein sample (purple mesh) and for the apoprotein without betaine (green mesh).

the presence of betaine. The ratio of $I(0)$, scaled by sample concentration, for R67 DHFR in the presence of betaine to that in the absence of betaine was taken to ensure that the protein in 20% deuterated betaine was not aggregating. A ratio near 1 indicated that R67 DHFR in the presence of deuterated betaine was not aggregating under our experimental conditions.

SASSIE analysis of this SANS data set was performed using the same set of 461000 conformers that were used for the apoprotein. The χ^2 versus R_g plot (Figure S8A) also shows a “U” shape, indicating intermediate rather than very compact or very extended states fit the data well. Of all the structures generated using MD and Monte Carlo sampling, 58277 frames fit to the experimental SANS data with χ^2 values that are <10. The lowest χ^2 value was 3. Figure S8A shows a $\chi^2 = 10$ cutoff for the good fits. The number of structures that fit the SANS data well has greatly increased (compared to that of apo R67 in buffer), again suggesting a more inflated structure. An overlay

of the theoretical SANS profiles for the best and worst fits and the corresponding structures for these fits are shown in panels B and C of Figure S8, respectively. The R_g for the best structure is 23.79 Å, while that for the worst is 29.58 Å. The density plot (see Figure S8D) for the good fits (purple mesh) seems to occupy most of the space sampled by our set of 461000 frames (dark gray mesh). Also, the range of R_g values obtained indicates no (or transient) sampling of fully extended conformations for all four N-termini, which would have resulted in higher R_g values. The highest R_g sampled by Monte Carlo simulation is 29.49 Å, while our model of R67 DHFR with four fully extended N-termini has an R_g of 36.25 Å.

The R_g values for the good fits obtained using SASSIE ranged from 21.04 to 25.94 Å with an average R_g value of 22.78 ± 0.87 Å. The wider sampling range and higher average R_g both are consistent with the GNOM analysis, indicating that the N-termini sample extensive conformations in the presence of betaine. The COM point for Met1 in the four N-termini (see Figure S8E) indicates the termini sample many positions both near the core of the protein and farther from the surface. From these data and analyses, the disorder in the N-termini becomes more pronounced upon addition of betaine with the N-termini potentially acting as entropic bristles, sweeping out volume around the protein core.⁴⁴

Osmolytes Probe Preferential Hydration of R67 DHFR.

Hydration is important in the protein structure and function relationship. Three regions with different scattering-length densities are present in this experiment: (1) the protein, (2) the bulk solution of deuterated buffer containing hydrogenated osmolytes, and (3) the hydration shell surrounding the protein. Betaine or DMSO was used to probe the hydration of full-length R67 DHFR, while only betaine was used for our truncated R67 DHFR experiments. While a plot of R_g values obtained by GNOM analysis does not show any significant trend upon addition of an osmolyte (Figure S9), the zero-angle scattering intensity, $I(0)$, is sensitive to changes in hydration. No significant change in the R_g value is consistent with the N-termini remaining disordered upon addition of an osmolyte. Note that, earlier, the R_g value of 22.8 ± 0.3 Å was obtained from the SANS data for apo R67 DHFR in 20% deuterated betaine. As deuterated betaine was added to the deuterated buffer, the contrast between the hydration layer and bulk was masked and the R_g value represents the overall shape of the protein without any contributions from the hydration layer. With hydrogenated betaine, the R_g is dependent upon the volume of the hydration layer and the location of the waters in the hydration shell. Even though the hydration layer contrast will increase with the addition of osmolytes, thus causing an apparent decrease in the overall R_g , our deuterated betaine data indicate the intrinsic protein volume increases. Therefore, the expansion of the intrinsic volume along with the apparent decrease in size from hydration contrast may be compensating for each other, and the net result is a relatively consistent R_g for R67 DHFR in the presence of hydrogenated osmolytes.

Decreasing $I(0)$ values for both full-length and truncated R67 DHFR were observed with increasing concentrations of osmolytes. As shown in Figure 5, the data were fit to eq 2. The fits yield the volume of the hydration layer for R67 DHFR in the presence of osmolytes. The number of water molecules in the hydration layer is determined by dividing the observed water volume by the volume of a single water molecule (30 \AA^3). The number of osmolyte-excluding water molecules associated with the full-length protein is 1285 ± 214 or 1253 ± 199 using

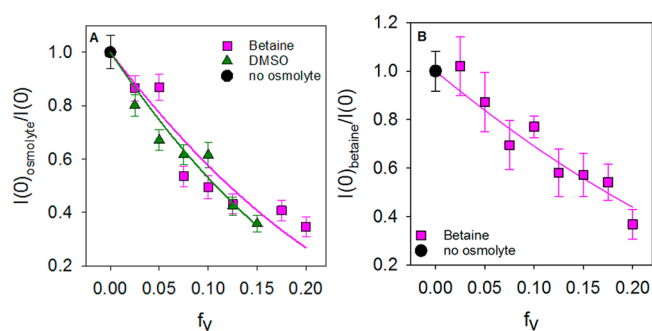


Figure 5. Preferential hydration of the full-length and truncated R67 DHFR in the presence of osmolytes. Small angle neutron scattering intensity ratios with and without osmolyte $[I(0)_s/I(0)]$ as a function of osmolyte concentration, f_v , for betaine (w/v) (magenta filled squares) and DMSO (v/v) (green filled triangles). Panels A and B show the data for full-length and truncated R67 DHFRs, respectively. Solid lines are fits to eq 2 to calculate the number of protein-associated waters, n_w .

betaine or DMSO, respectively, indicating similar numbers of water molecules. The number of water molecules (n_w) excluding betaine from the hydration shell of the truncated protein is 380 ± 100 . The difference in the number of hydrating waters (~ 900) between full-length and truncated R67 DHFR indicates that the disordered N-terminal tails span a large volume in solution and are extensively hydrated.

To compare the experimental hydration values with theoretical numbers, the solvent accessible surface area (ASA) of tetrameric, truncated apo R67 DHFR (2RH2) was calculated to be 11072 \AA^2 using the Molecular Operating Environment program (MOE 2015 version). If we assume the area of a water molecule to be 9 \AA^2 ,⁴⁵ this yields approximately 1230 water molecules hydrating the truncated protein. For another high-resolution crystal structure of truncated R67 DHFR (2GQV),⁹ the solvent accessible surface area was 11673 \AA^2 . This structure predicts ~ 1297 water molecules.

To obtain a theoretical n_w value associated with the full-length R67 DHFR, we used the 7936 good fits obtained from our SASSIE analysis. The ASA was determined using SurfaceRacer⁴⁶ for each of the 7936 frames to obtain the n_w values. An average of 1800 water molecules were predicted in the hydration layer. Table 2 compares our experimental results with the predicted values from the truncated crystal structures as well as the average value for the 7936 full-length models of R67 DHFR. The predicted ranges of values for waters in the hydration shell are higher than those measured by our SANS data.

Effect of Osmolytes on the Thermal Stability of R67 DHFR. DSC scans were performed to monitor the effects of betaine and DMSO on the thermal stability of R67 DHFR. This is an additional way to determine if osmolytes are excluded from the protein surface. Previous thermal denaturation studies of R67 DHFR at pH 8 have found reversible folding with a melting temperature of $70.95 \text{ }^\circ\text{C}$ and evidence of an intermediate state.⁴⁷ Our DSC scans are shown in Figure 6. The data were fit to a three-state model, giving two melting temperatures (T_M) that correspond to two events in the thermal unfolding of R67 DHFR. T_{M1} and T_{M2} values in the absence of an osmolyte are 66.8 and $68.7 \text{ }^\circ\text{C}$, respectively. Addition of 20% betaine increased the melting temperature of R67 DHFR by $2\text{--}3 \text{ }^\circ\text{C}$, while 15% DMSO decreased the T_M by $7\text{--}9 \text{ }^\circ\text{C}$ (see Table S2). Stabilization of R67 DHFR in the presence of betaine is consistent with preferential exclusion of

Table 2. Comparison of the Predicted and Experimental Numbers of Water Molecules (n_w) Hydrating R67 DHFR from the Crystal Structure, SASSIE Fits, and SANS Data

protein	source	predicted/experimental	no. of water molecules in the hydration layer (n_w)
truncated R67 DHFR	crystal structure (2RH2) ^a	predicted from ASA	1230 ^a
truncated R67 DHFR	crystal structure (2GQV) ⁹	predicted from ASA	1297 ^b
truncated R67 DHFR	crystal structure (2GQV) ⁹	experimental	340 in pore and first hydration shell
truncated R67 DHFR in the presence of betaine	SANS	experimental	380 ± 105
full-length R67 DHFR	frames from SASSIE analysis ($\chi^2 < 10$)	predicted from ASA averaged for 7936 frames	1800
full-length R67 DHFR in the presence of betaine	SANS	experimental	1285 ± 214
full-length R67 DHFR in the presence of DMSO	SANS	experimental	1253 ± 199

^aThe 2RH2 structure lacks 20 residues at the N-termini. ^bThe 2GQV structure lacks 19 residues at the N-termini. Serine 20 was removed for comparison with 2RH2.

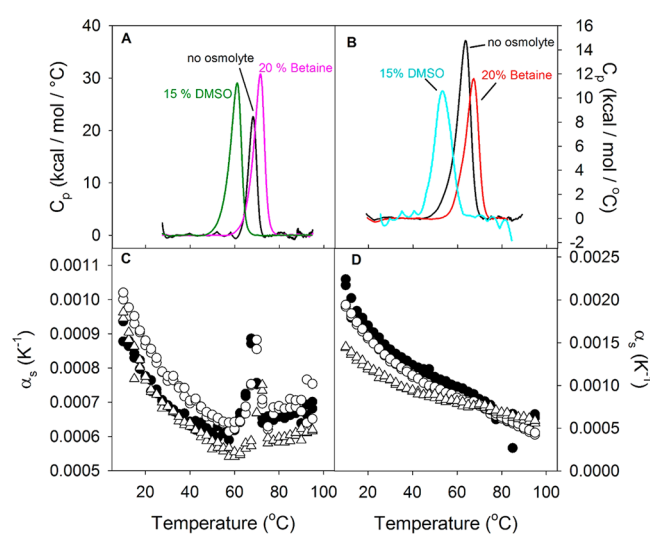


Figure 6. (A) Effects of osmolytes on thermal denaturation of full-length R67 DHFR. DSC scans were performed with 150–160 μ M R67 DHFR in MTA buffer with and without 20% betaine (magenta) and 15% DMSO (green). Betaine increases the melting temperature of the protein by 2–3 $^{\circ}$ C, whereas DMSO decreases it by 7–9 $^{\circ}$ C. (B) DSC thermograms for truncated R67 (50–90 μ M) DHFR in MTA buffer, pH 8.0 buffer (black line), or buffer with 20% betaine (red) or 15% DMSO (cyan). Panels C and D show our PPC results. Thermal expansion coefficients (α_s) were obtained from PPC using 2.5 $^{\circ}$ C increments between 10 and 95 $^{\circ}$ C for (C) full-length R67 DHFR (90–150 μ M) or (D) truncated R67 DHFR (99–150 μ M). Experiments were performed in 45 mM Na_2HPO_4 , pH 8.0 buffer alone (\bullet), buffer with 10% betaine (\circ), or buffer with 20% betaine (Δ).

betaine from the protein surface. DMSO slightly destabilizes R67 DHFR, which indicates the likely interaction of DMSO with R67 DHFR, in the native or unfolded state.

DSC was also performed on truncated R67 DHFR. The two T_M values were decreased 5 and 7 $^{\circ}$ C, respectively, compared to those of full-length protein, consistent with the disordered N-termini stabilizing the enzyme (see Table S2). Addition of 20% betaine to truncated R67 DHFR resulted in stabilization of both T_M values by 4 $^{\circ}$ C, while addition of 15% DMSO destabilized the protein by 7 $^{\circ}$ C. It has been reported that DSC of intrinsically disordered proteins or regions does not show an unfolding transition.^{48,49} If true, then DSC signals of full-length and truncated R67 DHFR should report on the unfolding of the core “doughnut” structure. Also, addition of solutes should have similar effects on full-length and truncated R67 DHFRs. This behavior is seen in Figure 6. As addition of betaine increases the T_M values for both R67 DHFR species by similar levels (Table S2), it stabilizes the protein core, most likely by being excluded. In other words, the protein favors interaction with water.^{22,50} In contrast, DMSO destabilizes the R67 core structure as the T_M values are lowered to similar degrees for both full-length and truncated protein species. This behavior suggests a preferential interaction mechanism for DMSO with the protein.^{22,50} Again, DSC appears to report on the core, folded structure.

Pressure Perturbation Calorimetry. Another avenue for exploring the molar volume of R67 DHFR in the presence of betaine uses pressure perturbation calorimetry. From PPC, the thermal expansion coefficient for R67 DHFR was $8.7 \times 10^{-4} \text{ K}^{-1}$ at 10 $^{\circ}$ C (Table 3), and it decreased as the temperature increased to 57 $^{\circ}$ C (Figure 6C). Structure-breaking polar groups on the surface of the protein are most likely responsible for the decrease in α_s .³⁸ The denaturation transition of R67

Table 3. Analysis of the PPC Data for Full-Length and Truncated R67 DHFRs in 45 mM Na_2HPO_4 , pH 8.0 Buffer with 0, 10, or 20% (w/w) Betaine

DHFR	[betaine] (%)	$\Delta V/V$	$\alpha_{s,10-40} (\times 10^{-4} \text{ K}^{-1})$	$T_M (^{\circ}\text{C})$	$\alpha_{s,10} (\times 10^{-4} \text{ K}^{-1})$	$\Delta\alpha_s (\times 10^{-5} \text{ K}^{-1})$
full-length R67 DHFR	0	0.0013	2.7	67.5	8.7	4.3
	10	0.00077	3.1	70.0	9.7	4.8
	20	0.00048	3.2	72.4	8.9	3.3
truncated R67 DHFR	0	0	9.7	nd ^a	22.1	nd ^a
	10	0	8.2	nd ^a	19.3	nd ^a
	20	0	5.1	nd ^a	14.5	nd ^a

^aNot determined as there was no thermal denaturation transition in the PPC thermogram.

DHFR between 60 and 75 °C caused an increase in α_s , which indicates an increase in the volume of R67 DHFR as the protein denatures. Integrating the area underneath this peak in the thermogram yielded a relative change in volume ($\Delta V/V$) of +0.0013. A T_M for the denaturation of R67 DHFR of 67.5 °C was calculated from the PPC. This value matches well with the conventional DSC analysis of R67 DHFR at pH 8 (Table S2).⁴⁷

As a control, PPC was also performed on the truncated form of the protein (see Figure 6D). The α_s at 10 °C for truncated R67 DHFR ($2.2 \times 10^{-3} \text{ K}^{-1}$) was twice that of the full-length protein (Table 3). Another interesting characteristic of the PPC for the truncated R67 DHFR was that no transition for denaturation was noted (Figure 6D). As there is a clear transition in the DSC data for truncated DHFR (see Figure 6B), the lack of a denaturation transition in the PPC thermogram is not due to the protein being unfolded. Additionally, we note the truncated enzyme was active, indicating that it was not unfolded.

A balance exists between elements that contribute to the negative volume change (i.e., loss of voids in the protein and the electrostriction of water around polar and charged groups that are more exposed upon unfolding) and those that contribute to a positive volume change (a larger thermal expansivity for the unfolded state vs the folded state and changes in the hydrophilic–hydrophobic balance of the exposed groups).^{38,51} Additional effects may be loss of clathrate water in the R67 active site pore⁹ and dissociation of a tetramer to four unfolded monomers. The relative contribution of these effects leads to the observed α value. For a positive $\Delta V/V$ (as in full-length R67 DHFR), the positive effects must predominate. For $\Delta V/V$ to be zero (as in truncated R67 DHFR), the various effects appear to be balanced. Most monomeric, globular proteins show negative $\Delta V/V$ values.^{51,52} Because the structural differences between the truncated and full-length R67 DHFRs are the four disordered N-termini, they appear to be the key determinant for the positive $\Delta V/V$ value seen for full-length R67.

PPC thermograms were also performed in the presence of betaine for both full-length and truncated R67 DHFRs (Figure 6). Addition of 10% (w/w) betaine to the full-length protein caused an increase in the α_s value at 10 °C ($9.7 \times 10^{-4} \text{ K}^{-1}$) relative to the protein in the absence of betaine (Table 3). This increase in α_s suggests there is an increase in the protein volume that is likely due to extension of the collapsed N-termini in the presence of betaine. Further increasing the betaine concentration to 20% (w/w) decreased the α_s at 10 °C to $8.9 \times 10^{-4} \text{ K}^{-1}$, similar to the value in the absence of betaine. This reduction most likely describes a decrease in the size of the solvation shell of the full-length protein. Similar effects of various solutes on the α values for RNase³⁸ and SNase⁵³ have been previously observed and ascribed to effects on the hydration shell. The increased volume at 10% betaine correlates with our SANS result of an increased R_g for full-length R67 DHFR in buffer containing 20% deuterated betaine.

For truncated R67 DHFR, the α_s value at 10 °C decreases with each increase in betaine concentration from $1.9 \times 10^{-3} \text{ K}^{-1}$ at 10% betaine to $1.5 \times 10^{-3} \text{ K}^{-1}$ at 20% betaine. The volume of the truncated protein, including its water shell, decreases as betaine is added, decreasing the concentration of water in the solution.

DISCUSSION

The R67 DHFR monomer is 78 amino acids long, and around 16–20 N-terminal residues are disordered; therefore, ~20–25% of its sequence is unstructured. R67 assumes a compact structure by forming a homotetramer. Chymotrypsin treatment of the folded protein results in a truncated product, which is almost fully active but 2.6 kcal/mol less stable.⁷ Expression of the truncated protein from a shorter gene sequence does not confer trimethoprim resistance. Thus, the N-termini are essential for protein expression and/or stability but not for catalysis. To understand the conformational space sampled by the N-termini of R67 DHFR, we characterized full-length and truncated R67 DHFR using SANS.

Apoprotein Analysis. The best fits for apo R67 DHFR indicate compaction of two N-termini on one side of the ordered tetramer core, whereas the other two N-termini prefer to remain partially extended (see Figure 2D and Figure S5). In many of these poses, the N-terminal residues interact with each other and/or with residues exposed on the monomer–monomer interface. Intramolecular and intermolecular interactions are both feasible. These interactions lead to compaction of the overall shape and seem likely to be why the N-termini provide 2.6 kcal/mol of stability to R67 DHFR.⁷

Data mining of the conformers fitting the SANS profile was accomplished using a Python script. Frequent interactions were identified by counting the number of times the COM of each amino acid occurs within 5 Å of the COM of every other residue. Figure S10A shows a heat map of the minimum distance between residues. Figure S10B provides a heat map of the number of these interactions versus the amino acid number (1–78 for the first monomer, 79–156 for the second monomer, 157–234 for the third monomer, and 235–312 for the last monomer). The symmetry of the structure provides an initial understanding of these plots as monomers nearby in space interact (A and C or B and D), while distant monomers do not. In Figure S10, intramolecular interactions can be visualized by the points near the diagonal while intermolecular interactions are indicated by the areas describing interactions between residues 1–78 and 157–234 (for example).

Using the symmetry of the core structure, the number of potential interactions was summed, using the rationale that a stabilizing interaction would occur in more than one monomer. Supplemental Excel sheet 1 lists these amino acid pairs. Three bins were noted: first, pairs that occur more than 1000 times (=22); second, intramolecular pairs that occur in all four monomers (=8); and third, intermolecular pairs that occur in all four monomers (=1). Hydrophobic, polar/uncharged, and charged residues are identified and colored in the excel sheet as described by Eisenberg et al.⁵⁴ In the pairs that occur >1000 times, hydrophobic residues occur 50% of the time while polar/uncharged amino acids occur 40% of the time and charged residues 10% of the time. These pairs mostly describe N-terminal to N-terminal interactions.

As the N-terminal sequence contains several hydrophobic side chains (M1, I2, V8, A13, F16, V17, and F18), these amino acids could also potentially form hydrophobic interactions with similar exposed side chains on the folded protein surface. In particular, each of the two-symmetry related W45 residues provides ~94 Å² of ASA for interaction. Short distances were observed from most of the hydrophobic residues mentioned above to W45 and its symmetry-related W201 residue. Also,

cation- π interactions could be transiently occurring as R3 often occurred nearby W45 as well as M1 (N-terminal residue).

Other residues contributing to hydrophobic surfaces near the monomer-monomer interface include A22, F24, M26, V30, V43, V71, A72, and I77. Spatial proximity was noted for several N-terminal residues with A22, M26, and V43. Interestingly, three 2-methylpentane-2,4-diol molecules per monomer were found close to the A22, W45, and L50 residues in the 2RH2 structure,⁸ suggesting a possible hydrophobic interaction hot spot. Figure 7 shows exposed hydrophobic residues on the ASA surface of R67 DHFR.

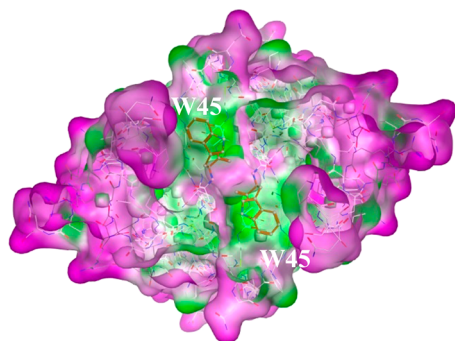


Figure 7. Surface of R67 DHFR (2RH2) shown with the exposed lipophilic surfaces colored green and hydrophilic surfaces colored magenta along the monomer-monomer interface. The surface was made partially transparent so the side chains could be seen. Two symmetry-related W45 residues are labeled in the center of the figure. This figure was composed using the lipophilic surface option in MOE (version 2015.10).

The crystal structure of truncated R67 DHFR shows exact 222 symmetry.⁴ While this symmetry could also apply to each of the disordered N-termini, it is more likely that they impart asymmetry via their disorder.

Analysis of Binary and Ternary Complexes. Table 2 summarizes the various R_g values obtained from GNOM and SASSIE analysis. No substantial effect of ligand binding was observed in the conformations sampled by the N-termini as those frames that provided the best fits to the SANS data for both the binary and ternary complexes mostly overlap with those sampled by the apoprotein. GNOM analysis yielded comparable R_g values for the apoprotein and ligand-bound protein samples, and the conformers obtained from our SASSIE analysis placed the disordered tails near the sides of the active site pore.

Data mining of the ternary complex conformers that fit the SANS data was also performed. Figure S11A plots the minimum distance between the COM of residues. A pattern similar to that seen in apo R67 DHFR is observed. Supplementary Excel sheet 2 lists those amino acid pairs whose centers of mass are ≤ 5 Å apart. Three bins were again considered: 47 pairs occur more than 1000 times, while nine intramolecular and two intermolecular pairs occur in all four monomers. The same type of interactions are observed as in the apo conformers with hydrophobic residues occurring 56% of the time, polar/uncharged 32% of the time, and charged 13% of the time. One difference is that the N-terminal methionines now very frequently interact with W45 or W201 (symmetry-related residues).

Effects of Osmolytes. The main difference in our data arises when betaine is added, which leads to a more swollen

state of R67 DHFR. Osmolytes that are excluded from protein surfaces are known to stabilize the protein via the preferential exclusion mechanism.²² The ability of TMAO to force folding of a modified RNase was attributed to its preferential exclusion from the peptide backbone (also termed the solvophobic effect).^{21,55} While R67 DHFR was found to be stabilized upon addition of betaine by our DSC studies, no disorder to order transition was observed for the disordered tails from our analysis of the SANS data. On the contrary, SASSIE finds the addition of betaine results in greater conformational sampling of the disordered tails, from being collapsed near the sides of the protein to being partially extended. In our previous studies of the interaction of betaine with folate and other compounds, we found betaine can compete with water to form stable interactions.⁵⁶ Betaine prefers to interact with aromatic surfaces as well as cationic and amide nitrogen atoms, while water prefers to interact with carboxylate, phosphate amide and hydroxyl oxygens.^{56,57} For the case of R67 DHFR, betaine may interact with some residues in the N-termini, hindering the collapsed conformations from being sampled.

Another possible explanation for the extensive sampling of the disordered tails upon addition of betaine may be attributed to changes in the solvent structure. Studies have characterized the effects of solutes on the structure of bulk as well as hydrating water molecules around proteins.^{37,38} The nature and extent of these alterations depend on the chemical properties of the solutes. Polar and hydrophilic surfaces were found to be water structure breakers, whereas hydrophobic surfaces were described as water structure makers.³⁷ Stabilization of RNase A by 1.5 M sucrose was previously observed, while accompanying pressure perturbation calorimetry studies showed nonlinear effects on α , the apparent coefficient of thermal expansion. Specifically, RNase is less compact in 0.5 M sucrose, as indicated by an increased α , than in the absence of sucrose, while the protein becomes more compact at 1.5 M sucrose, yielding a decreased α .³⁸ The differences in α were attributed to changes in protein hydration.

Hydration Studies. Experiments that have examined protein hydration have used varying techniques. A typical approach calculates the accessible surface area (ASA) and divides the value by 9 \AA^2 to predict the number of solvent waters in the hydration shell. This yields a high value. In contrast, experimental approaches often yield smaller numbers of hydration waters. For lysozyme, ASA calculations predict ~ 900 waters of hydration.²⁷ Experimental techniques for studying lysozyme hydration include NMR,⁵⁸ excess heat capacity,⁵⁹ dielectric relaxation,⁶⁰ and X-ray diffraction.⁶¹ These experimental approaches yield 121–900 hydration waters, indicating the value is sensitive to the technique used as well as the experimental conditions employed. A previous SANS study of hydration in lysozyme used different osmolytes.²⁷ With added betaine, triethylene glycol, PEG400, or PEG1000, 84 ± 5 , 114 ± 24 , 156 ± 8 , or 347 ± 11 hydration waters were observed, respectively, along with different water shell thicknesses. The increase in the number of waters (n_w) may be due to osmotic stress effects combined with volume exclusion as the size of the osmolyte increases.^{45,62,63} Alternatively, fewer waters may be observed if the osmolyte interacts with the protein surface. Both factors likely play a role in observation of an n_w value that is lower than the predicted upper limit.

In our SANS studies of R67 DHFR, we used the osmoprotectant, betaine, as it is often excluded from the

protein surface.⁶⁴ Our SANS experiments allow three areas of different contrast to be delineated: the protein, the bulk solution containing osmolytes, and the hydration shell that excludes osmolytes. The number of water molecules (n_w) responsible for the exclusion of betaine from the truncated R67 DHFR surface was found to be 380 ± 105 . This value is smaller than the predicted value of 1230–1297 waters from ASA calculations from the crystal structures (2RH2⁸ and 2GQV⁹).

Because of the high resolution and low temperature factors of the 1.1 Å resolution structure of R67 DHFR (2GQV), 85 waters per monomer were identified in the first hydration shell (e.g., formation of a H-bond with the protein surface) and 106 in higher-level shells.⁹ This yields 340 waters in the first hydration shell of the tetramer. This value compares to that from our SANS experiment with truncated R67 DHFR that yields an n_w value of 380 ± 100 . Thus, both SANS and crystallography appear to measure polar bound waters.

When SANS was performed on full-length R67 DHFR, addition of both betaine and DMSO yielded n_w values of ~ 1250 waters hydrating the protein surface. This value is lower than the average n_w value of 1800 waters predicted using ASA calculations on PDB files generated by MD and directed Monte Carlo analyses in SASSIE. We also counted the average number of waters associated with the R67 DHFR tetramer in our MD trajectories and found an average of 1647 (range of 1446–1879). Again, the experimental value is lower than the predicted upper limit, suggesting some level of interaction of the osmolyte with the protein surface.

When the n_w values for truncated (380) and full-length R67 DHFRs (1200) are compared, the difference is 900 waters. This indicates each N-terminus is well-hydrated by ~ 225 waters.

To test whether betaine and DMSO were interacting with R67 DHFR, we performed DSC experiments. Excluded osmolytes typically increase the stability of proteins by increasing the level of hydration, while interacting osmolytes decrease protein stability.^{21,65} Additionally, DSC experiments of intrinsically disordered proteins (IDPs) typically lack cooperative structural transitions,^{48,49,66} so our results appear to report on the effects of the osmolyte on the structural core of the protein. This idea is supported by a similar 4–5 °C increase in T_M when betaine is added to either truncated or full-length R67 DHFR. Thus, betaine appears to be mostly excluded from the surface of the core of the R67 DHFR structure (supported by DSC results), while there is some level of interaction of the osmolyte with the disordered N-termini (supported by our deuterated betaine SANS and PPC results).

Addition of 20% betaine increased the T_M values by 4–5 °C for both full-length and truncated R67 DHFRs, while addition of 20% DMSO decreased the T_M values by 5–7 °C. These results were surprising given that the numbers of hydrating waters for these two osmolytes were within error as measured by our SANS experiments. Though the n_w values are similar, the water location may vary. DMSO can form hydrophobic interactions, whereas betaine interacts with aromatic, amide, and cationic nitrogens exposed on the protein. Thus, both osmolytes may lead to the exclusion of water from different protein surfaces, which can in turn result in the variable effects on protein stability.

CONCLUSION

While it is confounding that disordered regions can provide some level of stability to a folded protein, we find this is the case for R67 DHFR. From our SANS data, we find the

disordered N-termini prefer to sample conformational space near the sides of the apoprotein. This allows both N-termini to interact with themselves as well as the monomer–monomer interface, providing 2.6 kcal/mol of stability to R67 DHFR.⁷

According to van der Lee et al.,⁶⁷ entropic chains are a form of IDP that remain disordered. This applies to the N-termini of R67 DHFR as they do not fold upon addition of a ligand. Addition of betaine to R67 DHFR results in a larger R_g and SASSIE fits that predict a wider sampling volume. These results suggest the N-termini are responsive to their environment. It is tempting to speculate that in the cell, the N-termini may interact in a similar fashion with small molecules or macromolecules and provide an entropic bristle function where they sweep out volume around the protein core. This would prevent large molecules from entering this space but allow penetration of small molecules.⁶⁸ Entropic bristles have also been proposed to enhance protein solubility and prevent aggregation.^{44,69–71} Finally, we note there are several R-plasmid DHFRs that differ only in the sequence of their N-termini.^{10–12} All have N-termini of similar lengths, which indicates it may be the length of the N-termini more than the sequence that is important for its function. A longer disordered sequence (as in our His-tag constructs cloned in pRSETB with an additional 30 amino acids) leads to ~ 2 -fold increases in K_m values for NADPH and DHF.⁷² Another study found the R67 DHFR N-terminal sequence was essential for evolvability.⁷³ These various observations suggest that R67 DHFR's disordered N-termini play roles in stability, solubility, evolvability, and substrate access.

Finally, both the betaine studies and the preferential hydration measurements indicate the disordered tails are highly hydrated, consistent with large, exposed surface areas. These data sets support the importance of water and solutes in the R67 structure–function relationship. As the disordered segments are exposed and our SANS results show the polar regions are well hydrated, it seems likely that betaine can compete well with water for solvation of aromatic groups. Indeed, Uversky suggested intrinsically disordered proteins (IDPs) are “multifarious interactors”.⁶⁸ While he meant IDP can often interact with various protein partners, here we wonder if disordered regions can interact with different solutes, which can subtly change their behavior. This would add another layer of complexity to the role of IDP and disordered regions in the cell.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.biochem.7b00822.

A 20% SDS–PAGE gel indicating clear separation of full-length and truncated R67 DHFRs, a flow diagram of our steps using MD and SASSIE to find N-terminal conformers that fit the SANS data, another figure depicting the SANS profile and GNOM analysis for truncated R67 DHFR as well as a dimensionless Kratky plot comparing full-length and truncated R67, the COM position for the N-terminal methionine of apo R67 DHFR conformers that fit the SANS data, SASSIE analysis of the data of the binary complex (R67 DHFR–NADP⁺), the ternary complex (R67 DHFR–NADP⁺–DHF), and R67 DHFR in the presence of 20%

deuterated betaine, plots of R_g for full-length and truncated R67 DHFR probed after addition of betaine or DMSO, data mining plots for apo and ternary complex fits, and two tables listing R_g values from the various programs and T_M values from DSC data (PDF) Excel sheet from the data mining of apo conformers (XLSX)

Excel sheet from the data mining of ternary conformers (XLSX)

AUTHOR INFORMATION

Corresponding Author

*Department of Biochemistry and Cellular and Molecular Biology, University of Tennessee, Knoxville, TN 37996-0840. Phone: 865-974-4507. Fax: 865-974-6306. E-mail: lzh@utk.edu.

ORCID

Pratul Agarwal: 0000-0002-3848-9492

Christopher Stanley: 0000-0002-4226-7710

Elizabeth E. Howell: 0000-0001-6157-433X

Funding

This work was supported by National Institutes of Health Grants GM 110669 (to E.E.H.) and GM105978 (to P.A.).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

A portion of this research at ORNL's Spallation Neutron Source was sponsored by the Scientific User Facilities Division, Office of Basic Energy Sciences, U.S. Department of Energy. The authors thank Susan Krueger from the National Institute of Standards and Technology (NIST) for her help with SASSIE analysis. This work benefitted from CCP-SAS software developed through a joint EPSRC (EP/K039121/1) and NSF (CHE-1265821) grant.

ABBREVIATIONS

ASA, accessible surface area; COM, center of mass; DHF, dihydrofolate; DHFR, dihydrofolate reductase; DMSO, dimethyl sulfoxide; DSC, differential scanning calorimetry; IDP, intrinsically disordered protein; $I(0)$, zero-angle scattering intensity; MD, molecular dynamics; MTA buffer, 100 mM MES, 50 mM Tris, and 50 mM acetic acid; NADP⁺ and NADPH, oxidized and reduced nicotinamide adenine dinucleotide phosphate, respectively; NMR, nuclear magnetic resonance; NTA, nitrilotriacetic acid; n_w , number of water molecules; PDB, Protein Data Bank; PEG, polyethylene glycol; PPC, pressure perturbation calorimetry; q , scattering angle; R_g , radius of gyration; SANS, small angle neutron scattering; THF, tetrahydrofolate; T_M , melting temperature.

REFERENCES

(1) Toulouse, J. L., Edens, T. J., Alejandre, L., Manges, A. R., and Pelletier, J. N. (2017) Integron-Associated DfrB4, a Previously Uncharacterized Member of the Trimethoprim-Resistant Dihydrofolate Reductase B Family, Is a Clinically Identified Emergent Source of Antibiotic Resistance. *Antimicrob. Agents Chemother.* 61, e02665-16.

(2) Bastien, D., Ebert, M. C., Forge, D., Toulouse, J., Kadnikova, N., Perron, F., Mayence, A., Huang, T. L., Vanden Eynde, J. J., and Pelletier, J. N. (2012) Fragment-based design of symmetrical bis-benzimidazoles as selective inhibitors of the trimethoprim-resistant, type II R67 dihydrofolate reductase. *J. Med. Chem.* 55, 3182–3192.

(3) Toulouse, J. L., Abraham, S. M. J., Kadnikova, N., Bastien, D., Gauchot, V., Schmitzer, A. R., and Pelletier, J. N. (2017) Investigation of Classical Organic and Ionic Liquid Cosolvents for Early-Stage Screening in Fragment-Based Inhibitor Design with Unrelated Bacterial and Human Dihydrofolate Reductases. *Assay Drug Dev. Technol.* 15, 141–153.

(4) Narayana, N., Matthews, D. A., Howell, E. E., and Xuong, N. (1995) A plasmid-encoded dihydrofolate reductase from trimethoprim-resistant bacteria has a novel D2-symmetric active site. *Nat. Struct. Mol. Biol.* 2, 1018–1025.

(5) Feng, J., Grubbs, J., Dave, A., Goswami, S., Horner, C. G., and Howell, E. E. (2010) Radical redesign of a tandem array of four R67 dihydrofolate reductase genes yields a functional, folded protein possessing 45 substitutions. *Biochemistry* 49, 7384–7392.

(6) Matthews, D. A., Smith, S. L., Baccanari, D. P., Burchall, J. J., Oatley, S. J., and Kraut, J. (1986) Crystal structure of a novel trimethoprim-resistant dihydrofolate reductase specified in *Escherichia coli* by R-plasmid R67. *Biochemistry* 25, 4194–4204.

(7) Reece, L. J., Nichols, R., Ogden, R. C., and Howell, E. E. (1991) Construction of a synthetic gene for an R-plasmid-encoded dihydrofolate reductase and studies on the role of the N-terminus in the protein. *Biochemistry* 30, 10895–10904.

(8) Krahn, J. M., Jackson, M. R., DeRose, E. F., Howell, E. E., and London, R. E. (2007) Crystal structure of a type II dihydrofolate reductase catalytic ternary complex. *Biochemistry* 46, 14878–14888.

(9) Narayana, N. (2006) High-resolution structure of a plasmid-encoded dihydrofolate reductase: pentagonal network of water molecules in the D2-symmetric active site. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 62, 695–706.

(10) Amyes, S. G., and Smith, J. T. (1976) The purification and properties of the trimethoprim-resistant dihydrofolate reductase mediated by the R-factor, R388. *Eur. J. Biochem.* 61, 597–603.

(11) Flensburg, J., and Steen, R. (1986) Nucleotide sequence analysis of the trimethoprim resistant dihydrofolate reductase encoded by R plasmid R751. *Nucleic Acids Res.* 14, 5933.

(12) Alonso, H., and Gready, J. E. (2006) Integron-sequestered dihydrofolate reductase: a recently redeployed enzyme. *Trends Microbiol.* 14, 236–242.

(13) Strader, M. B. (2003) Identifying the Catalytic and Ligand Binding Roles of Active Site Residues in Homotetrameric R67 Dihydrofolate Reductase. Ph.D. Dissertation, University of Tennessee, Knoxville, TN.

(14) Duff, M. R., Jr., Chopra, S., Strader, M. B., Agarwal, P. K., and Howell, E. E. (2016) Tales of Dihydrofolate Binding to R67 Dihydrofolate Reductase. *Biochemistry* 55, 133–145.

(15) Schmitzer, A. R., Lepine, F., and Pelletier, J. N. (2004) Combinatorial exploration of the catalytic site of a drug-resistant dihydrofolate reductase: creating alternative functional configurations. *Protein Eng., Des. Sel.* 17, 809–819.

(16) Bradrick, T. D., Shattuck, C., Strader, M. B., Wicker, C., Eisenstein, E., and Howell, E. E. (1996) Redesigning the quaternary structure of R67 dihydrofolate reductase. Creation of an active monomer from a tetrameric protein by quadruplication of the gene. *J. Biol. Chem.* 271, 28031–28037.

(17) Dam, J., and Blondel, A. (2004) Effect of multiple symmetries on the association of R67 DHFR subunits bearing interfacial complementing mutations. *Protein Sci.* 13, 1–14.

(18) Dam, J., Rose, T., Goldberg, M. E., and Blondel, A. (2000) Complementation between dimeric mutants as a probe of dimer-dimer interactions in tetrameric dihydrofolate reductase encoded by R67 plasmid of *E. coli*. *J. Mol. Biol.* 302, 235–250.

(19) Yachnin, B. J., Colin, D. Y., Volpato, J. P., Ebert, M., Pelletier, J. N., and Berghuis, A. M. (2011) Novel crystallization conditions for tandem variant R67 DHFR yield a wild-type crystal structure. *Acta Crystallogr., Sect. F: Struct. Biol. Cryst. Commun.* 67, 1316–1322.

(20) Ebert, M. C., Morley, K. L., Volpato, J. P., Schmitzer, A. R., and Pelletier, J. N. (2015) Asymmetric mutations in the tetrameric R67 dihydrofolate reductase reveal high tolerance to active-site substitutions. *Protein Sci.* 24, 495–507.

- (21) Baskakov, I., and Bolen, D. W. (1998) Forcing thermodynamically unfolded proteins to fold. *J. Biol. Chem.* 273, 4831–4834.
- (22) Parsegian, V. A., Rand, R. P., and Rau, D. C. (2000) Osmotic stress, crowding, preferential hydration, and binding: A comparison of perspectives. *Proc. Natl. Acad. Sci. U. S. A.* 97, 3987–3992.
- (23) Biedermannova, L., and Schneider, B. (2016) Hydration of proteins and nucleic acids: Advances in experiment and theory. A review. *Biochim. Biophys. Acta, Gen. Subj.* 1860, 1821–1835.
- (24) Rani, P., and Biswas, P. (2015) Diffusion of Hydration Water around Intrinsically Disordered Proteins. *J. Phys. Chem. B* 119, 13262–13270.
- (25) Rani, P., and Biswas, P. (2015) Local Structure and Dynamics of Hydration Water in Intrinsically Disordered Proteins. *J. Phys. Chem. B* 119, 10858–10867.
- (26) Park, H. (1997) Creation and characterization of asymmetric mutations in R67 dihydrofolate reductases. Ph.D. Dissertation, University of Tennessee, Knoxville, TN.
- (27) Stanley, C., Krueger, S., Parsegian, V. A., and Rau, D. C. (2008) Protein structure and hydration probed by SANS and osmotic stress. *Biophys. J.* 94, 2777–2789.
- (28) Arnold, O., Bilheux, J. C., Borreguero, J. M., Buts, A., Campbell, S. I., Chapon, L., Doucet, M., Draper, N., Ferraz Leal, R., Gigg, M. A., Lynch, V. E., Markvardsen, A., Mikkelsen, D. J., Mikkelsen, R. L., Miller, R., Palmen, K., Parker, P., Passos, G., Perring, T. G., Peterson, P. F., Ren, S., Reuter, M. A., Savici, A. T., Taylor, J. W., Taylor, R. J., Tolchenov, R., Zhou, W., and Zikovskiy, J. (2014) Mantid-Data analysis and visualization package for neutron scattering and mu SR experiments. *Nucl. Instrum. Methods Phys. Res., Sect. A* 764, 156–166.
- (29) Svergun, D. I. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* 25, 495–503.
- (30) Whitten, A. E., Cai, S., and Trehwella, J. (2008) MULCh: modules for the analysis of small-angle neutron contrast variation data from biomolecular assemblies. *J. Appl. Crystallogr.* 41, 222–226.
- (31) Curtis, J. E., Raghunandan, S., Nanda, H., and Krueger, S. (2012) SASSIE: A program to study intrinsically disordered biological molecules and macromolecular ensembles using experimental scattering restraints. *Comput. Phys. Commun.* 183, 382–389.
- (32) Webb, B., and Sali, A. (2014) Comparative Protein Structure Modeling Using MODELLER. *Current Protocols in Bioinformatics* 47, 5.6.1–5.6.32.
- (33) Berendsen, H. J. C., Grigera, J. R., and Straatsma, T. P. (1987) The missing term in effective pair potentials. *J. Phys. Chem.* 91, 6269–6271.
- (34) Ramanathan, A., Savol, A. J., Langmead, C. J., Agarwal, P. K., and Chennubhotla, C. S. (2011) Discovering conformational sub-states relevant to protein function. *PLoS One* 6, No. e15827.
- (35) Case, D. A., Babin, V., Berryman, J. T., Betz, R. M., Cai, Q., Cerutti, D. S., Cheatham, T. E., III, Darden, T. A., Duke, R. E., Gohlke, H., Goetz, A. W., Gusarov, S., Homeyer, N., Janowski, P., Kaus, J., Kolossváry, I., Kovalenko, A., Lee, T. S., LeGrand, S., Luchko, T., Luo, R., Madej, B., Merz, K. M., Paesani, F., Roe, D. R., Roitberg, A., Sagui, C., Salomon-Ferrer, R., Sebabra, G., Simmerling, C. L., Smith, W., Walker, R. C., Wang, J., Wolf, R. M., Wu, X., and Kollman, P. A. (2014) *AMBER14*, University of California, San Francisco.
- (36) Bradrick, T. D., Beechem, J. M., and Howell, E. E. (1996) Unusual binding stoichiometries and cooperativity are observed during binary and ternary complex formation in the single active pore of R67 dihydrofolate reductase, a D₂ symmetric protein. *Biochemistry* 35, 11414–11424.
- (37) Lin, L. N., Brandts, J. F., Brandts, J. M., and Plotnikov, V. (2002) Determination of the volumetric properties of proteins and other solutes using pressure perturbation calorimetry. *Anal. Biochem.* 302, 144–160.
- (38) Mitra, L., Smolin, N., Ravindra, R., Royer, C., and Winter, R. (2006) Pressure perturbation calorimetric studies of the solvation properties and the thermal unfolding of proteins in solution—experiments and theoretical interpretation. *Phys. Chem. Chem. Phys.* 8, 1249–1265.
- (39) Keller, S., Vargas, C., Zhao, H., Piszczek, G., Brautigam, C. A., and Schuck, P. (2012) High-precision isothermal titration calorimetry with automated peak-shape analysis. *Anal. Chem.* 84, 5066–5073.
- (40) Timson, M. J., Duff, M. R., Jr., Dickey, G., Saxton, A. M., Reyes-De-Corcuera, J. L., and Howell, E. E. (2013) Further studies on the role of water in R67 dihydrofolate reductase. *Biochemistry* 52, 2118–2127.
- (41) Receveur-Brechot, V., and Durand, D. (2012) How random are intrinsically disordered proteins? A small angle scattering perspective. *Curr. Protein Pept. Sci.* 13, 55–75.
- (42) Svergun, D. I., Richard, S., Koch, M. H., Sayers, Z., Kuprin, S., and Zaccai, G. (1998) Protein hydration in solution: experimental observation by x-ray and neutron scattering. *Proc. Natl. Acad. Sci. U. S. A.* 95, 2267–2272.
- (43) Rambo, R. P., and Tainer, J. A. (2013) Accurate assessment of mass, models and resolution by small-angle scattering. *Nature* 496, 477–481.
- (44) Santner, A. A., Croy, C. H., Vasawala, F. H., Uversky, V. N., Van, Y. Y., and Dunker, A. K. (2012) Sweeping away protein aggregation with entropic bristles: intrinsically disordered protein fusions enhance soluble expression. *Biochemistry* 51, 7250–7262.
- (45) Fried, M. G., Stickle, D. F., Smirnakis, K. V., Adams, C., MacDonald, D., and Lu, P. (2002) Role of hydration in the binding of lac repressor to DNA. *J. Biol. Chem.* 277, 50676–50682.
- (46) Tsodikov, O. V., Record, M. T., Jr., and Sergeev, Y. V. (2002) Novel computer program for fast exact calculation of accessible and molecular surface areas and average surface curvature. *J. Comput. Chem.* 23, 600–609.
- (47) Zhuang, P., Eisenstein, E., and Howell, E. E. (1994) Equilibrium folding studies of tetrameric R67 dihydrofolate reductase. *Biochemistry* 33, 4237–4244.
- (48) Permyakov, S. (2012) Differential Scanning Microcalorimetry of Intrinsically Disordered Proteins. In *Intrinsically Disordered Protein Analysis, Vol. 2, Methods and Experimental Tools* (Uversky, V. L., and Dunker, A. K., Eds.) pp 283–296, Springer, Dordrecht, The Netherlands.
- (49) Permyakov, S. E., Bakunts, A. G., Denesyuk, A. I., Knyazeva, E. L., Uversky, V. N., and Permyakov, E. A. (2008) Apo-parvalbumin as an intrinsically disordered protein. *Proteins: Struct., Funct., Genet.* 72, 822–836.
- (50) Parsegian, V. A., Rand, R. P., and Rau, D. C. (1995) Macromolecules and water: probing with osmotic stress. *Methods Enzymol.* 259, 43–94.
- (51) Royer, C. A. (2002) Revisiting volume changes in pressure-induced protein unfolding. *Biochim. Biophys. Acta, Protein Struct. Mol. Enzymol.* 1595, 201–209.
- (52) Zhai, Y., Okoro, L., Cooper, A., and Winter, R. (2011) Applications of pressure perturbation calorimetry in biophysical studies. *Biophys. Chem.* 156, 13–23.
- (53) Ravindra, R., Royer, C., and Winter, R. (2004) Pressure perturbation calorimetric studies of the solvation properties and the thermal unfolding of Staphylococcal nuclease. *Phys. Chem. Chem. Phys.* 6, 1952–1961.
- (54) Eisenberg, D., Weiss, R. M., Terwilliger, T. C., and Wilcox, W. (1982) Hydrophobic moments and protein structure. *Faraday Symp. Chem. Soc.* 17, 109–120.
- (55) Wang, A., and Bolen, D. W. (1997) A naturally occurring protective system in urea-rich cells: mechanism of osmolyte protection of proteins against urea denaturation. *Biochemistry* 36, 9101–9108.
- (56) Bhojane, P. P., Duff, M. R., Jr., Bafna, K., Rimmer, G. P., Agarwal, P. K., and Howell, E. E. (2016) Aspects of Weak Interactions between Folate and Glycine Betaine. *Biochemistry* 55, 6282–6294.
- (57) Capp, M. W., Pegram, L. M., Saecker, R. M., Kratz, M., Riccardi, D., Wendorff, T., Cannon, J. G., and Record, M. T., Jr (2009) Interactions of the osmolyte glycine betaine with molecular surfaces in water: thermodynamics, structural interpretation, and prediction of m-values. *Biochemistry* 48, 10372–10379.
- (58) Torres, A. M., Grieve, S. M., Chapman, B. E., and Kuchel, P. W. (1997) Strong and weak binding of water to proteins studied by NMR

triple-quantum filtered relaxation spectroscopy of (17)O-water. *Biophys. Chem.* 67, 187–198.

(59) Yang, P. H., and Rupley, J. A. (1979) Protein–water interactions. Heat capacity of the lysozyme–water system. *Biochemistry* 18, 2654–2661.

(60) Pethig, R. (1992) Protein-water interactions determined by dielectric methods. *Annu. Rev. Phys. Chem.* 43, 177–205.

(61) Kodandapani, R., Suresh, C. G., and Vijayan, M. (1990) Crystal structure of low humidity tetragonal lysozyme at 2.1-Å resolution. Variability in hydration shell and its structural consequences. *J. Biol. Chem.* 265, 16126–16131.

(62) Shimizu, S., and Matubayasi, N. (2014) Preferential solvation: dividing surface vs excess numbers. *J. Phys. Chem. B* 118, 3922–3930.

(63) Zhou, H. X., Rivas, G. N., and Minton, A. P. (2008) Macromolecular crowding and confinement: Biochemical, biophysical, and potential physiological consequences. *Annu. Rev. Biophys.* 37, 375–397.

(64) Courtenay, E. S., Capp, M. W., Anderson, C. F., and Record, M. T., Jr (2000) Vapor pressure osmometry studies of osmolyte-protein interactions: implications for the action of osmoprotectants in vivo and for the interpretation of “osmotic stress” experiments in vitro. *Biochemistry* 39, 4455–4471.

(65) Santoro, M. M., Liu, Y., Khan, S. M., Hou, L. X., and Bolen, D. W. (1992) Increased thermal stability of proteins in the presence of naturally occurring osmolytes. *Biochemistry* 31, 5278–5283.

(66) Tantos, A., Szmka, K., Szabo, B., Bokor, M., Kamasa, P., Matus, P., Bekesi, A., Tompa, K., Han, K. H., and Tompa, P. (2013) Structural disorder and local order of hNopp140. *Biochim. Biophys. Acta, Proteins Proteomics* 1834, 342–350.

(67) van der Lee, R., Buljan, M., Lang, B., Weatheritt, R. J., Daughdrill, G. W., Dunker, A. K., Fuxreiter, M., Gough, J., Gsponer, J., Jones, D. T., Kim, P. M., Kriwacki, R. W., Oldfield, C. J., Pappu, R. V., Tompa, P., Uversky, V. N., Wright, P. E., and Babu, M. M. (2014) Classification of intrinsically disordered regions and proteins. *Chem. Rev.* 114, 6589–6631.

(68) Uversky, V. N. (2013) The most important thing is the tail: multitudinous functionalities of intrinsically disordered protein termini. *FEBS Lett.* 587, 1891–1901.

(69) Uversky, V. N. (2009) Intrinsically disordered proteins and their environment: effects of strong denaturants, temperature, pH, counter ions, membranes, binding partners, osmolytes, and macromolecular crowding. *Protein J.* 28, 305–325.

(70) Abeln, S., and Frenkel, D. (2008) Disordered flanks prevent peptide aggregation. *PLoS Comput. Biol.* 4, No. e1000241.

(71) Grana-Montes, R., Marinelli, P., Reverter, D., and Ventura, S. (2014) N-terminal protein tails act as aggregation protective entropic bristles: the SUMO case. *Biomacromolecules* 15, 1194–1203.

(72) Strader, M. B. (1998) Constructing a hybrid of R67 dihydrofolate reductase to study asymmetric mutations in the active site. M.S. Thesis, University of Tennessee, Knoxville, TN.

(73) Martinez, M. A., Pezo, V., Marliere, P., and Wain-Hobson, S. (1996) Exploring the functional robustness of an enzyme by in vitro evolution. *EMBO J.* 15, 1203–1210.