Complexes of Native Ubiquitin and Dodecyl Sulfate Illustrate the Nature of Hydrophobic and Electrostatic Interactions in the Binding of Proteins and Surfactants

Native ubiquitin-poly-dodecyl sulfate complexes define cooperative hydrophobic and electrostatic interactions in protein-surfactant binding.

Bryan F. Shaw[1,2‡], Grégory F. Schneider[1,3‡], Haribabu Arthanari[4‡], Max Narovlyansky[1], Demetri Moustakas[4], Armando Durazo[5], Gerhard Wagner[2], and George M. Whitesides[1]*

[1]*Department of Chemistry and Chemical Biology, Harvard University;*

[2]*Chemistry and Biochemistry, Baylor University*

[3]*Kavli Institute of Nanoscience, Delft University of Technology, The Netherlands*

[4]*Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School;*

[5]*Department of Chemical and Environmental Engineering, University of Arizona*

[6]*Department of Infection, Computational Sciences, AstraZeneca Pharmaceuticals LP*

[‡]These authors contributed equally to conducting this research and preparing this manuscript.

*Corresponding Author

Telephone Number: (617) 495-9340

Fax Number: (617) 495-9857

E-mail

Address: gwhitesides@gmwgroup.harvard.edu

Abstract: A previous study, using capillary electrophoresis (CE) [*J. Am. Chem. Soc.* **2008**, *130*, 17384-17393], reported that six discrete complexes of ubiquitin (UBI) and sodium dodecyl sulfate (SDS) form at different concentrations of SDS along the pathway to unfolding of UBI in solutions of SDS. One complex (that formed at concentrations between 0.8 and 1.8 mM SDS) consisted of UBI having *native* structure, but with approximately 11 molecules of SDS associated in some fashion. The current study used CE and $^{15}N/^{13}C$-^{1}H heteronuclear single quantum coherence (HSQC) NMR spectroscopy to identify residues that associate with SDS at 0.8-1.8 mM SDS. The ability of the surface charge and hydrophobicity of folded UBI to affect the association with SDS (below the CMC) was also studied, using CE, by converting lys-ε-NH_3^+ to lys-ε-$NHCOCH_3$ groups. Capillary electrophoresis demonstrated that the acetylation of lysine residues inhibited the binding of 11 SDS (< 2 mM SDS) and decreased the number of identifiable complexes of composition UBI-$(NHAc)_8$•SDS_n that form on the pathway of unfolding of UBI-$(NHAc)_8$ in SDS. A comparison of ^{15}N-^{1}H HSQC data at 0 mM and 1 mM SDS with calculated electrostatic surface potentials of folded UBI (e.g., solutions to the non-linear Poisson-Boltzmann equation) suggested that SDS binds at residues in UBI (at 1 mM SDS) that are both hydrophobic and contained within regions with positive electrostatic surface potential. With regard to the first intermediate that is stable during CE, HSQC demonstrated that SDS interacted with surface-exposed leucine and isoleucine residues, and not cationic residues, suggesting that the accommodation of a large fraction of *n*-alkyl tail is essential for binding. Amino acid residues in UBI did not interact with SDS, at 1 mM SDS, according to their 2° structure: residues in the four β-strands interacted with SDS as strongly as residues in the single α-helix although residues in hydrogen-bonded loops did *not* bind SDS. No correlation was observed between the association of an amino acid with SDS and the solvent accessibility of the residue in folded UBI, or the rate of exchange of its amide hydrogen. The majority of resonances of aliphatic protons were attenuated by dynamic broadening resulting from interactions with *n*-alkyl tails of SDS. Among resonances from protons with intrinsically low rates of relaxation a subset exhibited chemical shifts indicative of specific interactions and correlated with greater chemical shifts of amide resonances. This study establishes a few (of perhaps several) factors that control the simultaneous molecular recognition of multiple anionic amphiphiles by a *folded* cytosolic protein.

Introduction

This paper describes a study of the association of SDS with two small proteins – ubiquitin (UBI; also abbreviated (UBI-$(NH_3^+)_8$), and its peracetylated derivative (UBI-$(NHAc)_8$), in which all lysine groups and the amino terminus are acetylated. This acetylation neutralizes ~70% of the positively charged residues of UBI. The work has three objectives: i) to determine if surface-accessible cationic residues on UBI are involved in specific electrostatic association with SDS, ii) to clarify the relative importance of electrostatic and hydrophobic interactions in the interaction of an anionic surfactant with proteins, and iii) to determine how the structural environment of residues in folded ubiquitin (i.e., secondary structure and solvent accessibility) correlates with the interaction of those residues with SDS (at concentrations of SDS that do not unfold UBI).

The association between a protein and a charged surfactant involves both hydrophobic and electrostatic interactions, but the relative importance of these interactions has not been defined. We believe that correlating the pathway of these associations (i.e., the number of surfactants that bind to the protein, as a function of the concentration of the surfactants) with the biophysical and physical-organic properties of proteins (for example, net charge, electrostatic surface potential, hydrophobicity, secondary and tertiary structure, conformational flexibility, and fold family, to name a few) will help in understanding the interaction between surfactants and proteins, and the unfolding or aggregation of proteins that often results from this association. (1, 2) The interaction of lipids and proteins has, for example, been hypothesized to be a crucial step in promoting (or preventing) the aggregation of proteins and the pathogenesis of protein aggregation diseases. (1-3)

Ubiquitin binds up to eleven SDS molecules without unfolding. Using CE, we have previously demonstrated that UBI unfolds in SDS *via* an unexpectedly complicated pathway that

involves the stepwise binding of discrete numbers of molecules of SDS and, therefore, the formation of distinct complexes along the unfolding pathway: UBI-$SDS_{\sim 11}$, UBI-$SDS_{\sim 25}$, and UBI-$SDS_{\sim 33}$. We sometimes refer to these complexes as *intermediates*, although some involve completely folded UBI (i.e., UBI-$SDS_{\sim 11}$), and others completely unfolded UBI (i.e., UBI-$SDS_{\sim 33}$).

The complexes of UBI that form at low concentrations of SDS (< 2 mM, i.e., UBI-SDS_{11}) are remarkable for their conservation of native structure (as measured by circular dichroism and by rates of hydrogen-deuterium exchange (4)). The complexes that form at higher concentrations of SDS (> 2 mM; e.g., UBI-$SDS_{\sim 25}$, UBI-$SDS_{\sim 33}$, UBI-SDS_{42}) show more α-helical structure than the native protein (as measured by circular dichroism), but retain very little tertiary structure, as demonstrated by the rapid rate of hydrogen-deuterium exchange. (4) Above the critical micellar concentration (CMC), which for SDS is 3.4 mM (in 25 mM Tris, 192 mM glycine, pH 8.4, 25 °C) (4), UBI is 'unfolded' and saturated with approximately 42 molecules of SDS. This stoichiometry (UBI-SDS_{42}) translates to approximately one SDS for every two amino acids.

Most of the proteins that we have studied with CE either do not form complexes of intermediate stoichiometry with SDS (e.g., most cannot bind SDS and remain natively *folded*), or do not form as many different complexes as does UBI. (5) We have, so far, been unable to predict which proteins interact with SDS in ways that form stable aggregates of protein and surfactant, and that retain much or all of the tertiary structure of the protein. Nevertheless, because UBI is folded in *some* of the UBI-SDS complexes that form (i.e., UBI-SDS_{11}) and because multiple interactions (as many as eleven) between the *folded* protein and surfactant can be studied simultaneously, the UBI-SDS system could be particularly useful for acquiring general information on the interactions of folded proteins with surfactants (and perhaps with other small

molecules in general).

Native ubiquitin-poly-dodecyl sulfate complexes represent a useful system in the study of molecular recognition. The binding of SDS to folded UBI below the CMC does *not* involve the scenario encountered in many other protein-ligand systems—that is, the binding of a ligand *inside* a pocket. The binding of SDS to native UBI at 1 mM SDS, for example, appears to occur on the *surface* of the protein.(6) It must be remembered, however, that the binding of SDS to the surface of UBI can be both strong and specific below the CMC: the complexes that form at ~ 1 mM SDS (e.g., UBI-SDS_{11}) are equilibrium complexes. These complexes are characterized during CE by a single peak that is unimodal (and as sharp as the peak representing UBI-SDS_0), which suggests that the interconversion of UBI-SDS_{11} between, for example, UBI-SDS_{10} or UBI-SDS_{12}, is slower than the time scale of a CE experiment. Native ubiquitin-poly-surfactant complexes (prepared with SDS or with other surfactants) could, therefore, become a useful system for studying protein-ligand interactions, and a system that could yield new insights into the physical-organic chemistry that underlies both the hydrophobic effect, and biomolecular recognition.

In this paper we have investigated, in particular, the importance of surface-accessible, cationic, lysine ε-NH_3^+ residues in the pathway of association of UBI with SDS, and we have identified residues in folded ubiquitin that bind SDS at concentrations that are below the CMC. The two analytical tools used in the present study are capillary electrophoresis (CE) and heteronuclear single quantum coherence (HSQC) NMR spectroscopy. Capillary electrophoresis provides information about the stoichiometry of protein-SDS interactions: the electrophoretic mobility of low- to medium-molecular-weight proteins (< 60 kDa) in a bare capillary is sensitive to the binding of a single surfactant molecule. (4) Because the native fold of UBI is retained in

the first UBI-SDS complex that forms (at ~ 1mM SDS), we can identify residues that interact with SDS below the CMC by shifts of amide N-H peaks in ^{15}N-^{1}H HSQC and C-H peaks in ^{13}C-^{1}H HSQC.

Our results show that the binding of SDS to folded UBI (and the pathway of unfolding of UBI) can be changed by altering the electrostatic surface potential and hydrophobicity of UBI. The results of our analysis of complexes of UBI and SDS with CE and HSQC also suggests, unexpectedly, that the negatively charged sulfate group of SDS does not, generally, associate selectively with the positively charged ammonium group of lysine (or the guanidinium group of arginine) below the CMC. The binding of SDS by UBI (below the CMC) occurs, instead, at hydrophobic patches of the protein that have positive electrostatic surface potential.

Experimental design

Selection of ubiquitin. We used ubiquitin (UBI) because it is a heat-stable, small protein (76 amino acids, 8565 Da), which has both α-helix and β-sheet elements,(7) and which forms a variety of complexes with SDS that can be distinguished by CE. (4) UBI is a tractable and widely used model protein for experimental (and sometimes theoretical) studies of protein folding.(7) UBI contains seven lysines and one N-terminal methionine α-NH_3^+ group (all of these amino groups can be acetylated by acetic anhydride); one C-terminal α-COO^- group; four arginines; one histidine; five aspartates and six glutamates; nine leucines; seven isoleucines; and three prolines. The theoretical isoelectric point of ubiquitin is 6.56. The theoretical value of the net charge of UBI (Z_{SEQ}) at pH 8.4 was calculated, from its sequence, to be $Z_{SEQ} = 0$ (assuming that Z = +1 for the N-terminus; Z = -1 for the C-terminus; Z = -1 for Asp and Glu; Z = +1 for Arg and Lys). The value of net charge of UBI that we estimated experimentally with CE (Z_{CE}) at pH 8.4 (25 mM

Tris, 192 mM glycine) was Z_{CE} = -0.2 (in our previous study (4)). The supporting information includes a discussion of the possible causes of this small difference between Z_{CE} and Z_{SEQ}.

Studying the binding of SDS to peracylated UBI with CE: are cationic residues on the surface of UBI required for surfactant binding? Our previous study (4) concluded that the first stable complex of UBI with SDS (a complex that forms at [SDS] = 1.0 mM), had eleven molecules of SDS bound to each molecule of UBI. Since this number roughly corresponds to the number of positively charged amino acids in UBI at pH 8.4, we hypothesized that cationic sites might nucleate the condensation of SDS molecules on the protein at [SDS] < 1.0 mM. Here we tested this hypothesis by neutralizing all NH_3^+ moieties into $NHCOCH_3$ by acetylation with acetic anhydride, and comparing the pathways for the association of SDS with unacetylated UBI (referred to as "UBI-$(NH_3^+)_8$") and the peracetylated derivative (e.g., "UBI-$(NHAc)_8$"). We have, so far, neglected possible changes in the structure of the UBI as a result of acetylation, based on the previous work of Makhatadze et al. which demonstrated (by site-directed mutagenesis or carbamylation) that the neutralization of positively charged residues on UBI left the protein properly folded. (8) Peracetylated UBI was prepared as previously described for the acetylation of other proteins that we have studied (i.e., α-amylase, carbonic anhydrase II in detail (9, 10), and many others in survey (11)).

Using Heteronuclear Single Quantum Coherence (HSQC) NMR spectroscopy to identify electrostatic or hydrophobic interactions between specific residues in folded ubiquitin and SDS. Because a specific UBI-SDS complex is predominant at 0.8-1.2 mM SDS, and because the UBI proteins with up to 11 bound SDS have the same structure as UBI•SDS_0 (4), we can use ^{15}N-^1H HSQC NMR spectroscopy to identify which residues in UBI associate with SDS at 0.8-1.2 mM.

The binding of a ligand (such as a surfactant) to residues in a protein can be conveniently detected by a shift in the amide ^{15}N-^{1}H resonance of the interacting residue along the ^{15}N or ^{1}H axes (12) (e.g., a *chemical shift perturbation*, denoted CSP or $\Delta\delta$), and in some cases, by a reduction of the intensity of the ^{15}N-^{1}H resonance).(1) Chemical shift perturbations have been widely used to characterize the interaction between proteins and charged ligands including: fatty acids (i.e., palmitic acid (13)), anionic detergents (14, 15), nucleic acids (16), oligonucleotides (17), lipids (1), carbohydrates (18), and small drug-like molecules (19). The specific residues whose ^{15}N-^{1}H resonance shift upon titration with a ligand in solution are often the same residues that are binding directly to the ligand according to X-ray crystallography—so long as no change in the conformation of the protein occurs (in solution) upon the binding of ligand.

We expect that electrostatic interaction between the $ROSO_3^-$ head group of SDS, and any positively charged functional group in UBI (i.e., ε-NH_3^+ or δ-$NH(C{=}NH_2^+)NH_2$), might be detected by movement in ^{15}N-^{1}H resonance of the interacting residue along the ^{15}N or ^{1}H axes ($\Delta\delta$). We base these predictions on previously published reports of chemical shift perturbations of residues (in structurally well-characterized proteins such as ubiquitin and nuclease) that engage in electrostatic interactions with negatively charged ligands (and micellar surfaces) that bear phosphate, carboxylic, and sulfate groups. (14, 19, 20) The application of NMR spectroscopy to study the interaction of folded proteins with surfactants in detergent micelles (e.g., membrane proteins) has been reviewed by Wüthrich.(21)

Approximating the electrostatic surface potentials of residues in native ubiquitin that are accessible to solvent or ligand. In order to determine if the sulfate group of SDS interacts with residues in UBI as a function of the local electrostatic environment of the residue (as opposed to, or in addition to the *formal* charge of the residue), and in order to determine how

the positive charge on some residues might be screened or dampened by solvation—and therefore not interact with R-OSO_3^-—we approximated the electrostatic surface potential of residues in folded UBI and mapped these potentials onto the solvent-accessible surface of ubiquitin (based upon the X-ray crystal structure; PDB:1UBI). The electrostatic potential of residues at the solvent-accessible surface of UBI was approximated by solving the non-linear Poisson-Boltzmann (PB) equation (22), using an implicit solvent model (23) (e.g., treating solvent as a continuum with a uniform dielectric). A more detailed description of the Poisson-Boltzmann implicit solvent model (e.g., the assumptions and weaknesses of the model) can be found in supporting information. One utility of numerical solutions to the PB equation is that they provide electrostatic information about the surface of a protein that is not obvious from an examination of the location of positively and negatively charged residues in the three dimensional structure.(24) The validity of these models for use in rationalizing the molecular recognition of charged ligands and proteins is largely empirical: proteins that bind DNA, for example, generally have patches of positive electrostatic surface potential (according to numerical solutions of PB equations) located at the site of binding for DNA (25, 26); phospholipases have patterns of electrostatic surface potential (at the site of lipid binding) that are thought to facilitate the binding of negatively charged lipids (27); many other examples of complementary electrostatic surface potentials between macromolecules that interact with one another, or with ligands, has been reviewed.(28)

Results and Discussion.

Neutralizing Lys-ε-NH_3^+ and α-NH_3^+ of UBI with acetylation prevents the formation of folded ubiquitin-poly-dodecyl sulfate complexes. The electropherograms displayed in Figure 1 show the binding of different numbers of SDS molecules to UBI-$(NH_3^+)_8$ and UBI-$(NHAc)_8$ at

different concentrations of SDS. For example, at 0 mM SDS, the electrophoretic mobility of UBI-$(NH_3^+)_8$ is similar to the neutral marker (dimethylformamide, DMF) because the net charge of UBI-$(NH_3^+)_8$ (denoted Z_0) is close to zero (at pH 8.4, $Z_0 = -0.2$). The peak for UBI-$(NH_3^+)_8$ shifts to a higher mobility as the concentration of SDS is increased from 0.0 to 0.2 and 0.4 mM because the binding of SDS increases the net negative charge of UBI-$(NH_3^+)_8$; the neutral marker peak remains at $\mu = 0$ cm^2 kV^{-1} min^{-1}.

The progressive binding of SDS to UBI-$(NH_3^+)_8$ is illustrated in the capillary electropherograms by a transition from the native protein (denoted N) to discrete complexes that we describe as groups "G" depending of their mobility. The first *discrete* and well resolved complex (denoted G_2) forms via a set of complexes (denoted G_1*; Figure 1a). Although G_2 represents the first *discrete* complex with a defined stoichiometry, G_1* represents, in fact, the *first* complex *per se*—actually a set of complexes. We estimate the net charge of the complex UBI-$(NH_3^+)_8$•$SDS_{\sim 11}$ to be –7.4 (assuming that each SDS molecule increases the negative charge by 0.9 units(29)). The supporting material contains more details about the progression from N to G_2 (see Figure S1).

The electropherograms of UBI-$(NHAc)_8$ in different concentrations of SDS illustrates how SDS binds to UBI-$(NHAc)_8$ (Figure 1b). The mobility of UBI-$(NHAc)_8$ at 0 mM SDS (μ = 15 cm^2 kV^{-1} min^{-1}) is significantly larger than for UBI-$(NH_3^+)_8$ at 0 mM SDS because UBI-$(NHAc)_8$ (Z = -7.4) is much more negatively charged than UBI-$(NH_3^+)_8$ (Z=-0.2).

The binding of SDS to UBI-$(NHAc)_8$ is quite different from that of UBI-$(NH_3^+)_8$. For example, UBI-$(NHAc)_8$ migrates during CE—at concentrations up to 1.8 mM SDS —as a single peak that has a mobility indistinguishable from UBI-$(NHAc)_8$ at 0 mM SDS (N', $\mu \sim 15$ cm^2 kV^{-1} min^{-1}, Figure 1b). Therefore, SDS does not appear to be binding to UBI-$(NHAc)_8$ at

concentrations of SDS below 1.8 mM. The inhibition of the recognition of SDS by UBI (at <1.8 mM SDS) by the neutralization of its lysine-ε-NH_3^+ groups (and N-terminus) is remarkable and demonstrates that electrostatic interactions are involved, somehow, in the binding of SDS with UBI—and that hydrophobic interactions are *not* so dominant that they can completely explain the association of SDS with UBI in the absence of electrostatic interactions.

The different capacities of UBI-$(NHAc)_8$ and UBI-$(NH_3^+)_{8+}$, to bind SDS (at < 1.8 mM SDS) caused us to hypothesize that the R-OSO^{3-} group of SDS formed specific electrostatic interactions with positively charged groups on the surface of UBI-$(NH_3^+)_8$. The neutralization of 8 ammonium groups (7 lysines and the N-terminus) with acetic anhydride, however, prevented the binding of 11 molecules of SDS at < 1.8 mM SDS, and not 8 molecules of SDS, as might be expected if SDS molecules were associating selectively and specifically with Lys-ε-NH_3^+. The four guanidinium groups of UBI-$(NHAc)_8$ remained positively charged in UBI-$(NHAc)_8$, after peracetylation (presumably), but these Arg residues did not bind SDS below a concentration of 1.8 mM. Therefore, there is *not* a 1:1 relationship between the number of positively charged functional groups in UBI-$(NHAc)_8$ (e.g., four guanidinium groups) and the number of SDS bound at 0-1.8 mM SDS (e.g., zero SDS). Moreover, there is not a 1:1 relationship between the number of *amino* groups in UBI-$(NH_3^+)_8$ and the maximum number of SDS (11) that bind at < 1.8 mM. Thus, we conclude that the electrostatic interactions between UBI-$(NH_3^+)_8$ and 11 molecules of SDS (at < 1.8 mM) are not occurring, exclusively, between the sulfate group of SDS and positively charged amino and guanidinium groups of the protein (although we cannot rule out the possibility that *some* lysine-ε-NH_3^+ groups are interacting with the SDS). We conclude, instead, that peracetylation inhibited the binding of SDS (at < 1.8 mM) because it increased the net negative charge of UBI, or decreased the *local* positive electrostatic surface potential of regions

on the surface of UBI that *were* binding SDS (possibly Lys residues or residues nearest the 7 lysines or N-terminus).

Association of UBI-$(NHAc)_8$ with SDS is reversible. The formation of protein-protein aggregates (transient or irreversible) is a problem that complicates the study of interactions between proteins and chaotropic agents (e.g., *protein folding*). Protein aggregation is hypothesized to be the cause of numerous artefacts observed during studies of protein folding.(30) The aggregation of UBI, for example, has been hypothesized to cause its *apparent* three-state folding pathway, observed in guanidinium hydrochloride (a two-state pathway is observed in other studies).(31) We have previously demonstrated that the formation of complexes of SDS and UBI-$(NH_3^+)_8$ is reversible in tris-glycine buffer (at room temperature), and that very little protein is lost to aggregation during dialysis in different concentrations of SDS.(4) In order to prove that the association of SDS with UBI-$(NHAc)_8$ is also reversible, we prepared complexes of SDS and UBI-$(NHAc)_8$ from UBI-$(NHAc)_8$ that was initially *unfolded,* and also from UBI-$(NHAc)_8$ that was initially *folded,* and analyzed each with CE. Unfolded UBI-$(NHAc)_8$ was prepared by dialyzing solutions of UBI-$(NHAc)_8$ against SDS (up to 10 mM), as previously described for UBI-$(NH_3^+)_8$(4). The two sets of electropherograms from these experiments are shown in Figure S2 (see supporting information), and their similarity demonstrates that the unfolding of UBI-$(NHAc)_8$ is reversible in SDS (pH 8.4).

In order to determine whether the peracetylation of UBI significantly perturbed its structure or conformational stability, we also characterized UBI-$(NHAc)_8$ (in the absence of SDS) with circular dichroism (Figure S3, see supplemental information), differential scanning calorimetry, and amide hydrogen-deuterium exchange (Figure S4, see supplemental information). The similarity in the circular dichroic spectra of UBI-$(NH_3^+)_8$ and UBI-$(NHAc)_8$ indicated that the

over-all structures of UBI-$(NH_3^+)_8$ and UBI-$(NHAc)_8$ were similar. Peracetylation did, however, increase the rate of amide hydrogen-deuterium exchange, and did lower the thermostability of the protein (T_m= 67 °C for UBI-$(NHAc)_8$; T_m > 90 °C for UBI-$(NH_3^+)_8$), however, both differential scanning calorimetry and hydrogen-deuterium exchange indicated that the peracetylated protein remained folded.

UBI-$(NH_3^+)_8$·$(SDS)_{\sim 11}$ unfolds at a lower concentration of SDS than UBI-$(NHAc)_8$. UBI-$(NHAc)_8$ begins to form a complex denoted G_4 at a higher concentration of SDS (2.5 mM) than UBI-$(NH_3^+)_8$, which begins to forms G_4 complex at concentrations as low as 1.4 mM SDS, although as a mixture with G_2 and G_3 intermediates, Figure 1. The G_4 complex of both UBI-$(NH_3^+)_8$·$(SDS)_{\sim 11}$ and UBI-$(NHAc)_8$ exists up to 3.5-4.0 mM SDS, and is the only thermodynamically complex stable at 3.0 mM SDS. Our previous study(4) of UBI-$(NH_3^+)_8$ showed that the UBI-SDS complex G_4 is almost completely unfolded based on the rate of H/D exchange at 3.0 mM SDS. The increase of α-helical content, measured by circular dichroism (CD) at 220 nm in both G_4 complexes, suggests they exist as mixed micelles with cylindrical shape. The lowered dielectric permittivity of the micelle surrounding the unfolded polypeptide increases the stability of H-bonds and promotes the formation of local folds (Figure S5, see supplemental information). The rearrangement of the native β-sheet in UBI into α-helix in mixed micelles is a slow and irreversible process, that yields complex G_4 upon extensive dialysis (170 hours) of ubiquitin with SDS. Electrophoresis of unmodified UBI-$(NH_3^+)_8$ in the presence of SDS forms some G_4, but only as a mixture of complexes at higher concentration of SDS (> 5.0 mM) during a 10 minute reaction in the capillary, Figure 2.

One explanation for the observation that UBI-$(NH_3^+)_8$ forms the G_4 complex at a lower concentration of SDS than UBI-$(NHAc)_8$ (by ~ 1 mM) is that UBI-$(NH_3^+)_8$•$(SDS)_{11}$ is more

hydrophobic than UBI-$(NHAc)_8$•SDS_0 (because of the associated $C_{12}H_{25}$ chains) and UBI-$(NH_3^+)_8$•$(SDS)_{11}$ might, therefore, interact with additional molecules of SDS more favorably (thermodynamically) than UBI-$(NHAc)_8$•SDS_0. (32, 33, 34) Regardless of the explanation of this observation, G_4 intermediates, UBI-$(NH_3^+)_8$•SDS_{11} and UBI-$(NHAc)_8$•SDS_0, exist exclusively above 2.5 mM SDS (< 3.5 mM) although they differ in charge by 4 units. We conclude that the presence of ε-NH_3^+ groups is important for the formation of the first complex of UBI and SDS (e.g., UBI-SDS_{11}), but that ε-NH_3^+ groups are *not* playing a major role thermodynamically during the intermediate stages of binding to SDS, and hydrophobic effects dominate electrostatics (although the latter are crucial in the latest stages of denaturation).

UBI-$(NH_3^+)_8$•$(SDS)_{11}$ and UBI-$(NHAc)_8$ unfold *via* the binding of indistinguishable numbers of SDS molecules. The mobilities of complexes of UBI-$(NH_3^+)_8$ and UBI-$(NHAc)_8$ with SDS are approximately equal at concentration of SDS = 3.0 mM ($\mu \sim 20$ cm^2 kV^{-1} min^{-1}) (Figure 1). The stoichiometry of the complexes at this concentration are, however, quite different: UBI-$(NH_3^+)_8$•$(SDS)_{\sim 25}$ versus UBI-$(NHAc)_8$•$(SDS)_{\sim 14}$. The formation of G_4 complexes, therefore, involves binding 14 molecules of SDS by UBI-$(NHAc)_8$ and UBI-$(NH_3^+)_8$•$(SDS)_{\sim 11}$.

The electrophoretic mobility of a complex of UBI and SDS can be approximated (Eq. 1-2) as a function of its net charge (Z), and molecular weight (M). In Eq. 2, Z_0 is the net charge of the native UBI (Z_0(UBI-$(NH_3^+)_8$) ~ -0.2),(4) p is the number of acetylations, n is the number of equivalents of SDS bound to UBI-$(NHAc)_p(NH_3^+)_{8-p}$, ΔZ is the charge increment resulting from one acetylation or from the binding of one molecule of SDS, M_{Ac} is the increase in mass due to acetylation, M_{DS}- is the molecular weight of dodecyl sulfate ions, and C_p is a protein-specific constant (35) and α has a typical value of 2/3 for globular proteins.(36) This value is consistent with the hydrodynamic drag being proportional to the surface area.

$$\mu = C_p \cdot \frac{Z}{M^{\alpha}} \qquad (1)$$

$$\mu = C_p \cdot \frac{Z_0 + (p+n)\cdot \Delta Z}{(M_{UBI} + p \cdot M_{Ac} + n \cdot M_{DS^-})^{\alpha}} \qquad (2)$$

We were able to estimate the stoichiometry of the complex making up G_4 for UBI-$(NH_3^+)_8$ in our previous study (G_4 is UBI-$(NH_3^+)_8$•$(SDS)_{\sim 25}$).(4) The mobility of UBI-$(NH_3^+)_8$•$(SDS)_{\sim 25}$ can be expressed using Eq. 1 (Z_0 = -0.2, p = 0, n = 25, ΔZ = - 0.9, M_{UBI} = 8565 Da and M_{DS^-}=265.4 Da) and can be set equal to the analogous expression for G_4 complexes of UBI-$(NHAc)_8$ (with p=8, M_{Ac}=42 Da, yielding n~14, and suggesting that fourteen SDS are bound to UBI-$(NHAc)_8$ at [SDS] = 3.0 mM. This number (n~14) is equal to the number of SDS molecules that we showed previously to bind to UBI-$(NH_3^+)_8$ during the transition $G_2 \rightarrow G_4$. Complexes of UBI-$(NHAc)_8$ with SDS in G_4, therefore, have the composition UBI-$(NHAc)_8$•$(SDS)_{\sim 14}$, while complexes of UBI-$(NH_3^+)_8$ in G_4 have composition UBI-$(NH_3^+)_8$•$SDS_{\sim 25}$. The reason why identical number of 14 SDS molecules associated with both UBI-$(NH_3^+)_8$•$(SDS)_{\sim 11}$ during the transition $G_2 \rightarrow G_4$ and with UBI-$(NHAc)_8$ during the transition $N \rightarrow G_4$ is unclear. The G_2 complex of UBI (e.g. UBI-$(NH_3^+)_8$•$(SDS)_{\sim 11}$) is much more hydrophobic than UBI-$(NHAc)_8$, because the former has more aliphatic methylene groups on the surface of the protein compared with UBI-$(NHAc)_8$, and this greater hydrophobicity of UBI-$(NH_3^+)_8$•$(SDS)_{\sim 11}$ should render it more favourable for interactions with SDS than UBI-$(NHAc)_8$. The UBI-$(NH_3^+)_8$•$(SDS)_{\sim 11}$ complex is also, however, more negatively charged than UBI-$(NHAc)_8$, and this greater negative charge could effectively cancel out any favorable hydrophobic interactions between UBI-SDS_{11} and additional molecules of SDS.

The binding of SDS to UBI-$(NHAc)_8$ and UBI-$(NH_3^+)_8$ is similar above the CMC; the shape and mobilities of the electropherograms are almost undistinguishable for UBI-$(NH_3^+)_8$ and UBI-$(NHAc)_8$ (Figure 1, [SDS] > 3.5 mM). Calculating the number of SDS molecules that are bound to UBI-$(NHAc)_8$ and UBI-$(NH_3^+)_8$ at > 3.5mM (with Equation 2) demonstrated, however, that UBI-$(NHAc)_8$ has ~11 fewer SDS molecules bound than UBI-$(NH_3^+)_8$ under denaturing conditions (> CMC).

A kinetically-stable ubiquitin-poly-dodecyl sulfate complex (e.g., UBI-$(NH_3^+)_8$•$(SDS)_7$) forms on the path from N to G_2. The CE experiments that we have done have involved complexes of UBI and SDS that are formed at equilibrium, by prolonged exposure of UBI to solutions of SDS. We are also interested, however, in studying the *kinetics* of binding of SDS to UBI. We investigated the kinetics of binding of SDS to UBI-$(NH_3^+)_8$ in the transitions $N \rightarrow G_1^* \rightarrow G_2$ by using a technique we refer to as SurfCE (e.g. Surfactant Capillary Electrophoresis). SurfCE is a technique that makes it possible to observe both the rate and stoichiometry of association of SDS with proteins, *in situ* (e.g., during capillary electrophoresis).(5) In SurfCE, the protein is injected in its native, folded form into a capillary that is filled with a solution of SDS at a selected concentration; the reaction time therefore corresponds to the time of electrophoresis (e.g., ~7 min with a 60-cm capillary). Rapid association of proteins with SDS is reflected in an immediate conversion of the peak corresponding to N into a new peak with a higher mobility. The slow association of SDS to proteins or a slow equilibration between multiple aggregates with different compositions will result in a broad peak.

Analysis of the interaction of UBI-$(NH_3^+)_8$ with SDS by SurfCE (Figure 2), revealed several additional complexes with mobilities between the peaks N and G_2 (that is $1.5 < \mu < 14$ cm^2 kV^{-1}

min^{-1}); these ubiquitin-poly-dodecyl sulfate complexes are not present in the equilibrium electropherograms for UBI-$(NH_3^+)_8$ (Figure 1). The peak at a mobility μ ~11.5 cm^2 kV^{-1} min^{-1} is present over a broad range of SDS concentrations (1.2 < [SDS] < 3.0 mM). The constant mobility of this peak suggests that a particular species with a discrete stoichiometry is formed along the path from N to G_2. Since this peak is absent in electropherograms of equilibrated mixtures (Figure 1a), we infer that this complex is kinetically labile, and forms along the pathway from N to G_2. The comparison of the mobility of this peak with the mass-calibrated mobilities of the rungs of UBI charge ladder(4) (rulers overlaid on top of the electropherograms in Figure 2), indicates that this complex has the composition UBI-$(NH_3^+)_8(SDS)_{\sim 7}$.

Monitoring the binding of SDS to residues in UBI-$(NH_3^+)_8$ with ^{15}N-^{1}H HSQC. Figure 3 summarizes the thermodynamic pathway of association of SDS with UBI-$(NH_3^+)_8$ and UBI-$(NHAc)_8$. The complex UBI-$(NH_3^+)_8(SDS)_{\sim 7}$ described in the previous section is an unstable intermediate on the pathway of formation of the first stable complex UBI-$(NH_3^+)_8(SDS)_{\sim 11}$. In order to determine which residues in UBI-$(NH_3^+)_8$ interact with SDS at concentrations below the CMC, we studied UBI-SDS complexes with ^{15}N-^{1}H HSQC NMR. We focused on interactions between SDS and UBI-$(NH_3^+)_8$ at 1.0 mM SDS because the protein retains its fold at this concentration and forms only a single complex, UBI-$(NH_3^+)_8(SDS)_{\sim 11}$ according to CE experiments.

With ^{15}N-^{1}H HSQC, we were able to monitor the association of SDS to amino acid residues of UBI-$(NH_3^+)_8$ in 1 mM SDS by measuring the movement in the ^{15}N-^{1}H correlation signal ($\Delta\delta$) of each residue (Figures 6 and 7). The value $\Delta\delta$ represents the Euclidian distance (adjusted for the characteristic frequency of precession of N and H) that each ^{15}N-^{1}H correlation signal was shifted, calculated by Equation 3:

$$\Delta\delta = \sqrt{[\Delta^{15}N/5]^2 + [\Delta^{1}H]^2} \qquad (3)$$

One important limitation of ^{15}N-^{1}H HSQC NMR as a tool for studying SDS binding to UBI at 1 mM is that an interaction between a specific residue and surfactant will not be detected if the association occurs at a rate faster than the rate of the ^{15}N-^{1}H HSQC scan, which occurs on a millisecond time-scale. For example, a portion of the dodecyl chain of an SDS molecule might be mobile (to some degree) when bound to the surface of UBI and different gauche rotamers might interact with different residues. If dynamic rotameric interactions are occurring, and a particular SDS molecule is populating a myriad rotamers, then we would only be able to detect those rotameric interactions that are populated longer than milliseconds. The majority of ^{15}N-^{1}H resonances were well-resolved, however, a similar difficulty was manifest for ^{13}C-^{1}H spectra as discussed below.

HSQC spectra of ^{15}N-^{1}H resonances of UBI-$(NH_3^+)_8$ in the presence of 1 mM SDS were still characterized by a set of well resolved, dispersed peaks, similar to the spectrum of UBI-$(NH_3^+)_8$ at 0 mM SDS, indicating that UBI-$(NH_3^+)_8(SDS)_{\sim 11}$ was folded, but the structure was modestly perturbed relative to free UBI-$(NH_3^+)_8$(Figure 4a, b). We were able to collect data on 63 of the 76 amino acids in UBI-$(NH_3^+)_8$ (including all 7 lysine residues, 2 of 4 arginines, but not the N-terminal methionine). The ^{15}N-^{1}H correlation signal for two residues (Glu24, Ala46) were too weak to be detected at 0 mM or 1 mM SDS; the chemical shift for three residues (Arg42, Gln49, and Arg72) could not be determined at 1 mM SDS because the ^{15}N-^{1}H correlation signal in 1 mM SDS disappeared entirely, or could not be assigned to a residue with confidence. In addition, the three proline residues (Pro19, Pro37, Pro38) were also not observable because proline does not have a backbone amide N-H; and ^{15}N-^{1}H correlation signals for four other residues in UBI are unassigned at 0 mM SDS and pH 7.4 (Met1, Thr9, Gly10, and Arg74).

The values of Δδ for several residues of UBI-$(NH_3^+)_8$ at 0-1 mM SDS were close to zero (e.g., $0.00 < \Delta\delta < 0.02$ ppm; Lys 6, Thr12, Glu16, Val17, Glu18, Ser20, Asp21, Thr22, Asp39, Gln40, Leu56, Tyr59, and Asn60; Figure 5); we conclude that these residues did not bind SDS, or only bound SDS transiently. The ^{15}N-^{1}H correlation signals for other residues of UBI, however, shifted considerably on going from 0 mM SDS to 1 mM SDS, with $0.15 < \Delta\delta < 0.30$ ppm, and we conclude that these residues specifically bind SDS (i.e., Val5, Leu8, Lys11, Ile13, Thr14, Leu43, Ile44, Leu50, His68, Val70, and Leu71; Figure 5)). Table 1 shows a list of the residues of UBI-$(NH_3^+)_8$ with the largest and smallest values of Δδ in 0-1mM SDS (a complete list for all 63 residues is contained in supplemental information).

A comparison of the values of Δδ with a map of the secondary structure of UBI (Figure 5) revealed that the residues contained in hydrogen bonded loops in UBI were least affected by SDS (in comparison to residues in the β-sheet or α-helix), suggesting that SDS does not interact specifically with residues in loops (e.g., residues 18-22, 39-40, and 56-59; residues 39-40 and 56-59 have also been described by some as a 3_{10} helix).

Most of the residues in UBI-$(NH_3^+)_8$ with the largest values of Δδ are hydrophobic residues with no formal charge (Figure 5, Table 1). Some residues that have a formal positive charge (and are expected to *be* positively charged in UBI) did exhibit large changes in chemical shift in the presence of 1 mM SDS (i.e., Lys11 and Lys29) and we conclude that these residues are specifically interacting with SDS; several residues with a formal negative charge showed little or no perturbation (i.e., Glu18, Asp21, and Asp39). This correlation, however, was *not* uniformly observed for ionic residues: the least perturbed residue in UBI-$(NH_3^+)_8$ at 1 mM SDS, for example, was Lys6 (e.g., $\Delta\delta = 0.00180$) and we conclude that Lys6 did not associate with SDS, whereas a large change in chemical shift was observed for Asp32 from 0 mM to 1 mM SDS (e.g.,

$\Delta\delta = 0.14069$). An analysis of the structure of UBI, obtained by either NMR or X-ray crystallography, indicates that Lys6 is solvent exposed and that the $\varepsilon\text{-}NH_3^+$ is not hydrogen bonded with any nearby residues (Supplemental Figure S6). The reasons why SDS does not associate with this solvent-exposed, positively charged residue, are unknown. We *do* expect that an electrostatic interaction between the $\varepsilon\text{-}NH_3^+$ of Lys6 and the $ROSO_3^-$ of SDS can be detected with $^1H\text{-}^{15}N$ HSQC. Flynn and co-workers have demonstrated, in studies that aimed to mimic molecular confinement and crowding(14), that Lys6 of UBI has the fifth largest value of $\Delta\delta$ (of any residue), when the reverse micelles of sodium bis-(2-ethylhexyl) sulfosuccinate are decreased in size (by varying the amount of water in the organic-aqueous-micellar mixture). We hypothesize that this large value of $\Delta\delta$ for Lys6 is likely to be the result of a direct interaction between $R\text{-}OSO_3^-$ and $Lys\text{-}\varepsilon\text{-}NH_3^+$.

The $^{15}N\text{-}^1H$ HSQC experiments provided incomplete data on the association of arginine residues with SDS. One of the four arginines (Arg 74) could not be assigned at 0 mM SDS, and correlation signals for two others (Arg43, Arg72) could not be assigned at 1 mM SDS, suggesting that these amides interact with surfactant on an intermediate time-scale. A fourth arginine, however, (Arg54) was observed at 0 mM and 1 mM SDS, and the value of $\Delta\delta$ was small ($\Delta\delta = 0.02433$), indicating that Arg54 did not associate with SDS, or at least did not associate with SDS as often, or as tightly, as did other residues in UBI with positive formal charge (i.e., Lys11; $\Delta\delta = 0.17593$).

We did not analyze UBI-$(NHAc)_8$ with HSQC because the acetylation of lysine residues (and the amino terminus) will shift the $^{15}N\text{-}^1H$ correlation signals of residues in unpredictable directions, so that most $^{15}N\text{-}^1H$ correlation signals would need to be completely re-assigned to each residue (the experimental work required to do these assignments is beyond the scope of this

study). Nevertheless, our analysis of UBI-$(NH_3^+)_8$ at 0-1 mM SDS with HSQC shows that three solvent exposed, cationic residues—Lys6, Lys48, and Arg54—have a small value of $\Delta\delta$ (i.e., $\Delta\delta < 0.03$), indicating that these residues do not specifically bind SDS. Therefore, the results of our analysis of UBI-$(NH_3^+)_8$ with HSQC do *not* support the hypothesis—based on the stoichiometry inferred from CE experiments (Figure 1), that the negatively charged sulfate group of SDS associates specifically with positively charged functional groups (lys-ε-NH_3^+ more strongly than Arg-NHC(=NH_2^+)NH_2) on the surface of UBI-$(NH_3^+)_8$ in approximately 1:1 stoichiometry at 0.8-1.4 mM SDS. The results of HSQC do, however, suggest that some cationic residues (i.e., Lys11 and Lys29) are associating, in some fashion, with SDS.

Monitoring the interaction of side chains of UBI and SDS with ^{13}C-^1H HSQC.

We carried out ^{13}C-^1H HSQC experiments to supplement the results of ^{15}N-^1H HSQC experiments that could potentially miss interactions between UBI and SDS. For example, ^{15}N-^1H HSQC suggests that SDS is not interacting with Lys6, however, the backbone amide of Lys6 could possibly be unperturbed by electrostatic interactions between the R-OSO_3^- and ε-NH_3^+ groups. The overlay of ^{13}C-^1H HSQC spectra of UBI with 0 and 1 mM SDS is plotted in Figure 8. This plot of ^{13}C-^1H HSQC spectra reveals a striking attenuation of the intensities of the majority of resonances in the presence of 1 mM SDS. The resonances corresponding to majority of α-CH and β-CH are completely attenuated in the presence of 1 mM SDS, while the remaining few include the α-CH of Ile23 and β-CH of Ala28 and Ala46. The resonance of methyl group of Ala28 is virtually stationary, while that of Ala46 and α-CH of Ile23 are significantly shifted in the presence of 1 mM SDS. These results suggest that Ala46 and Ile23 are involved in specific interactions with SDS, but Ala28 is not. Correspondingly, the amide chemical shift of I23($\Delta\delta$ = 0.11) is significantly greater than of A28($\Delta\delta$ = 0.02). Similarly, for a pair of valines, movement

of any ^{13}C-^{1}H resonances correlates with greater amide shifts: in Figure 8, γ-CH resonances of V17, and V26 move only slightly (^{15}N-^{1}H $\Delta\delta$ = 0.012, 0.02), while γ-CH of V70 shifts dramatically and corresponds to a greater amide shift ($\Delta\delta$ = 0.15).

Several of the few remaining resonances in the HSQC spectra of UBI in 1 mM SDS correspond to the methylenes of lysines; ε-CH, δ–CH and γ-CH. Although the latter do not exhibit significant chemical shifts, we can infer that underlying interactions with SDS result in bound states with lifetimes longer than milliseconds. Existence of such resonances (*e.g.,* γ-CH's of K29, K33, and K63) correlates with larger $\Delta\delta$ of amide HSQC. The resonances of HSQC signals for K6, however, are neither shifted not significantly in the presence of 1 mM SDS: we easily resolved the δ–CH, γ-CH, and ε-CH of K6. This lack of perturbation or attenuation of K6 when studied by either ^{13}C-^{1}H or ^{15}N-^{1}H HSQC provides support for the conclusion that SDS does not interact with this highly solvent exposed lysine residue. Other lysines, such as K27, also produced very small amide chemical shifts, but the ε-CH was completely attenuated. The alkyl protons of SDS contribute to spin-lattice relaxation of aliphatic protons of UBI. The similar energy levels of these protons on the protein and SDS allow for undesirable transfers of magnetization, but these require proximity. The loss of the vast majority of α-CH resonances, including those from the hydrophobic core of UBI in the presence of 1 mM SDS, indicates that the backbone is entirely exposed to interactions with the *n*-alkyl tail. This result suggests that the *n*-alkyl tails of the surfactant transiently penetrate the hydrophobic core of the protein, pushing the strands of the backbone of UBI apart (*e.g.*, the protein-SDS complex undergoes "breathing" motions). This result suggests that the surfactant transiently penetrates the hydrophobic core of the protein, which must swell to accommodate the *n*-alkyl chain between weaves of the tightly backed backbone.

The greater effect of SDS on relaxation of ^{13}C-^{1}H magnetization than of ^{15}N-^{1}H suggests that the alkyl tails make a unique contribution to relaxation of aliphatic protons. This relaxation must come from the undesirable mode of relaxation of magnetization – spin-spin or transverse (T2 relaxation) rather than the typically faster spin-lattice (longitudinal or T1 relaxation). The longitudinal relaxation, strongly dependent on the size of the protein, would affect both carbon and nitrogen equally. The ^{13}C-^{1}H spectrum at 1 mM SDS shows resonances for methyl moieties from valine, leucine, isoleucine, and threonine residues, as well as a number of methylene groups. These groups are known to have slow rates of T1 relaxation, because of free and rapid rotation, so we conclude that dynamic, low-affinity (non-specific) interactions of SDS contribute to spin-spin relaxation of aliphatic protons in UBI. This dynamic interaction of 1 mM SDS affects both Hα and side-chains, but does not disrupt the H-bonding network, or the secondary structure of UBI according to circular dichroism and hydrogen-deuterium exchange (4).

These observations suggests that the effects of SDS on the side-chains of UBI falls in two classes – weak, transient interactions with aliphatic methyl and methylene protons, as well as stable interactions with sub-millimolar affinity and millisecond lifetimes.

SDS (1 mM) interacts preferentially with hydrophobic residues in UBI-(NH_3^+) that have a positive electrostatic surface potential. We hypothesized that a better understanding might emerge as for why SDS associates with some surface residues in UBI-$(NH_3^+)_8$ (i.e., hydrophobic residues such as Ile13 or Leu43), and not other surface residues (i.e., ionic residues such as Lys6 or Asp39; hydrophobic residues such as Val17, Thr22)—and why the acetylation of lysine inhibits the binding of SDS to UBI—when we took into consideration the local electrostatic potential at the surface of folded protein. We used a non-linear Poisson-Boltzmann calculation to approximate the electrostatic potential of the solvent-accessible surface of UBI. Figure 6 shows a

comparison of the electrostatic surface potential of the solvent-accessible surface of UBI-$(NH_3^+)_8$ with the value of Δδ for each surface residue at 0-1 mM SDS. Red color indicates negative electrostatic surface potential; white indicates a zero electrostatic surface potential (i.e., electrostatically neutral regions that are not highly solvated, and negative or positively charged regions whose charge is screened by solvation); blue indicates a positive electrostatic surface potential. We used an X-ray crystal structure of UBI (PDB code 1UBI) for the electrostatic calculations, and for making a color-coded surface-rendering of UBI that illustrates the magnitude of Δδ for each surface residue of UBI at 0-1 mM SDS.

We have also colored the surface of UBI according to the magnitude of the chemical shift Δδ at 0-1 mM SDS. Figure 6 shows the correlation between the local electrostatic surface potential of native UBI, and the magnitude of Δδ at 0-1 mM SDS. In general, a residue contained within patches of UBI-$(NH_3^+)_8$ with positive electrostatic potential interacted with SDS more tightly (or strongly) than a residue with a negative electrostatic surface potential—irrespective of the formal charge of the residue (Figure 6a,b). It is worth noting that UBI-$(NH_3^+)_8$ is, electrostatically speaking, a Janus-faced protein.(37, 38) The electrostatic potential of surface residues of UBI-$(NH_3^+)_8$ seems to reveal why several residues, including hydrophobic residues with no formal charge, have larger values of Δδ than other residues, such as lysine or arginine, which have a positive formal charge. The nine hydrophobic residues of UBI-$(NH_3^+)_8$ that have the largest values of Δδ in 1 mM SDS (e.g., Ile13, Leu43, Leu71, Thr14, Leu50, Leu8, Gly47, Val70, Ile44; $0.15967<\Delta\delta<0.31160$) are located in regions that have a local positive electrostatic potential (Figure 6a)—that is in regions where positive charge is not screened by solvent. In fact, Leu8, Val70, Leu71 and Ile44 are all clustered together in the same hydrophobic face of UBI-$(NH_3^+)_8$ with a positive electrostatic potential (Figure 6a, top left image). We note that this hydrophobic

face (e.g., the 'Ile44 face') of UBI is comprised of a β-sheet and has been indentified, previously, to be the site of recognition of UBI by multiple families of proteins including UBAs(39) (ubiquitin-associated domains), UIMs(40) (ubiquitin-interacting motifs), CUE (41) (coupling of ubiquitin to ER degradation), and NZFs (42) (N1p14 zinc fingers), as well as sites of UBI-UBI interations of polyubiquitin chains.(43)

In contrast, residues that have the lowest values of $\Delta\delta$ at 0-1 mM SDS are located on the opposite face of UBI, and in a region with negative electrostatic potential (e.g., Val17, Thr22, Tyr59, Asn60; $0.01098<\Delta\delta<0.01468$). We also point out that the only arginine residue for which we have data, Arg54, has a small value of $\Delta\delta$ ($\Delta\delta = 0.02433$), and that moreover, Arg54 does not have a positive electrostatic surface potential (according to the Poisson-Boltzmann calculation). The approximately zero electrostatic surface potential of Arg54 (Figure 6a), that was calculated from the PB equation, is caused by (according to the model, and its approximations) a high degree of solvation of the guanidino group of Arg54; the highly solvated nature of Arg54 might explain why SDS does not have a specific electrostatic interaction with Arg54. Arg54 is, moreover, flanked on three sides by residues with a negative electrostatic potential, which will contribute to the neutralization of the surface potential of Arg54 (via screened Coulombic interactions). A previous study by Flynn and co-workers has shown that Arg54 has the second lowest chemical shift perturbation of any residue in UBI, when folded UBI was confined inside reverse micelles of sodium bis-(2-ethylhexyl) sulfosuccinate (in this study, $\Delta\delta < 0.005$ for Arg54). We also rationalize these previous observations in terms of the zero electrostatic surface potential of Arg54, and its surrounding by anionic residues.

The major finding of Figures 6-8—that SDS interacts preferentially (below the CMC) with hydrophobic residues that have positive electrostatic surface potential—is consistent with the

result shown in Figure 1, that the acetylation of 7 lysine-ε-NH_3^+ (and one α-NH_3^+) inhibits the binding of 11 SDS below the CMC. We conclude that the peracetylation of UBI-$(NH_3^+)_8$ inhibits the binding of SDS to (primarily) hydrophobic residues on the surface of UBI, and to a lesser degree, cationic or polar residues, because peracetylation affects the electrostatic surface potential of these hydrophobic residues, and not because peracetylation neutralized specific R-NH_3^+ sites that could have been points of contact with the sulfate group of SDS.

The electrostatic surface potential of amino acid residues in UBI cannot, however, entirely explain the degree with which these residues associate with SDS: Thr12 and Lys6, for example, both have positively charged surfaces, but have low values of Δδ (Lys6 = 0.01109 and Thr12 = 0.00180); Asp32 has one of the largest values of Δδ for any residue (0.14069) but has a negative electrostatic potential, and is surrounded by many negatively charged residues (Figure 6).

The general observation that hydrophobic residues in UBI-$(NH_3^+)_8$ have larger values of Δδ than cationic residues (at 0-1 mM SDS) demonstrates that hydrophobic interactions between the dodecyl chain of SDS are a large driving force for the specific binding of SDS to UBI. In order to begin to quantify how the hydrophobicity of a residue in UBI-$(NH_3^+)_8$ affects its association with SDS, we plotted the average value of Δδ for each type of amino acid (*i.e.,* lysines, leucines, aspartates, etc.) as a function of their hydrophobicity (Figure 7). The values of hydrophobicity that we used for each amino acid were determined previously (44) by measuring the partitioning of derivatives of each amino acid into octanol ($C_8H_{17}OH$). In this plot, we only included amino acids that were highly represented in the sequence of ubiquitin (e.g., residues that were located at > 3 positions), without regard to the location of these residues in folded UBI (e.g., on the surface, or hydrophobic interior). The hydrocarbon chain of SDS is linear and saturated, and is therefore

similar, structurally and chemically, to the alkyl chain of octanol (with the exception of four additional methylene groups).(44)

We observed a significant correlation between the interaction of well-represented amino acids in UBI with 1 mM SDS and their hydrophobicity, Figure 7. This plot shows that the ^{15}N-^{1}H resonance signals of hydrophobic amino acids are more perturbed by 1 mM SDS than hydrophilic residues. A linear least-squares regression of each data set in Figure 7 yielded a $R^2 \approx 0.75$. We infer from this correlation that the SDS associates preferentially with hydrophobic sidechains, not hydrophilic or cationic amino acids.

The chemical shifts of amides and modeling the surface potential suggest that the highest affinity sides are hydrophobic residues in regions of positive electrostatic potential, such as those involved in formation of the first stable intermediate UBI-$(NH_3^+)_8(SDS)_{\sim 11}$.

A recent study, by Xiao et al., has used ^{1}H-^{15}N HSQC NMR spectroscopy to monitor the binding of sodium perfluorooctanoate (SPFO, an anionic surfactant) to UBI (0.5 mM, pH 5.7), below and above the CMC (the CMC of SPFO was reported to be 31.0 mM).(15) This study reported that the structure of UBI is largely unaffected by concentrations of SPFO < 3 mM. Even though the Xiao et al. study involved a surfactant with a perfluorinated tail, and studied UBI below its isoelectric point, we find that our observations are qualitatively similar. The previous study by Xiao et al. supports, for example, our observation that Lys6 and Arg54 are *not* associating with SDS at 1 mM SDS: the ^{1}H-^{15}N correlation signal for Lys6 and Arg54 were not affected—either in intensity or position—by the presence of 2.9 mM SPFO.

The conformational flexibility and solvent accessibility of a residue in UBI-$(NH_3^+)_8$ and its association with SDS. We also suspected that other biophysical properties of residues in UBI, such as the conformational flexibility, or solvent accessibility, might also contribute to the

binding of SDS with UBI below the CMC. Because UBI is a model protein that has been studied in depth by structural biologists, there are biophysical and structural data available for most amino acids of the folded protein. In order to determine how the structural properties of residues of UBI-$(NH_3^+)_8$ might influence their association with SDS—and perhaps explain the Δδ values of some residues (i.e., Lys6, Thr12, Asp32)—we compared the values of Δδ from ^{15}N-^{1}H HSQC at 0-1 mM SDS with some of these available data, including: i) the solvent accessibility of the side chain (SA_{Side}) and main chain (SA_{Main}) of each residue, as estimated from the crystal structure of UBI, and ii) the rate of amide hydrogen/deuterium (H/D) exchange of each residue in UBI (measured with NMR). Figure S7 (supplemental information) shows a comparison of Δδ of each residue in 0-1 mM SDS, and the static solvent accessibility of the side chain and main chain of each residue in folded UBI. A comparison of the Δδ of each residue in 0-1 mM SDS, and the rate of amide hydrogen-deuterium exchange of each residue is also shown in Figure S7. We did not observe any general correlation between the Δδ of a residue, its accessibility to solvent, or the rate of amide hydrogen-deuterium exchange of that residue. A detailed discussion of these results is included in the supplemental material.

Characteristics of the binding sites of SDS on the surface of folded ubiquitin. Our analysis of UBI at 0 and 1 mM SDS with ^{15}N-^{1}H and ^{13}C-^{1}H HSQC, and a comparison of these results with the electrostatic surface potential, secondary structure, hydrophobicity, solvent accessibility, and rate of H/D exchange of residues in UBI has shown that the first stable intermediate with SDS stably bound to UBI, such as occurs at 1 mM SDS, largely involves regions where hydrophobic side chains are adjacent to positively charged residues. The association of SDS to specific amino acids in folded UBI is not strongly influenced by type of secondary structure – whether located in a β-sheet, or an α-helix, however the loops are

characterized by small chemical shifts (Figure 6). Presumably, the loops provide a less-stable docking platform for SDS than either a β-sheet, or an α-helix. The results of this analysis are summarized in Table 1, which lists some of the structural, biophysical and physical-organic properties of the twelve residues that were most affected by 1 mM SDS (e.g., $\Delta\delta > 0.14$), and the properties of the twelve residues that were the least affected by 1 mM SDS (e.g., $\Delta\delta < 0.02$).

There are few structures available in the Protein Data Bank of SDS coordinated to folded proteins. The only example that we could find is the recent X-ray data of lysozyme(45) obtained from protein crystals that were grown in the presence of 2 mM SDS: a single dodecyl sulfate molecule was observed to be bound at the surface of lysozyme.(45) Details of this X-ray crystal structure, and details of X-ray crystal structures of other proteins coordinated with sulfatide, are included in the supporting information.

Predicting the structure of SDS molecules that are bound to the surface of UBI (i.e., at 1 mM SDS) with computational docking methods—for example, predicting where the sulfate group binds, and the conformation of the alkyl tail—is likely to be difficult because the binding of each additional SDS molecule will alter both the electrostatic surface potential of UBI, and the hydrophobicity. From considering the results of Figure 7—that hydrophobic functional groups bind SDS more favorably than ionic or polar groups—we find no reason to exclude the possibility that the alkyl chains of different SDS molecules, which are bound to the surface of folded UBI, are interacting with one another.

Electrostatic and hydrophobic interactions between SDS and UBI occur cooperatively. Does UBI recognize the hydrophobic tail of SDS *more* than the anionic head, at concentrations of 1 mM SDS? Our CE and HSQC results demonstrate that native UBI-$(NH_3^+)_8$ provides a type of surface (hydrophobic, but with a positive electrostatic surface potential) that allows electrostatic

and hydrophobic interactions between SDS and UBI to occur cooperatively. The residues in UBI that were found to have the largest values of Δδ (at 1 mM SDS) were hydrophobic residues that had the greatest magnitude of positive electrostatic surface potential (i.e., Ile13, Leu 43; Table 1). The majority of the residues with the least interaction with SDS were hydrophobic residues that had the greatest magnitude of negative electrostatic surface potential (i.e., Val17, Thr22; Table 1). We infer from the HSQC analysis that salt-bridges are not forming between the sulfate group of SDS and positively charged groups of lysine and arginine (although salt bridges between *some* lysine or arginine residues might be forming). This absence of a general interaction between SDS and Lys or Arg is not so surprising when the electrostatic surface potential of UBI (Figure 6) is taken into account: lysine and arginine residues in UBI do not always have the largest positive electrostatic surface potentials (i.e., Arg54 ≈ 0 eV; Ile13 ≈ 4 eV) because, presumably, solvent is screening the charge of these residues. Therefore, the location of a salt-bridge—if one were to form between UBI and SDS—would be with a residue bearing a poorly solvated charge.

Conclusions.

The strength of association of residues in ubiquitin and SDS at sub-denaturing concentrations (< 1.0 mM), seems to integrate three properties of the amino acid residues: i) the hydrophobicity of the amino acid residue, ii) its local electrostatic environment, and iii) the 2° structure of the residue (e.g., a hydrogen-bonded loop versus α-helix or β-strand; Table 1). None of these three factors, alone, seem sufficient to explain the binding of SDS to residues in ubiquitin below the CMC. We infer that—at least for UBI—the presence of positive charges facilitates the binding of molecules of SDS at concentrations that are below the CMC, and that these surfactants bind preferentially to regions with positive electrostatic surface (this location may or may not be at

cationic residues, depending upon solvation); salt bridges do not necessarily form between the sulfate group and cationic groups in the protein (although some might be forming, for example, between Lys11 and SDS). We conclude that the acetylation of lysine residues in UBI inhibits the binding of the initial ~ 11 molecules of SDS, below the CMC of SDS, primarily because acetylation increases the negative surface potential of UBI, and *not* because acetylation eliminates amino functionalities that might have been sites of binding for the sulfate group of SDS. In other words, changing the electrostatic surface potential affects the affinity of binding sites for SDS due to delocalized interactions of surface charges and sulfate groups. In general, the two amino residues in UBI that interacted most favorably with SDS, below the CMC, were the branched hydrophobic residues leucine and isoleucine, so long as they were adjacent to areas of positive electrostatic surface potential, and were not part of a hydrogen-bonded loop. Specific interactions of SDS with aliphatic protons of side-chains resulted in shifting of several ^{13}C-^{1}H resonances, but the majority of side-chains and all of Hα's (with the exception of the flexible N-terminus) were attenuated by dynamic broadening of resonances resulting from rapid kinetics of binding.

Previous work has suggested that β-stranded proteins are, in general, more kinetically resistant to unfolding by SDS than are α-β proteins(46), but we did not find with ubiquitin that similar amino acids residues bind SDS differently in β-strands and α-helices. Moreover, the solvent accessibility of the *side-chain* of a residue, and the conformational flexibility of the residue, do not seem to correlate with its association with SDS at 1mM SDS, although Table 1 suggests that peptide *backbones* that are solvent exposed interact with SDS slightly more than backbones that are buried (Table 1).

Our results also reveal a few points regarding the unfolding of UBI in SDS. Importantly, the difference between the native and peracetylated UBI of eleven molecules of bound SDS is observed at concentrations of SDS that are below, and above the CMC; this observation suggests that the eleven SDS molecules bound in the folded state remain bound in the unfolded state. Nevertheless, the pathway of association of SDS with UBI in the *unfolding* regime is not influenced by the initial number of cationic residues: according to CD spectroscopy both UBI-$(NH_3^+)_8$ and UBI-$(NHAc)_8$ unfold *via* the binding of indistinguishable numbers of SDS molecules (e.g., fourteen SDS; see the circular dichroism spectra in Figure S5).

We have previously used capillary electrophoresis to study the unfolding of several proteins in SDS (5), but ubiquitin (e.g., UBI-$(NH_3^+)_8$) is unique among the 18 that we studied in that UBI binds multiple equivalents of SDS below the CMC without unfolding. Most proteins that we have studied do not associate with large numbers of SDS below the CMC; most proteins do not bind SDS *without* unfolding, and most do not follow a complex pathway of unfolding that involves multiple stable, distinct complexes with defined stoichiometry. We still do not understand why UBI forms a stable intermediate with SDS below the CMC, without unfolding, and why UBI forms more distinct complexes with SDS, and populates more intermediates during unfolding than other proteins that we have studied. This study has shown, however, that the peracetylation of UBI causes the protein to unfold in a manner that *is similar* to most of the other proteins that we have studied: peracetylation eliminates the ability of the protein to coordinate SDS without unfolding and causes UBI to unfold in SDS via a two-state transition.

In studies of a subject—the association of proteins with surfactants in particular, and molecular recognition in general—that is important, but *very* complicated, this work represents a step toward a more detailed understanding. The combination of capillary electrophoresis and

protein charge ladders is certainly a useful, new technique to investigate quickly (and reproducibly) various hypotheses such as the influence of the structure or the surface chemistry of proteins, on the binding pathway of surfactants and other charged molecules to proteins.

Acknowledgments. This research was funded by NIH award GM 051559; BFS was supported by a NIH Ruth L. Kirchstein NRSA post-doctoral fellowship (GM081055) and GFS was supported by a "Lavoisier Générale" post-doctoral fellowship (Ministère des Affaires Etrangères Français). Molecular graphics images of UBI were produced using the UCSF Chimera package from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIH P41 RR-01081).

Supporting Information Available: Circular dichroism spectra of UBI-$(NH_3^+)_8$ and UBI-$(NHAc)_8$, high resolution CE electrophorograms of the transition $N \rightarrow G_1^* \rightarrow G_2$, electropherograms of BCA dialyzed against SDS.

Materials and Methods.

Chemicals. Ubiquitin from bovine erythrocytes (all experiments with the same lot #075K7405), Tris-NH_2, Glycine, sodium dodecyl sulfate, acetic anhydride, dioxane, N,N-dimethylformamide, and HEPBS were purchased from Sigma-Aldrich. Dialysis microtubes (Slide-A-Lyzer MINI, MWCO 3.5k) were purchased from Pierce and rinsed several times with ultra-pure water prior to use. SDS was recrystalized three times from hot ethanol, and the absence of any alcohol precursor (dodecanol) was checked by NMR. Tris-glycine buffer (25 mM Tris, 192 mM glycine) was prepared from salts purchased from Aldrich and were filtered through a 0.22-μm membrane prior to use. Preparation of UBI in solutions of SDS by dialysis were carried out as described in a previous study.(4)

Synthesis of Peracylated UBI. The acylation of UBI (20 μM) with acetic anhydride was carried out in 100 mM HEPBS at pH 9.0 (HEPBS: N- (2-Hydroxyethyl)piperazine-N′-(4-butanesulfonic acid)). Neat acetic anhydride was diluted 1:10 into dioxane, 100 stoichiometric equivalents (with respect to lysine and N-terminus) were added to buffered solutions of UBI and the reaction was allowed to proceed overnight at room temperature. The concentration of dioxane in the aqueous solution of UBI did not exceed 5%. Dioxane was eliminated by extensive dialysis against Tris-glycine. The peracetylated protein was then concentrated using Centricon centrifugal filtration devices (Millipore, 5 kDa MWCO) to a concentration of 50 μM.

Capillary Electrophoresis (CE): We used a Beckman Coulter P/ACE electrophoresis apparatus, with a capillary length of 60.2 cm (50 cm to the detector), and composed of fused silica (inner diameter of 50 μm; Polymicro Technologies TSP050375, Part#2000017, lot LHQ02A). Buffers used were composed of Tris-glycine (pH 8.4) containing concentrations of SDS (from 0 to 10 mM). CE was performed at an applied voltage of 30 kV for 10 minutes. The capillary was equilibrated before each electrophoresis by a sequential rinsing with i) methanol, ii) HCl (1M), iii) NaOH (0.1 M), iv) ultrapure water (Millipore Advantage Ultrapure Water System, conductivity of 18.2 MΩ.cm and total organic content not exceeding 4 ppb), and v) the running buffer for the electrophoresis (3 minutes at 20 psi prior to each injections of protein). The solutions containing UBI-SDS complexes (~50 μM; we used an extinction coefficient for UBI of ε(280) = 1490 L/(mol.cm))(47) were injected in the capillary at 0.5 psi for 20 s. Samples were injected at a temperature of 4 °C and the temperature of the capillary was maintained at 25 °C during electrophoresis. Each series of experiments was carried out with the same capillary, equilibrated after each set of 20 runs.

The mobility of a UBI-SDS complex was calculated according to Eq. 4 where μ is the

mobility of the migrating species (in $cm^2\ kV^{-1}\ min^{-1}$), L_T is the total length of the capillary, L_D the length up to the detector, V the applied voltage, t_{eof} the time of the neutral marker, and t the time of the migrating species:

$$\mu = \frac{L_D \cdot L_T}{V}\left(\frac{1}{t_{eof}} - \frac{1}{t}\right) \qquad (4)$$

Circular Dichroism (CD) Spectroscopy. CD spectra were collected on a spectropolarimeter (Jasco Inc.) equipped with a cuvette of quartz with a 1-mm pathlength. For far-UV measurements (205-260 nm) the concentration of proteins was 50 μM (at 25 °C). Each sample was scanned 10 times at 20 nm/min with a step size of 0.5 nm. The molar ellipticity θ ($deg \cdot M^{-1} \cdot cm^{-1}$) was obtained by using Eq. 5 with L_1 and L_2 (the pathlength of the cuvette used for the measurement of the absorbance at 220 nm is L_1 = 1 cm; for the cuvette used to collect CD spectra, L_2 = 0.1 cm).

$$\theta(\text{deg} \cdot M^{-1} cm^{-1}) = \frac{\theta_{measured}(\text{deg}) \cdot \varepsilon_{220}(L.mol^{-1}.cm^{-1}) \cdot L_1}{A_{220} \cdot L_2} \qquad (5)$$

^{15}N-^{1}H and ^{13}C-^{1}H HSQC NMR. The NMR experiments were carried out on Bruker 600 MHz spectrometer equipped with a cryoprobe. In all experiments were conducted using 25 mM $(DOCD)_3CND_2$ (or D(11)-Tris) in 1:9 H_20:D_20 at pH 7.53 with the samples at room temperature (20 ºC). The per-deuterated form of Tris buffer was obtained from Cambridge Isotope Labs (Cambridge, MA), as well as ^{15}N- and ^{13}C-^{1}H human ubiquitin (76 amino acids). ^{15}N- ^{1}H HSQC spectra of uniformly ^{15}N-labelled ubiquitin (100 μM) were collected for 45 minutes, while those in the presence of 1 mM SDS were collected over a period of 12 hours and then averaged. ^{13}C HSQC spectra of uniformly ^{13}C- and ^{15}N-labelled ubiquitin (150 μM) in absence of SDS were collected for 2 hours, and those in 1 mM SDS was collected over a period of 4.5 hours, maintaining the samples at room temperature. The ^{15}N-^{1}H HSQC experiments were performed

using solutions with 150 μM UBI-$(NH_3^+)_8$. We chose this concentration of UBI (low for a typical protein ^{15}N-^{1}H HSQC experiment) so that the UBI:SDS ratio would be as comparable with CE experiments as possible (CE experiments were carried out at 50 μM UBI-$(NH_3^+)_8$).We assigned residues to ubiquitin at pD 7.53 and processed images using Cara (v 1.8.4). We used Sparky software to analyze the chemical shifts (Δδ).

Figure 1. Studying the binding of SDS to ubiquitin and peracetylated ubiquitin with capillary electrophoresis. CE electropherograms of UBI-SDS complexes prepared by dialyzing UBI-$(NH_3^+)_8$ (a) and UBI-$(NHAc)_8$ (b) against Tris-glycine (25 mM Tris, 192 mM glycine) containing different concentrations of SDS (0 – 10 mM). In these electropherograms, UBI-$(NH_3^+)_8$ at 0 mM SDS is denoted "N", and the successive complexes that form during the unfolding at higher concentrations of SDS (e.g., the UBI-SDS_n complexes) are denoted "G_x", where x = the order in which the complex is formed along the unfolding pathway. a) Because the net charge, Z_0, of UBI-$(NH_3^+)_8$ is so close to zero at pH 8.4 and 0 mM SDS ($Z_0 = -0.2$), the mobility of UBI-$(NH_3^+)_8$ at 0 mM SDS (e.g., "N") is similar to the mobility of the neutral marker dimethylformamide (DMF) at 0 mM SDS. Therefore, the peak for DMF appears as a left-handed shoulder, at 0 mM SDS, that is unresolved from the peak for UBI-$(NH_3^+)_8$ at 0 mM SDS. The peak for UBI-$(NH_3^+)_8$ shifts to a higher mobility as the concentration of SDS is increased (as a result of the binding of SDS to UBI), but the neutral marker remains at $\mu = 0$ $cm^2\,kV^{-1}\,min^{-1}$. b) UBI-$(NHAc)_8$ has a mobility of $\mu = 15$ $cm^2\,kV^{-1}\,min^{-1}$ at 0 mM SDS, and this mobility is constant up to ~2.0 mM SDS, whereas the mobility of UBI-$(NH_3^+)_8$ in (a) was increasing from 0 to 2 mM SDS as a result of its binding with SDS. In both (a) and (b), electrophoresis was performed at 30 kV using a capillary made out of fused silica (total length of 60.2 cm, length to the detector of 50 cm, temperature = 25 °C).

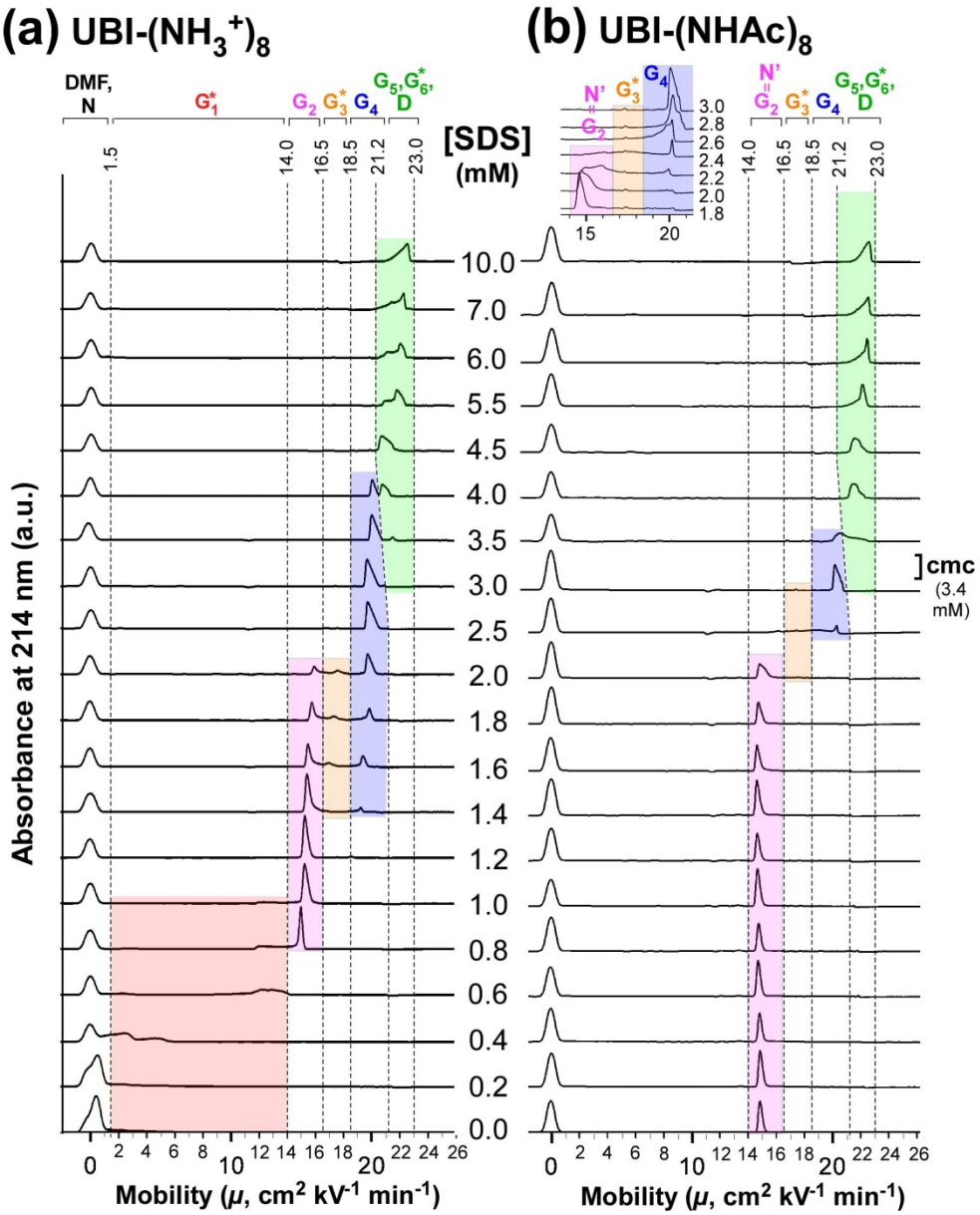
(a) UBI-(NH3+)8
(b) UBI-(NHAc)8
DMF, N
G1*
G2
G3*
G4
G5,G6*, D
N'
[SDS] (mM)
1.5
14.0
16.5
18.5
21.2
23.0
10.0
7.0
6.0
5.5
4.5
4.0
3.5
3.0
2.5
2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
3.0
2.8
2.6
2.4
2.2
2.0
1.8
cmc (3.4 mM)
Absorbance at 214 nm (a.u.)
Mobility (μ, cm2 kV-1 min-1)

FIGURE 1

Figure 2. SurfCE kinetic electropherograms of UBI-$(NH_3^+)_8$ in the presence of SDS. UBI is injected in its native form – N – into a capillary previously filled with Tris-glycine and SDS, at a concentration of SDS labeled on each electropherograms). The ruler that is overlaid on the top of the electropherograms corresponds to the mass-calibrated mobilities of a UBI charge ladder determined previously,(4) and enabled us to determine the stoichiometry (e.g., the number *n*) of UBI-SDS_n complexes. The labile G_1 complex (red box) is resolved due to the equilibrium of bound and free SDS in the capillary.

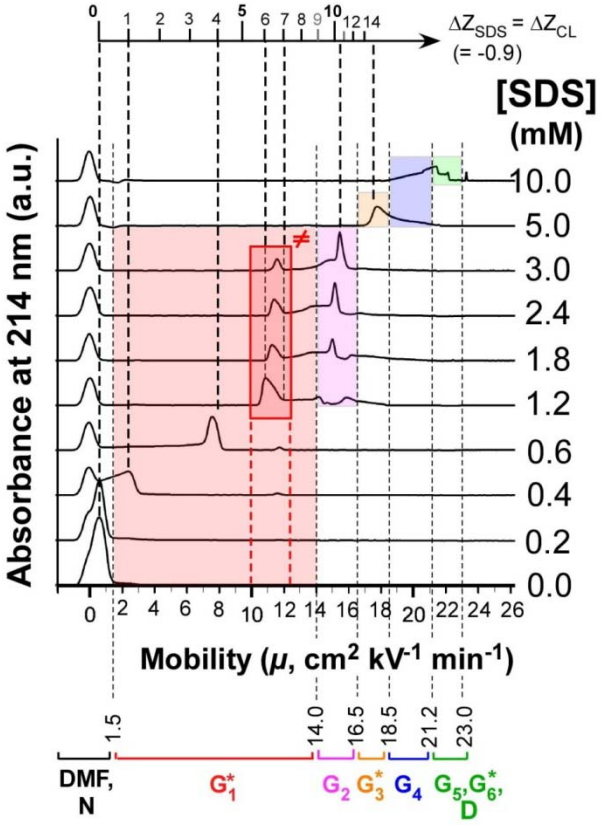
0 1 2 3 4 5 6 7 8 9 10 12 14
ΔZSDS = ΔZCL
(= -0.9)
[SDS]
(mM)
10.0
5.0
3.0
2.4
1.8
1.2
0.6
0.4
0.2
0.0
≠
Absorbance at 214 nm (a.u.)
0 2 4 6 8 10 12 14 16 18 20 22 24 26
Mobility (μ, cm2 kV-1 min-1)
1.5
14.0
16.5
18.5
21.2
23.0
DMF, N
G1*
G2 G3* G4 G5, G6*, D

FIGURE 2

Figure 3. Diagram summarizing the binding pathway of SDS to UBI-$(NH_3^+)_8$ (colored blue) and peracetylated UBI (colored red). The numbers of SDS molecules that are bound to each protein at different concentrations of SDS (0-10 mM SDS) are shown. The numbers of additional SDS that bind to each protein along the pathway of unfolding are also shown. Note: the structure of UBI:SDS complexes near the micellar regime are purely speculative, and may (in actuality) not involve the type of configurations that are shown (i.e., the so-called 'pearl necklace' or 'bead on a string' configuration).

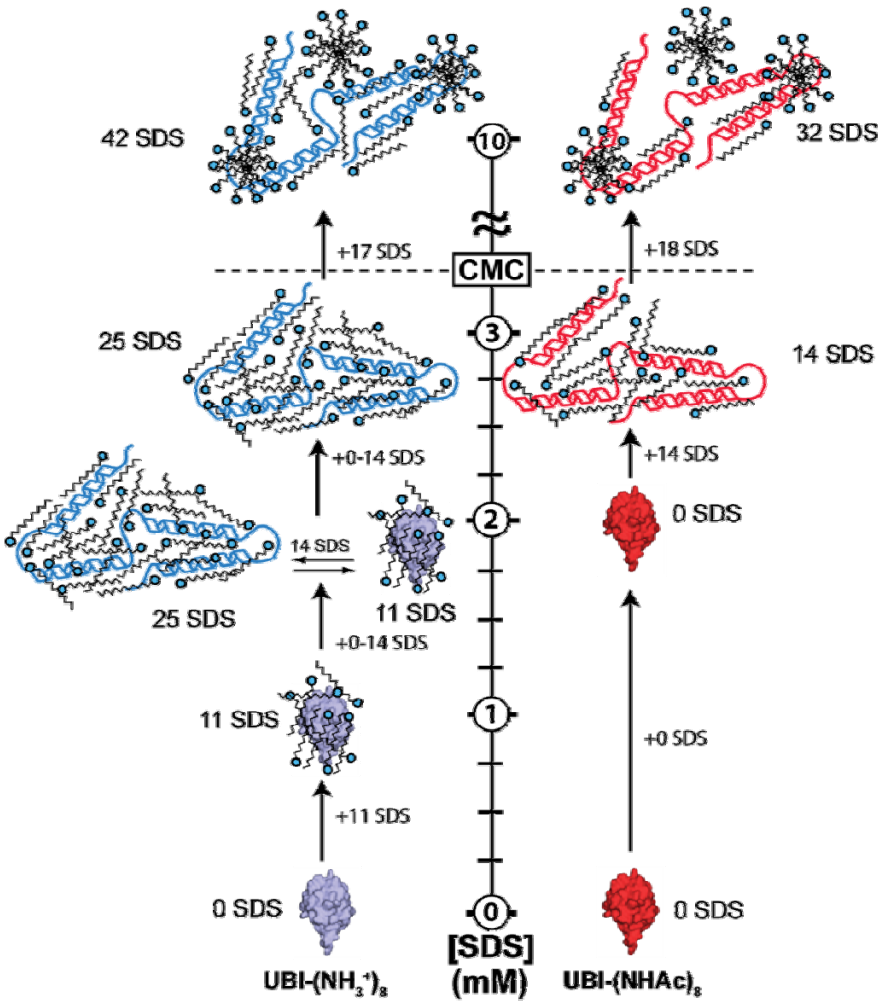
42 SDS
32 SDS
10
+17 SDS
CMC
+18 SDS
25 SDS
3
14 SDS
+0-14 SDS
+14 SDS
0 SDS
14 SDS
25 SDS
11 SDS
2
+0-14 SDS
11 SDS
1
+0 SDS
+11 SDS
0 SDS
0
0 SDS
[SDS]
(mM)
UBI-(NH3+)8
UBI-(NHAc)8

FIGURE 3

Figure 4. Studying the binding of SDS to UBI at 1 mM SDS with ^{15}N-^{1}H HSQC NMR. (a) HSQC spectra of human ubiquitin (labeled with ^{15}N) in the presence of 0 and 1 mM SDS (pH 7.53, 25 mM Tris); the concentration of ubiquitin was 150 μM. Expanded version showing Lys6, Leu67, and Val70. Residue labels are positioned next to the ^{15}N-^{1}H correlation signal for each residue at 0 mM SDS (blue). The correlation signal of Lys6 was not shifted by the presence of 1 mM SDS, indicating that Lys6 did not interact with SDS; Leu67 and Val70 do interact with SDS, as demonstrated by the shifting of their signals from 0 mM to 1 mM SDS. (b) Expanded version of full spectra shown in (a); residue labels are positioned next to the ^{15}N-^{1}H correlation signal for each residue at 0 mM SDS (blue).

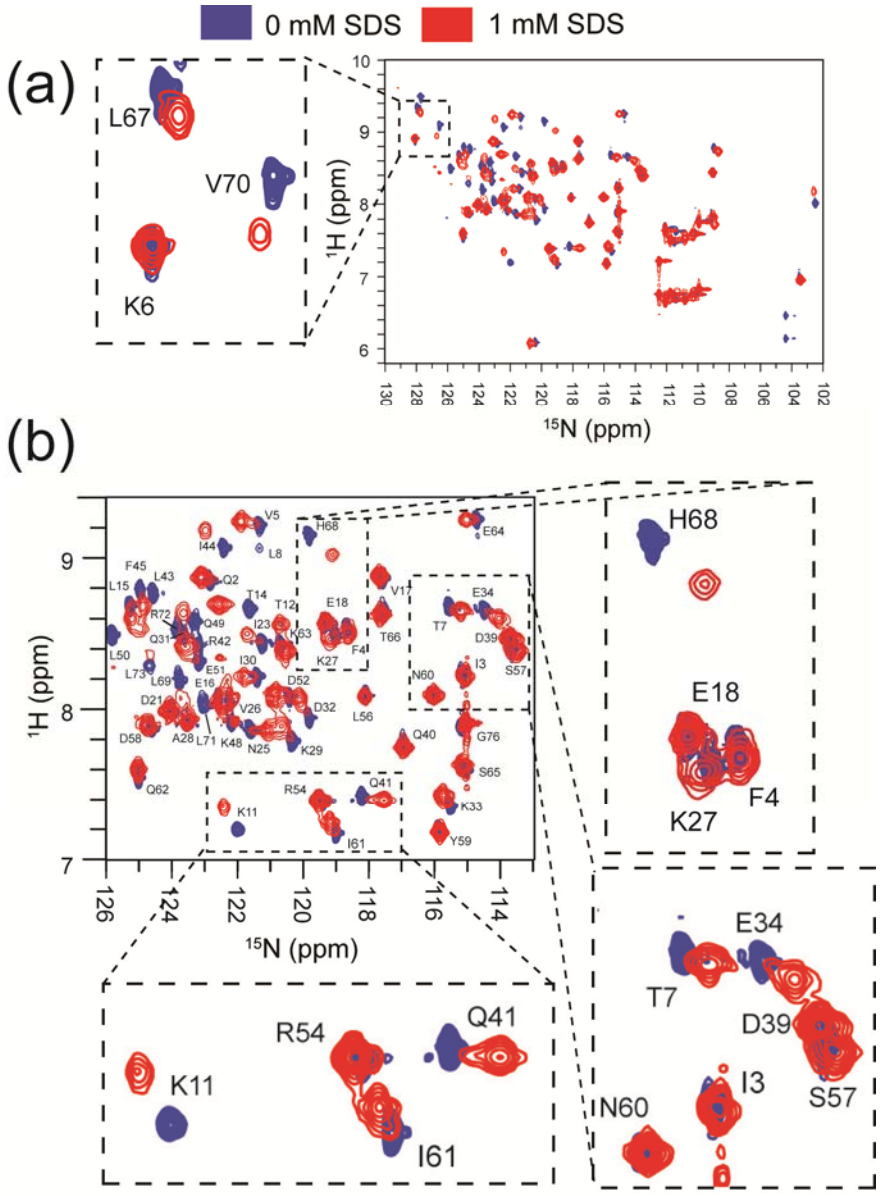
0 mM SDS
1 mM SDS
(a)
L67
V70
K6
^{1}H (ppm)
^{15}N (ppm)
(b)
H68
E18
F4
K27
E34
T7
D39
I3
S57
N60
R54
Q41
K11
I61
^{1}H (ppm)
^{15}N (ppm)

FIGURE 4

Figure 5. Bar graph showing the magnitude of change (Δδ) in the ^{15}N-^{1}H correlation signal of each residue in ubiquitin that resulted from the presence of 1 mM SDS. The bars for residues with formal positive charge (e.g., Lys, Arg) are colored blue; residues with formal negative charge (i.e., Asp and Glu) are red; the lone His is colored blue. The bars for all other residues are gray. The x-axis lists the residue number in the 76 residue UBI polypeptide. Residues with the largest values of Δδ are labeled with their amino acid label. The branched hydrophobic residues Ile and Leu are colored orange. A linear map of the secondary structure of folded UBI is listed below the x-axis. The least perturbed residue is Lys6; the most perturbed is Ile13. A vertical dashed line indicates that the peak disappeared from the spectrum of UBI in 1 mM SDS, or that its assignment was ambiguous. Asterisks denote the three residues that have not been assigned ^{15}N-^{1}H correlation signals in the HSQC spectrum of UBI (Met1, Thre9 and Gly10); double asterisks denote a signal that is too weak to detect; proline residues (denoted P) are not observable in this type of experiment. The map of the secondary structure of UBI is based upon the X-ray crystal structure (PDB code: 1UBI). Note: the hydrogen-bonded loops present at residues 56 and 57 are sometimes described as a 3_{10}-helix.

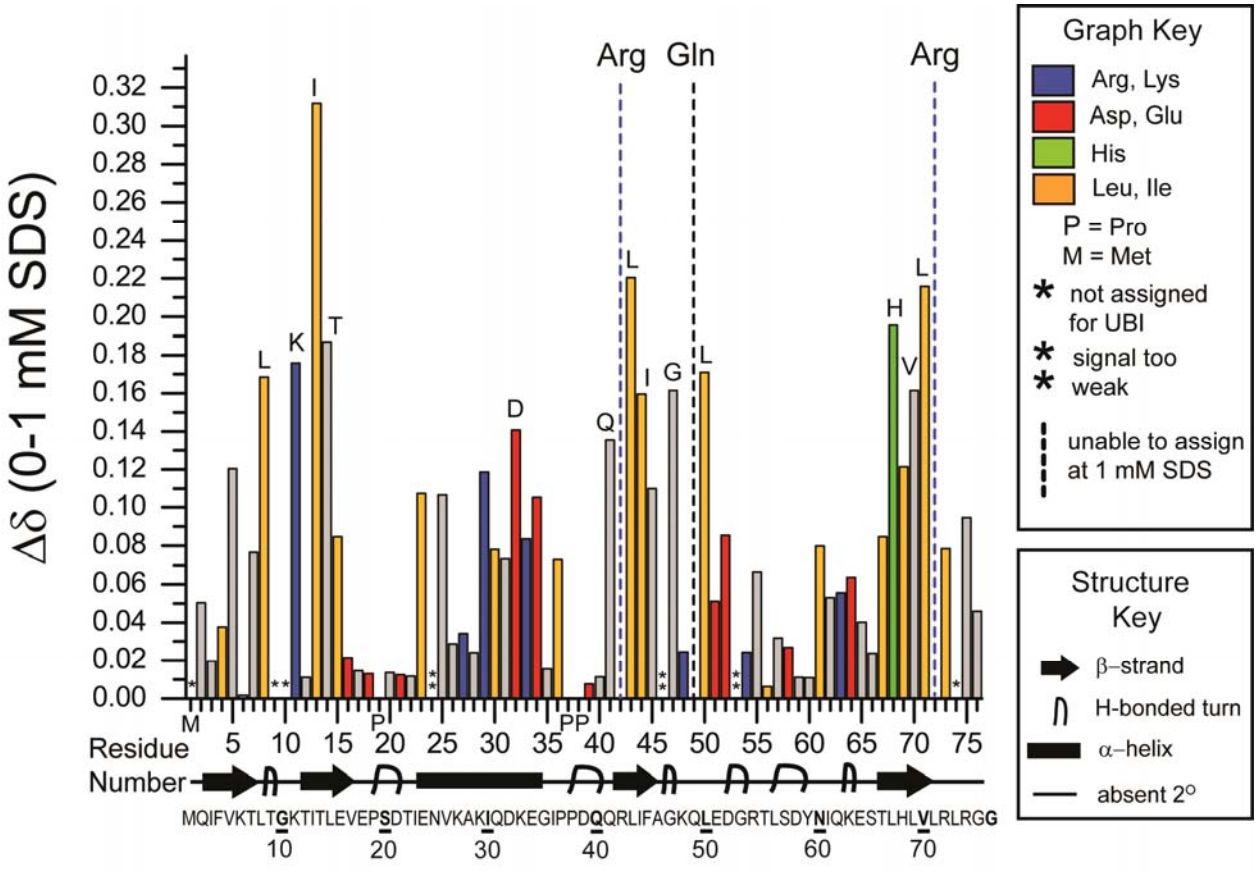
Δδ (0-1 mM SDS)
0.32
0.30
0.28
0.26
0.24
0.22
0.20
0.18
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
Arg
Gln
Arg
Residue Number
M 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75
MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDYNIQKESTLHLVLRLRGG
10 20 30 40 50 60 70
Graph Key
Arg, Lys
Asp, Glu
His
Leu, Ile
P = Pro
M = Met
* not assigned for UBI
** signal too weak
unable to assign at 1 mM SDS
Structure Key
β–strand
H-bonded turn
α–helix
absent 2°

FIGURE 5

Figure 6. SDS associates with non-polar residues on the surface of UBI that have positive electrostatic surface potential. (a) The electrostatic surface potential distribution (-4 to +4 eV) of the solvent accessible surface of ubiquitin was calculated using a non-linear Poisson-Boltzmann equation (based upon the X-ray crystal structure of ubiquitin; PDB code: 1UBI). The intensity of blue illustrates regions with positive electrostatic potential; red indicates negative electrostatic potential (white indicates zero electrostatic potential). Various surface exposed residues (hydrophobic, cationic, anionic) are labeled. (b) Surface rendering and (c) cartoon rendering of the crystal structure of ubiquitin (PDB code: 1UBI). The color-code in (a) and (b) illustrates the magnitude of the chemical shift perturbation ($\Delta\delta$) for each residue, during HSQC experiments at 0 and 1 mM SDS. Red represents residues with small chemical shift perturbations; white, intermediate chemical shift perturbations; blue, large chemical shift perturbations. Gray indicates residues that could not be studied (i.e., proline, or residues that could not be assigned a ^{15}N-^{1}H correlation signal). Surface-exposed lysine (K) residues and the single Histidine residue (H) are labeled. The surface of UBI that was most affected by 1 mM SDS was made up of non-polar, hydrophobic residues that existed in a positively charged environment: Leu71, Leu43, Leu8, Val70, Ile44, and Gly47. The surface of UBI that was least affected by SDS was a polar region, containing multiple glutamate and aspartate residues (and one valine), that was negatively charged: Glu16, Val17, Glu18, Asp21, and Thr22. The lone arginine that was observable in HSQC experiments (Arg54) had a small value of $\Delta\delta$ (0.02433 ppm) and is electrostatically neutral, but is also flanked on three sides by residues with a negative electrostatic surface potential. Remarkably, Lys6 and Lys48 also have a small value of $\Delta\delta$ (Lys6 has the smallest $\Delta\delta$ of any residue at 1 mM SDS), but both residues have a positive electrostatic surface potential, and are surrounded by other positively charged residues.

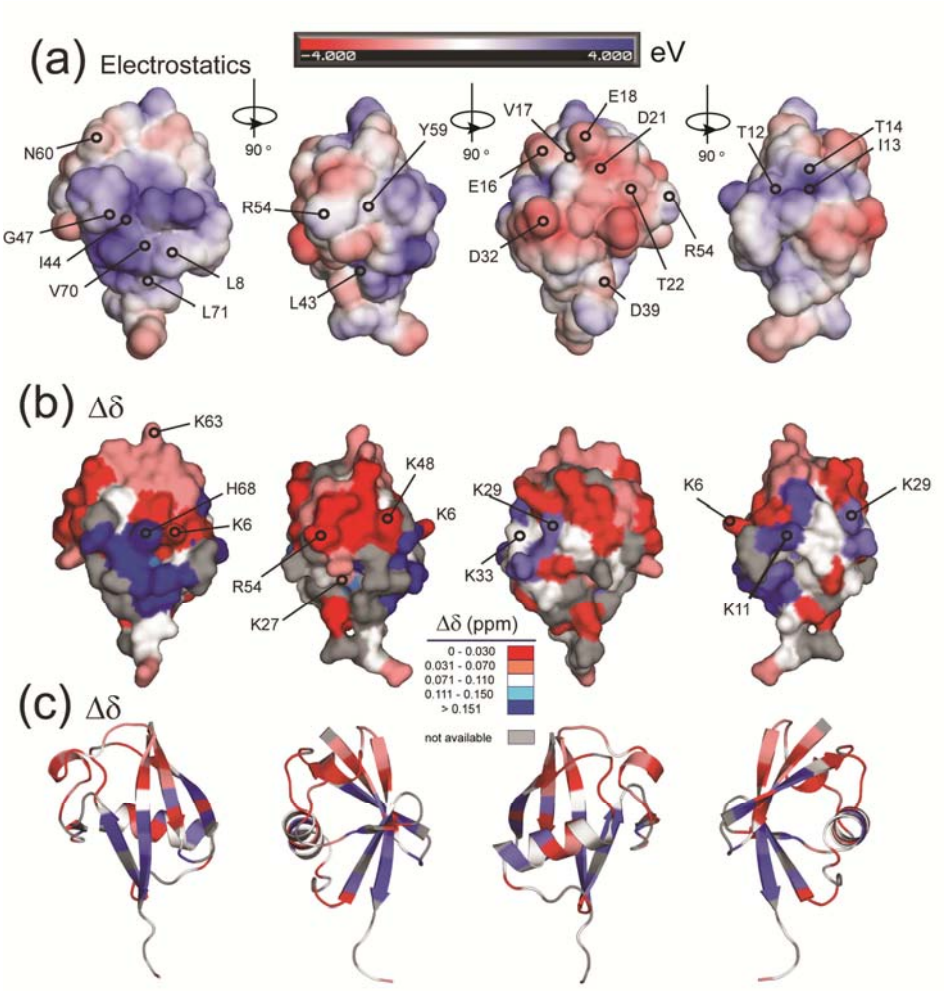
(a) Electrostatics
−4.000
4.000
eV
N60
G47
I44
V70
L8
L71
90 °
Y59
R54
L43
90 °
V17
E18
D21
E16
D32
R54
T22
D39
90 °
T12
T14
I13
(b) Δδ
K63
H68
K6
R54
K27
K48
K6
K29
K33
K6
K29
K11
Δδ (ppm)
0 - 0.030
0.031 - 0.070
0.071 - 0.110
0.111 - 0.150
> 0.151
not available
(c) Δδ

FIGURE 6

Figure 7. The hydrophobicity of an amino acid in UBI partly explains its interaction with SDS at 0-1 mM SDS. Average chemical shift perturbation for amino acids in ubiquitin at 0-1 mM SDS ($\Delta\delta$) plotted as a function of the hydrophobicity of the amino acid. Only amino acids that were included at > 3 positions in the amino acid sequence of ubiquitin were included in this set of data (arginine is not included because HSQC data is available for only Arg54). The values for the hydrophobicity of each amino acid were determined by Fauchere and Plisko (44) (more positive values are more hydrophobic). For each plot, m = slope and R^2 = correlation coefficient. The number of amino acid residues that are included in this set of data are: 9 leucine (L), 7 isoleucine (I), 7 lysine (K), 6 threonine (T), 5 aspartate (D), 5 glutamine (Q), 5 glutatmate (E), 4 glycine (G), and 4 valine (V) residues. This data set thus includes 52 of the 63 amino acids of UBI that are observable in HSQC experiments at both 0 and 1 mM SDS (e.g., 83% of the observable residues in UBI).

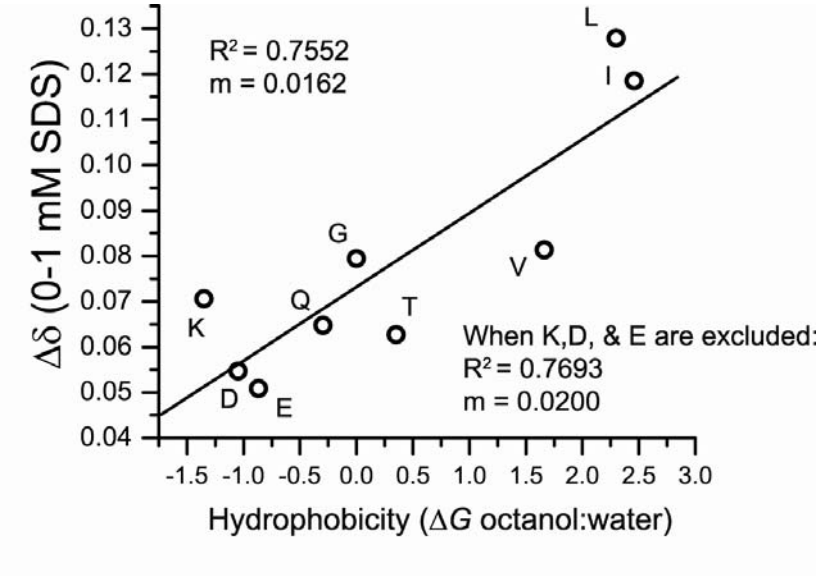
R² = 0.7552
m = 0.0162
When K,D, & E are excluded:
R² = 0.7693
m = 0.0200
Δδ (0-1 mM SDS)
Hydrophobicity (ΔG octanol:water)
0.13
0.12
0.11
0.10
0.09
0.08
0.07
0.06
0.05
0.04
-1.5 -1.0 -0.5 0.0 0.5 1.0 1.5 2.0 2.5 3.0
L
I
G
V
Q
T
K
D
E

FIGURE 7

Figure 8. Studying the binding of SDS to UBI at 1 mM SDS with ^{13}C-^{1}H HSQC NMR.

HSQC spectra of human ubiquitin (uniformly labeled with ^{13}C) in the presence of 0 and 1 mM SDS (pH 7.53, 25 mM Tris). The majority of ^{13}C-^{1}H signals are attenuated by the interaction with SDS. Several peaks remain and are unchanged in 1 mM SDS (i.e., ε-CH and γ-CH of several Lys, and δ-CH of Arg54 and Arg72) indicating no interaction with SDS. Several γ-CH and δ-CH of Ile and Leu remain, but are significantly shifted (denoted with curved arrows) indicating a specific interaction with SDS. Notably, the γ-CH, δ-CH, and ε-CH of Lys6 (a solvent exposed lysine residue) are not attenuated or shifted by the presence of 1 mM SDS. The resonance labeled with an asterisk (*) is likely the α-CH of Gly76 on the freely rotating C-terminus of UBI.

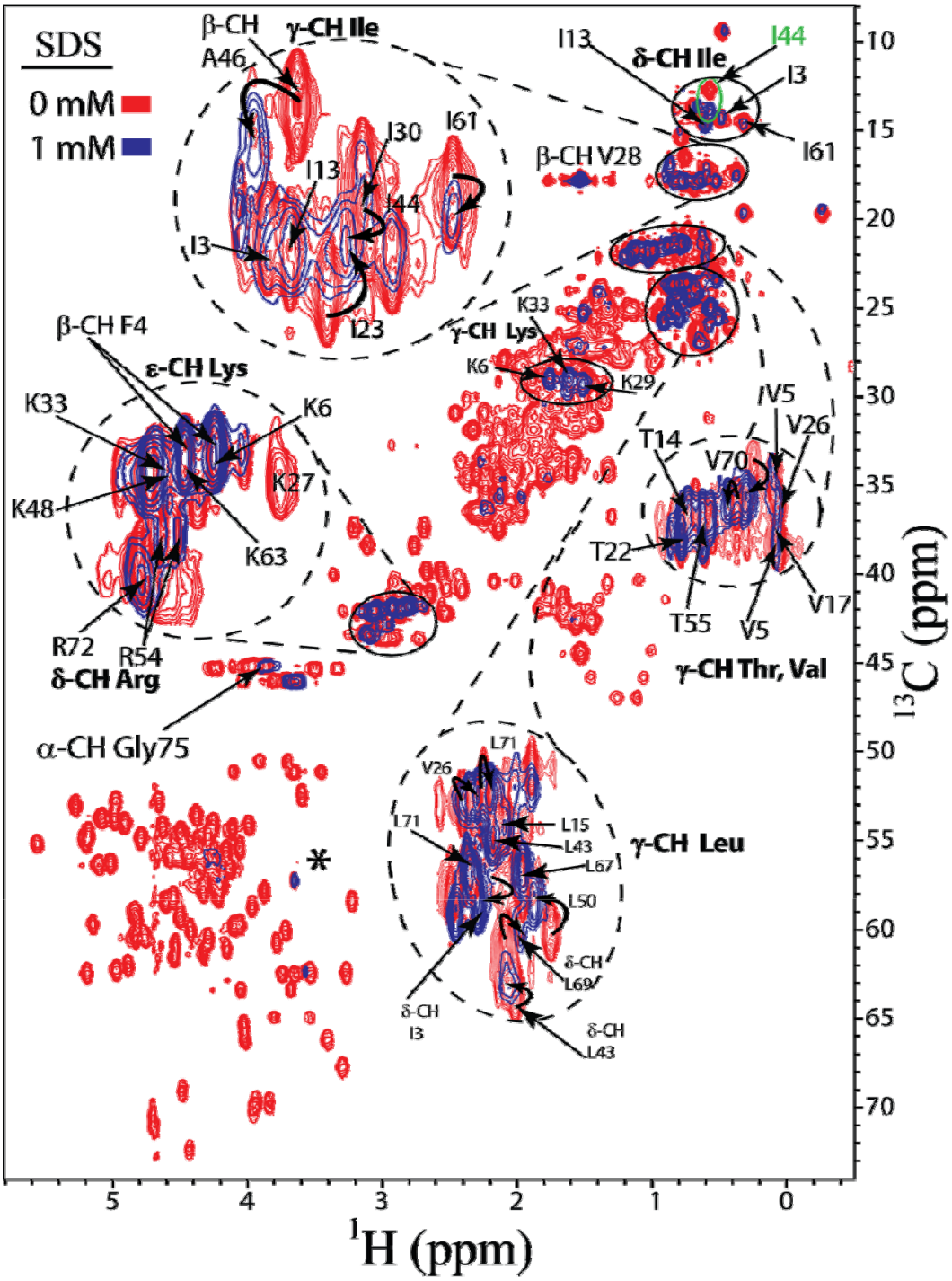
SDS
0 mM
1 mM
β-CH A46
γ-CH Ile
I30
I61
I13
I44
I3
I23
I13
δ-CH Ile
I44
I3
I61
β-CH V28
K33
γ-CH Lys
K6
K29
β-CH F4
ε-CH Lys
K33
K6
K27
K48
K63
R72
R54
δ-CH Arg
α-CH Gly75
V5
V26
T14
V70
T22
T55
V5
V17
γ-CH Thr, Val
L71
V26
L71
L15
L43
L67
γ-CH Leu
L50
δ-CH L69
δ-CH I3
δ-CH L43
10
15
20
25
30
35
40
45
50
55
60
65
70
^{13}C (ppm)
5
4
3
2
1
0
^{1}H (ppm)

FIGURE 8

Table 1. Biophysical properties of residues in UBI-$(NH_3^+)_8$ with the largest (top) and smallest (bottom) chemical shift perturbation (Δδ) in HSQC experiments at 0 mM and 1 mM SDS.

	eV_{Local}	2°	SA_{Main}	SA_{Side}	*HydroP*	$k_{H/D}$	***Δδ***
Largest Δδ							
Ile 13	+	β	0.0	3.7	2.46	0.46	0.31160
Leu 43	+	β	1.2	0.0	2.30	10.71	0.22049
Leu 71	+	β	63.7	39.5	2.30	0.52	0.21582
His 68	+	β	0.0	53.7	0.18	0.62	0.19558
Thr 14	+	β	17.6	52.8	0.35	0.63	0.18688
Lys 11	+	β	0.0	62.6	-1.35	2.83	0.17593
Leu 50	+	β	17.8	0.0	2.30	0.75	0.17105
Leu 8	+	∩	84.6	75.8	2.30	5.86	0.16858
Gly 47	+	∩	87.5	110.2	0.00	1.06	0.16164
Val 70	+	β	0.0	27.8	1.66	0.65	0.16162
Ile 44	+	β	0.0	23.9	2.46	0.11	0.15967
Asp 32	–	α	71.4	101.4	-1.05	0.40	0.14069
AVERAGE			**28.7**	**45.9**	**1.16**	**2.10**	**0.18912**
Smallest Δδ							
Glu 16	–	β	33.3	93.4	-0.87	N/A	0.02140
Val 17	–	∩	3.7	2.4	1.66	1.47	0.01468
Glu 18	–	∩	0.1	74.4	-0.87	0.68	0.01303
Asp 21	–	∩	0.0	29.5	-1.05	0.59	0.01244
Thr 22	–	∩	2.7	46.1	0.35	0.83	0.01174
Gln 40	0	β	0.0	30.6	-0.30	0.45	0.01130
Tyr 59	+/0	∩	22.5	16.3	1.31	0.41	0.01114
Thr 12	+	β	60.7	40.3	0.35	4.67	0.01109
Asn 60	–	∩	19.3	98.7	-0.79	0.80	0.01098
Asp 39	–	∩	11.5	77.6	-1.05	0.93	0.00778
Leu 56	N/A	∩	0.0	0.0	2.30	0.60	0.00640
Lys 6	+	β	2.5	58.9	-1.35	0.55	0.00180
AVERAGE			**13.0**	**47.4**	**-0.02**	**1.1**	**0.01115**

Column Legend: $\mathbf{eV_{Local}}$ refers to electrostatic potential in Figure 6; **2°** refers to α-helical or β-sheet secondary structure; ∩ denotes loop structure; **SA** refers to solvent accessibility of side chain or main chain, as approximated from the X-ray crystal structure; **HydroP** refers to hydrophobicity of each residue.(44) $\mathbf{k_{H/D}}$ is the rate of amide hydrogen-deuterium exchange (s^{-1}).(48) The change in chemical shift of each residue in the HSQC spectra at 0-1 mM SDS are

derived from data in Figure 4 and 5 and are denoted **Δδ.** A complete list of these values for all residues in UBI is included in the supplemental material.

References.

1. **Bodner CR, Dobson CM, & Bax A (2009) Multiple tight phospholipid-binding modes of alpha-synuclein revealed by solution NMR spectroscopy. *J Mol Biol* 390(4):775-790.**
2. **Vamvaca K, Volles MJ, & Lansbury PT, Jr. (2009) The first N-terminal amino acids of alpha-synuclein are essential for alpha-helical structure formation in vitro and membrane binding in yeast. *J Mol Biol* 389(2):413-424.**
3. **Bodner CR, Maltsev AS, Dobson CM, & Bax A (Differential phospholipid binding of alpha-synuclein variants implicated in Parkinson's disease revealed by solution NMR spectroscopy. *Biochemistry* 49(5):862-871.**
4. **Schneider GF, Shaw BF, Lee A, Carrilho E, & Whitesides GM (2008) Pathway for unfolding of ubiquitin in SDS, studied by capillary electrophoresis. *J. Am. Chem. Soc.* 130:17384-17393.**
5. **Gudiksen KL, Gitlin I, & Whitesides GM (2006) Differentiation of proteins based on characteristic patterns of association and denaturation in solutions of SDS. *Proc. Natl. Acad. Sci. U. S. A.* 103(21):7968-7972.**
6. **This point is supported by the absence of clefts or pockets on the surface of folded UBI, and the observation that no pockets appear to form in UBI upon its association with < 1.4 mM SDS (e.g., previous work has shown that the secondary and tertiary structure of UBI does not change upon the binding of the first 11 SDS.**
7. **Jackson SE (2006) Ubiquitin: a small protein folding paradigm. *Org. Biomol. Chem.* 4(10):1845-1853.**
8. **Loladze VV & GI (2002) Removal of surface charge-charge interactions from ubiquitin leaves the protein folded and very stable. *Protein Sci.* 11(1):174-177.**
9. **Gitlin I, Gudiksen KL, & Whitesides GM (2006) Peracetylated bovine carbonic anhydrase (BCA-Ac-18) is kinetically more stable than native BCA to sodium dodecyl sulfate. *J. Phys. Chem. B* 110(5):2372-2377.**
10. **Shaw BF*, et al.* (2008) Lysine acetylation can generate highly charged enzymes with increased resistance towards irreversible inactivation. *Protein Sci.* 17(8):1446-1455.**
11. **Colton IJ*, et al.* (1997) Formation of protein charge ladders by acylation of amino groups on proteins. *J. Am. Chem. Soc.* 119(52):12701-12709.**
12. **McCoy MA & Wyss DF (2002) Spatial localization of ligand binding sites from electron current density surfaces calculated from NMR chemical shift perturbations. *J Am Chem Soc* 124(39):11758-11763.**
13. **Konuma T, Sakurai K, & Goto Y (2007) Promiscuous binding of ligands by beta-lactoglobulin involves hydrophobic interactions and plasticity. *J Mol Biol* 368(1):209-218.**
14. **Van Horn WD, Ogilvie ME, & Flynn PF (2009) Reverse micelle encapsulation as a model for intracellular crowding. *J Am Chem Soc* 131(23):8030-8039.**

15. **Lu RC, Guo XR, Jin C, & Xiao JX (2009) NMR studies on binding sites and aggregation-disassociation of fluorinated surfactant sodium perfluorooctanoate on protein ubiquitin. *Biochim Biophys Acta* 1790(2):134-140.**
16. **Lima LM, *et al.* (2006) Structural insights into the interaction between prion protein and nucleic acid. *Biochemistry* 45(30):9180-9187.**
17. **Dupureur CM (2005) NMR studies of restriction enzyme-DNA interactions: role of conformation in sequence specificity. *Biochemistry* 44(13):5065-5074.**
18. **Ozen C, Norris AL, Land ML, Tjioe E, & Serpersu EH (2008) Detection of specific solvent rearrangement regions of an enzyme: NMR and ITC studies with aminoglycoside phosphotransferase(3')-IIIa. *Biochemistry* 47(1):40-49.**
19. **Stark J & Powers R (2008) Rapid protein-ligand costructures using chemical shift perturbations. *J Am Chem Soc* 130(2):535-545.**
20. **Simorellis AK & Flynn PF (2006) Fast local backbone dynamics of encapsulated ubiquitin. *J Am Chem Soc* 128(30):9580-9581.**
21. **Fernandez C & Wuthrich K (2003) NMR solution structure determination of membrane proteins reconstituted in detergent micelles. *FEBS Lett* 555(1):144-150.**
22. **Debye P & Huckel E (1923) The theory of the electrolyte II - The border law for electrical conductivity. (Translated from German) *Physikalische Zeitschrift* 24:305-325 (in German).**
23. **Feig M & Brooks CL, 3rd (2004) Recent advances in the development and application of implicit solvent models in biomolecule simulations. *Curr Opin Struct Biol* 14(2):217-224.**
24. **Honig B & Nicholls A (1995) Classical electrostatics in biology and chemistry. *Science* 268(5214):1144-1149.**
25. **Tsuchiya Y, Kinoshita K, & Nakamura H (2005) PreDs: a server for predicting dsDNA-binding site on protein molecular surfaces. *Bioinformatics* 21(8):1721-1723.**
26. **Tsuchiya Y, Kinoshita K, & Nakamura H (2004) Structure-based prediction of DNA-binding sites on proteins using the empirical preference of electrostatic potential and the shape of molecular surfaces. *Proteins* 55(4):885-894.**
27. **Gelb MH, Cho W, & Wilton DC (1999) Interfacial binding of secreted phospholipases A(2): more than electrostatics and a major role for tryptophan. *Curr Opin Struct Biol* 9(4):428-432.**
28. **Warshel A, Sharma PK, Kato M, & Parson WW (2006) Modeling electrostatic effects in proteins. *Biochim Biophys Acta* 1764(11):1647-1676.**
29. **Gitlin I, Carbeck JD, & Whitesides GM (2006) Why are proteins charged? Networks of charge-charge interactions in proteins measured by charge ladders and capillary electrophoresis. *Angew. Chem., Int. Ed.* 45(19):3022-3060.**
30. **Silow M, Tan YJ, Fersht AR, & Oliveberg M (1999) Formation of short-lived protein aggregates directly from the coil in two-state folding. *Biochemistry* 38(40):13006-13012.**
31. **Went HM & Jackson SE (2005) Ubiquitin folds through a highly polarized transition state. *Protein Eng Des Sel* 18(5):229-237.**
32. **Hansch C, *et al.* (1973) Aromatic Substituent Constants for Structure-Activity Correlations. *J. Med. Chem.* 16(11):1207-1216.**
33. **The Hansch hydrophobic parameter for SO_4^- is -4.76.**
34. **This rationalization assumes that the eleven SDS molecules bound to UBI-$(NH_3^+)_8$ in its G_2 state (e.g., UBI-$(NH_3^+)_8(SDS)_{11}$) render the surface of the protein more**

hydrophobic than the $NHCOCH_3$ groups introduced onto the surface of UBI-$(NH_3^+)_8$ by peracetylation (the binding of SDS also renders the surface more zwitterionic). The increase in surface hydrophobicity due to the acetylation of lysines in UBI-$(NHAc)_8$—that is, the addition of eight -$COCH_3$ groups—is certainly small compared to the increase in surface hydrophobicity reflected by the association of ~130 methylene units resulting from the binding of eleven molecules of SDS.

35. **Grossman PD (1992) *Capillary Electrophoresis: Theory and Practice* (Academic Press, San Diego, CA).**
36. **Basak SK & Ladisch MR (1995) Correlation of Electrophoretic Mobilities of Proteins and Peptides with Their Physicochemical Properties. *Anal. Biochem.* 226(1):51-58.**
37. **Wang C, Xi J, Begley TP, & Nicholson LK (2001) Solution structure of ThiS and implications for the evolutionary roots of ubiquitin. *Nat Struct Biol* 8(1):47-51.**
38. **The positively charged side of UBI is composed of an entire face of a β-sheet, and is hydrophobic. The negatively charged side contains a loop structure, and also contains the single α-helix; this α-helix contains 3 cationic residues (all lysine) but also contains 3 anionic residues and is close to being overall electrically neutral.**
39. **Wang Q, Goh AM, Howley PM, & Walters KJ (2003) Ubiquitin recognition by the DNA repair protein hHR23a. *Biochemistry* 42(46):13529-13535.**
40. **Fisher RD*, et al.* (2003) Structure and ubiquitin binding of the ubiquitin-interacting motif. *J Biol Chem* 278(31):28976-28984.**
41. **Ponting CP (2000) Proteins of the endoplasmic-reticulum-associated degradation pathway: domain detection and function prediction. *Biochem J* 351 Pt 2:527-535.**
42. **Alam SL*, et al.* (2004) Ubiquitin interactions of NZF zinc fingers. *Embo J* 23(7):1411-1421.**
43. **Verma R*, et al.* (2004) Ubistatins inhibit proteasome-dependent degradation by binding the ubiquitin chain. (Translated from eng) *Science* 306(5693):117-120 (in eng).**
44. **Wimley WC, Creamer TP, & White SH (1996) Solvation energies of amino acid side chains and backbone in a family of host-guest pentapeptides. *Biochemistry* 35(16):5109-5124.**
45. **Michaux C, Pouyez J, Wouters J, & Prive GG (2008) Protecting role of cosolvents in protein denaturation by SDS: a structural study. *BMC Struct Biol* 8:29.**
46. **Manning M & Colon W (2004) Structural basis of protein kinetic stability: Resistance to sodium dodecyl sulfate suggests a central role for rigidity and a bias toward beta-sheet structure. *Biochemistry* 43(35):11248-11254.**
47. **Gill SC & von Hippel PH (1989) Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* 182:319-326.**
48. **Brand T*, et al.* (2007) Residue-specific NH exchange rates studied by NMR diffusion experiments. *J Magn Reson* 187(1):97-104.**