BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

Disentangling the Coil: Modulation of Conformational and Dynamic Properties by Site-Directed Mutation in the Non-Native State of Hen Egg White Lysozyme

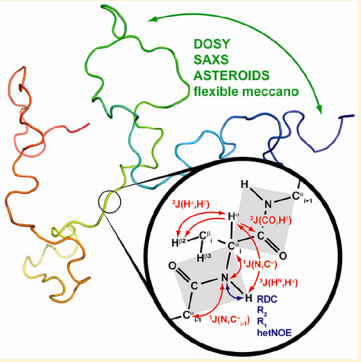
Friederike Sziegat,[†] Robert Silvers,[†] Martin Hähnke,[†] Malene Ringkjøbing Jensen,[‡] Martin Blackledge,[‡] Julia Wirmer-Bartoschek,*[,†] and Harald Schwalbe*[,†]

[†]Institute for Organic Chemistry and Chemical Biology, Center for Biomolecular Magnetic Resonance (BMRZ), Johann Wolfgang Goethe-University, Frankfurt a. M., Max-von-Laue-Straße 7, D-60438 Frankfurt, Germany

[‡]Protein Dynamics and Flexibility by NMR, Institut de Biologie Structurale Jean-Pierre Ebel, CEA-CNRS-UJF UMR 5075, 41 Rue Jules Horowitz, Grenoble 38027, France

Supporting Information

ABSTRACT: The conformational analysis of non-native states of proteins remains one of the most difficult problems in structural biology, because such states are represented by a superimposition of several states that are rapidly interconverting. Hence, model building of the conformational ensemble remains challenging, although many different biophysical observables can be determined in non-native states of proteins. Here, we present a comprehensive analysis of non-native states of wild-type and mutant forms of the model protein lysozyme by nuclear magnetic resonance spectroscopy. Relaxation rates, chemical shifts, backbone and side chain coupling constants, residual dipolar couplings, diffusion rate constants, and small-angle scattering data merged with computational approaches, such as *flexible meccano* and ASTEROIDS, allow the description of the non-native state of hen egg white lysozyme in unprecedented detail.



For decades, the biophysical investigation of unfolded states of proteins has significantly enhanced our understanding of protein folding and misfolding.[1] Among the repertoire of biophysical methods, multidimensional nuclear magnetic resonance (NMR) spectroscopy on uniformly ^{13}C- and ^{15}N-labeled samples in solution is particularly powerful in determining conformational and dynamical properties within ensembles representing the unfolded state of proteins.[2] Both global and residue-specific parameters can be measured to describe non-native states of proteins (Figure 1) over a broad range of time scales. To analyze the properties of the global ensemble, diffusion-ordered spectroscopy (DOSY) and small-angle X-ray scattering (SAXS) measurements can be performed, reporting on the hydration radius (R_h) and the radius of gyration (R_g), respectively. Residue-specific backbone flexibility and structural preferences can be deduced from heteronuclear ^{1}H−^{15}N relaxation rates and residual dipolar couplings (RDCs). Several scalar coupling constants can be used to obtain preferences for backbone and side chain torsions.

In this study, we are advancing the detection limits for residual nonrandom structure in non-native states of proteins combining ^{1}H and ^{13}C chemical shift deviations, an extensive set of coupling constants reporting on protein torsion angles ϕ, ψ, and χ_1, relaxation rate analysis, mutational analysis, residual dipolar couplings, ensemble compaction by DOSY and SAXS, and computational approaches.

One aspect of the study is further investigation of the local and global nonrandom structure in the model protein hen egg white lysozyme (HEWL). HEWL contains four disulfide bridges formed by eight cysteines. An unusually large number of tryptophans (W28, W62/63, W108, W111, and W123) can be found in lysozyme. In its disulfide-reduced state, the protein is unfolded even in the absence of denaturants.[3−5] For the investigation of the non-native states of HEWL, we replaced all cysteines with alanines following a similar strategy applied earlier for the investigation of unfolded states of α-lactalbumin.[6−8] After removal of the disulfide bridges by cysteine-to-alanine replacement (WT^{Ala}), the lysozyme polypeptide chain is unfolded even in the absence of denaturants as evidenced by the small chemical shift dispersion in the ^{1}H−^{15}N correlation spectra (Figures S1−S6 of the Supporting Information).

Structural and dynamic phenomena in non-native states of hen egg white lysozyme are linked to the presence of six hydrophobic clusters that correlate with the hydrophobicity of its polypeptide chain (Figure 2).

A second important aspect of the work is probing nonrandom structure and dynamics on various time scales. It is a unique aspect of NMR that differences in the time scales of averaging are directly observable in the experimental data and need to be taken into account for the prediction of random-coil NMR parameters. While ^{15}N relaxation rates are sensitive to

Received: February 17, 2012
Revised: March 30, 2012
Published: April 2, 2012

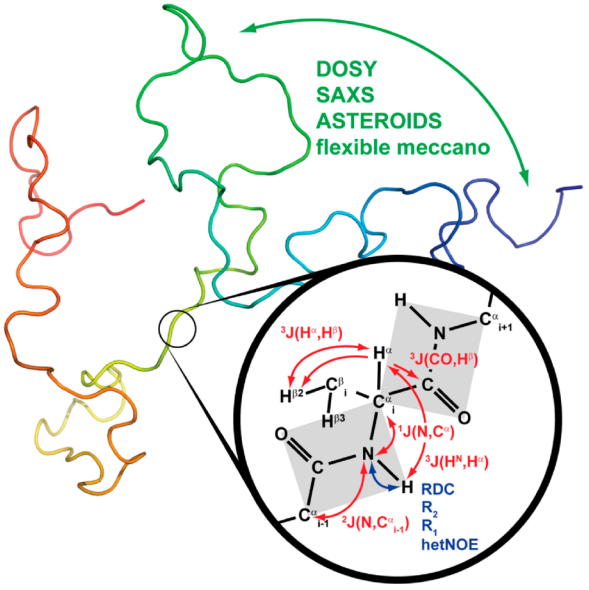


Figure 1. Schematic representation of data acquired for the non-native state of lysozyme. The unfolded state is represented by a random-coil structure that is rainbow color-coded for the sake of clarity. Global parameters are colored green; residue-specific coupling constants are colored red, and heteronuclear relaxation rates, hetNOEs, and RDCs are colored blue. Some atoms and bonds were omitted for the sake of clarity. Rigid peptide planes are colored gray.

motions faster than the overall tumbling time (τ_c) in the subnanosecond time regime, structure persisting on supra-τ_c time scales can be monitored by chemical shifts (δ), scalar coupling constants (J), and residual dipolar couplings (RDCs). Chemical shifts and J couplings reveal a unique dependence on local conformational properties and are thus sensitive indicators of nonrandom structure in unfolded polypeptide chains. Residual dipolar couplings are affected by local conformational properties and by the global reorientation of the protein relative to the alignment tensor. J couplings, chemical shifts, and RDCs are averaged over time ranges up to milliseconds.

Different approaches have been applied for the prediction of RDCs in unfolded protein states. Numerical random-flight chain models that treat the polypeptide chain as a homopolymer have been introduced; smaller values at the termini of the polymer and a leveling off in the center of the polypeptide chain are characteristics of such homopolymers.[9] Numerical models typically do not describe residue-specific modulations and do not take into account excluded volume effects.[10] Because the time scale for RDCs is much larger than for relaxation rates, the use of a model that treats the polymer as a heteropolymer is essential. Such an approach is often realized by the generation of conformational ensembles.[11,12] The *flexible meccano* algorithm generates statistical ensembles of coil conformations from databases of ϕ and ψ, and average sequence-specific values are calculated over these ensembles.[13] The detailed interpretation of deviations from these predicted values is still difficult; however, deviating RDCs have been detected and are interpreted in terms of residual structure, and correlations with conformational ensembles by molecular dynamics simulations can be obtained.[14−16] Utilizing such methodologies, residual structure and dynamics have been identified in a number of chemically or biochemically denatured proteins,[5,7,14−23] revealing that non-native states of proteins can vary widely in terms of the sampled conformations and dynamics. Only a few systems in which both relaxation rates and RDC data have been measured exist.[17−24] To the best of our knowledge, only for ACBP a joint analysis of relaxation data and RDC data has been published.

In this study, we cross-validate our NMR experimental findings with structural information derived from SAXS. From the joint analysis of the experimental data over six different single-point mutants of the protein, utilizing these two fundamentally different approaches, a decreasing degree of global compactness from WT^{Ala} to the most unstructured mutant, $W62G^{Ala}$, can be ascertained. Single-point mutations have substantial effects on local and long-range structure and dynamics in the unfolded states. They bias the ensemble supporting the notion of their impact on the rate of productive folding and unproductive misfolding as well as aggregation.[25]

MATERIALS AND METHODS

Cloning, Expression, and Purification of the Mutants. Gene sequences encoding the mutants $W28G^{Ala}$, $W62G^{Ala}$, $W108G^{Ala}$, $W111G^{Ala}$, and $W123G^{Ala}$ were generated by site-directed mutagenesis of the WT^{Ala} gene sequence[26] using the QuickChange Site-Directed Mutagenesis Kit (Stratagene). All genes exist as part of a pET11a expression plasmid. Gene sequences were verified by DNA sequencing (Scientific

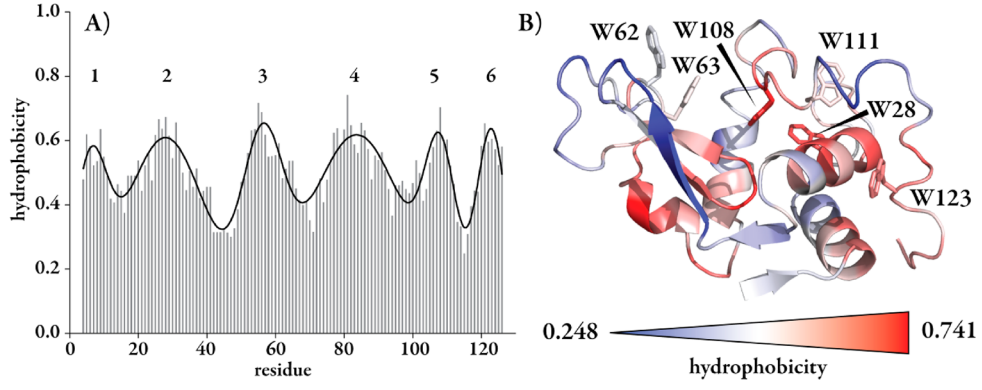


Figure 2. Hydrophobicity of HEWL. (A) Distribution of hydrophobicity according to the approach of Leo and Abraham[62] with a window size of 7. Data are normalized from 0 to 1 and fit by Gaussian clusters. The position of hydrophobic clusters is indicated. (B) Hydrophobic scale values mapped onto the structure of HEWL. Tryptophan residues are shown and indicated. Structures were generated using PyMOL (Protein Data Bank entry 6LYZ).

dx.doi.org/10.1021/bi300222f | *Biochemistry* 2012, 51, 3361−3372

Research and Development GmbH, Oberursel, Germany). ^{15}N-labeled protein was expressed in *Escherichia coli* strain BL21(DE3) (Novagen). Cells were grown in M9 minimal medium with ^{15}N-labeled ammonium chloride [$^{15}NH_4Cl$ (Martek)] as the sole nitrogen source. At an OD_{600} of 0.8, induction of protein expression was initiated by addition of 1 mM IPTG. After being induced for 3 h, the cells were harvested by centrifugation (6500*g* for 20 min). All HEWL mutants were produced by inclusion body purification: The cells were resuspended in 50 mM Tris-HCl, 25% sucrose, and 1 mM EDTA (pH 7.5), and cells were disrupted by sonication. After centrifugation (10000*g* for 30 min), the supernatant was discarded and the pellet was resuspended in 20 mM Tris-HCl, 1% Triton X-100, and 1 mM EDTA (pH 7.5). After centrifugation (10000*g* for 30 min), the supernatant was discarded and the pellet, which contains the inclusion bodies, was dissolved in 20 mM Tris-HCl, 50 mM NaCl, 5 mM EDTA, and 8 M urea (pH 7.5). After centrifugation (10000*g* for 30 min), the supernatant was loaded onto a CM-Sepharose (Sigma-Aldrich) ion exchange column, which was equilibrated with 50 mM Tris-HCl, 50 mM NaCl, 1 mM EDTA, and 4 M urea (pH 7.5). The protein was eluted with a linear gradient between buffer A [50 mM Tris-HCl, 50 mM NaCl, 1 mM EDTA, and 4 M urea (pH 7.5)] and buffer B [50 mM Tris-HCl, 300 mM NaCl, 1 mM EDTA, and 4 M urea (pH 7.5)]. Fractions with matching bands on a sodium dodecyl sulfate−polyacrylamide gel electrophoresis gel were pooled and dialyzed against water (pH 2) for several days (tubular membrane 3500 molecular weight cutoff, ZelluTrans, Roth). After purification via HPLC with a reversed phase Vydac Grace column, TFA buffer, and a linear water/acetonitrile gradient, the protein was freeze-dried. For NMR measurements, the protein was dissolved in sterile H_2O with 10% (v/v) D_2O at pH 2. All sample concentrations were 300 μM, determined by UV light absorption at 280 nm using the extinction coefficient calculated from the amino acid sequence (calculated via ExPASy, Swiss Institute of Bioinfomatics); for WT^{Ala}, the extinction coefficient (ε) is 37470 M^{-1} cm^{-1}, and for single tryptophan-to-glycine mutants, the extinction coefficient (ε) is 31970 M^{-1} cm^{-1}.

Backbone Assignment and Measurement of ^{15}N Relaxation Rates. The backbone assignment of all mutants was facilitated by the existing chemical shifts for wild-type WT^{Ala}. Therefore, measurements of ^{15}N HSQC[27] together with three-dimensional (3D) ^{15}N NOESY-HSQC[28,29] were sufficient. The 3D NOESY-HSQC spectrum was recorded with 2048 points in the ^{1}H dimension, 128 points in the ^{15}N dimension, and eight scans. Typically, the following maximal evolution times were deployed: t_2^{max} = 122 ms, and t_1^{max} = 71.25 ms. The mixing time (t_m) was 175 ms. All NMR relaxation experiments were conducted under identical conditions: 600 MHz (Bruker, DRX 600, probe, 5 mm TXI HCN xyz-gradient), 293 K, ^{15}N-labeled sample concentration of 300 μM (H_2O/10% D_2O, pH 2). NMR data were processed using TopSpin 2.1 (Bruker Biospin) and analyzed with Sparky (SPARKY 3, University of California, San Francisco, CA) and CARA[30] (http://www.nmr.ch). For the determination of R_2 rates, the ^{15}N CPMG-HSQC spectrum was recorded as a pseudo-three-dimensional spectrum with nine increments ranging from 17 to 306 ms, including three repeats in an interleaved manner. For $R_{1\rho}$ rates, 10 increments were measured ranging from 20 to 250 ms with two repeats. For the R_1 rates, eight different relaxation times have been used, ranging from 0.01 to 2 s, with two repeats. The SPARKY routine was used to extract relaxation rates from peak heights. R_2 relaxation rate profiles were fit by

$$R_2^{exp}(i) = R_{int}\sum_{j=1}^{N} e^{|i-j|/\lambda_0} + \sum_{cluster} R_{cluster} e^{-\left(\frac{i-x_{cluster}}{2\lambda_{cluster}}\right)^2} \quad (1)$$

where R_{int} is the intrinsic relaxation rate, which depends also on temperature and viscosity, λ_0 is the persistence length of the polypeptide chain (in terms of numbers and residues), N is the total chain length, $x_{cluster}$ is the position of the cluster, $\lambda_{cluster}$ is the cluster width, and $R_{cluster}$ is the rate. The first term characterizes the baseline, whereas the second term fits clusters yielding the deviation from the baseline relaxation profile.

CPMG relaxation experiments[31] yielded pseudo-three-dimensional spectra that were analyzed with TOPSPIN 2.0. A set of 16 experiments were performed with different CPMG field strengths ranging from 25 to 750 Hz, including three repeats. All spectra were recorded at 293 K with 128 real points and a sweep width of 22 ppm in the ^{15}N dimension at a protein concentration of 300 μM. For data analysis, the resulting effective relaxation rates (R_2^{eff}) were plotted against the CPMG field strength.

Measurements of $^1J(N_i,C^\alpha_i)$, $^2J(N_i,C^\alpha_{i-1})$, $^3J(H^N,H^\alpha)$, $^3J(H^\alpha,H^\beta)$, and $^3J(CO,H^\beta)$. All NMR experiments for *J*-modulated spectra were conducted on a 600 MHz spectrometer (Bruker, Avance II). The spectra were recorded at 293 K at a protein concentration of 300 μM (in H_2O/10% D_2O, pH 2). For the determination of $^1J(N_iC^\alpha_i)$ and $^2J(N_iC^\alpha_{i-1})$ coupling constants, a series of *J*-modulated ^{1}H−^{15}N HSQC spectra were measured with 15 increments (τ = 1−300 ms). To extract peak intensities, we used the SPARKY routine. The intensities were fitted with

$$I_{NC^\alpha}^{exp} = A\cos(\pi^1 J\tau)\cos(\pi^2 J\tau) \times e^{-\tau/T_2} \quad (2)$$

where I^{exp} values are the experimental peak heights (Figure S16 of the Supporting Information), A is a fitting factor, ^1J and ^2J are the couplings from N_i to C^α_i and $C^\alpha_{(i-1)}$, respectively,[32] T_2 is the transverse relaxation rate, and τ is the mixing time. For the determination of $^3J(H^N,H^\alpha)$ and $^3J(H^\alpha,H^\beta)$ coupling constants, a 3D ^{15}N HNHA spectrum and a 3D HN(COCA)-HAHB spectrum were recorded, respectively. The magnitude from relative peak intensities was calculated with[33]

$$\frac{I_{cross}}{I_{diag}} = -\tan^2(2\pi J_{HH}\tau) \quad (3)$$

where I_{diag} and I_{cross} are the intensities of diagonal peaks and cross-peaks, respectively, and τ is the J evolution delay. This analysis can be directly transferred for the $^3J(CO,H^\beta)$ couplings. They were measured with a 3D ^{15}N−^{13}C HN(CO)HB experiment[34] in which diagonal peaks represent the ^{13}C′ peak and the cross-peak represents the coupled H^β.

RDC Measurements and Simulations. The proteins were aligned in stretched 7% polyacrylamide gels (acrylamide:bisacrylamide ratio of 37.5:1)[35,36] at a concentration of 300 μM (in H_2O/10% D_2O, pH 2) to compare them to the isotropic samples in solution. The equipment for the preparation of the anisotropic samples was purchased from Newera Enterprises Inc. The HDO quadrupolar splitting was measured from the quadrupolar splitting that could be observed in the ^{2}H one-dimensional spectrum. RDCs were measured on a 600 MHz spectrometer (Bruker, DRX 600, probe head, 5 mm TXI HCN xyz-gradient) via the standard Bruker implementation of a

dx.doi.org/10.1021/bi300222f | *Biochemistry* 2012, 51, 3361−3372

two-dimensional (2D) ^{15}N IPAP-HSQC[37] extracting the sum of $^{1}J(H^{N},N)$ and $^{1}D(H^{N},N)$. $^{1}J(H^{N},N)$ values were measured using the same experiments for unaligned samples. For RDC simulation, 500000 structures were calculated for the wild type WT^{Ala} and each mutant using the *flexible meccano* approach as described previously.[13] RDCs were calculated from these ensembles using PALES.[38] For WT^{Ala}, the radius of gyration and the radius of gyration for a subset of 10000 structures were calculated using HYDRPRO.[39]

Chemical Shift Analysis and Secondary Structure Propensity. Secondary structure propensities (SSPs) were calculated from $^{1}H^{\alpha}$, $^{13}C^{\alpha}$, and $^{13}C^{\beta}$ chemical shifts as described previously.[40] A seven-residue weighted averaging was used. Chemical shift deviations from random-coil values (δ^{rc}) were calculated by $\Delta\delta = \delta^{exp} - \delta^{rc}$ with cutoffs for chemical shift indexing as described previously.[41,42] The values for $\Delta\delta H^{\alpha}$ and $\Delta\delta H^{N}$ were rounded using value 0 ranging from −0.049 to 0.049, value 0.1 ranging from 0.05 to 0.149, value 0.2 ranging from 0.15 to 0.249, etc. $\Delta^{3}J(H^{N},H^{\alpha})$, $\Delta^{2}J(N,C^{\alpha})$, and $\Delta^{1}J(N,C^{\alpha})$ were determined according to $\Delta J = J^{exp} - J^{rc}$. Random-coil coupling constants (J^{rc}) were calculated from a database of φ and ψ angles derived from random-coil regions of Protein Data Bank high-resolution structures as described previously.[11,12,43]

Diffusion-Ordered Spectroscopy (DOSY). Translational diffusion measurements[44] were recorded in D_2O at pH 2 with protein concentrations ranging from 300 to 100 μM. A 5-fold molar excess of dioxane was used as an internal standard. The standard Bruker pulse sequence for DOSY measurements was used where the lengths of all pulses were held constant while the strength of the diffusion gradient was varied between 5 and 95%. The spectra were recorded with 256 scans for 300 μM protein and up to 512 scans for 100 μM protein in 32 experiments. The one-dimensional spectra were optimized to give a signal decay of 85−90% and, thus, a stimulated echo of 150 ms; a diffusion gradient pulse of 6.0 ms was used. Data analysis was conducted using Topspin 2.0. Hydrodynamic radii were calculated by

$$R_h^{protein} = \frac{D_{dioxane}}{D_{protein}} R_h^{dioxane} \quad (4)$$

with a known dioxane radius $R_h^{dioxane}$ of 2.12 Å. $D_{dioxane}$ and $D_{protein}$ are the decay rates, and $R_h^{protein}$ is the hydrodynamic radius of the protein.[44]

Small-Angle X-ray Scattering (SAXS). The SAXS measurements were recorded on the ID14-3 BioSAXS beamline at the European Synchrotron Radiation Facility (ESRF, Grenoble, France). The sample−detector distance was 2.6 m; an X-ray wavelength of 0.931 Å was used (13.32 keV). Fifty microliters of each protein solution (as well as corresponding buffers) was loaded in a flow-through quartz capillary cell at 20 °C. The total exposure time was 100 s per sample. The 2D diffraction patterns were normalized to an absolute scale and azimuthally averaged to obtain the intensity profiles $I(q)$, within BSxCuBE (ESRF beamline data collection software). Solvent contributions (buffer backgrounds collected before and after every protein sample) were averaged and subtracted from the associated protein sample using PRIMUS.[45] Small-angle X-ray scattering was performed using protein concentrations from 300 to 100 μM in water at pH 2. For data analysis, two fitting routines were used, EOM (ensemble optimization method)[46] and the Porod-Debye relation.[47,48] EOM is based on averaging the individual scattering from every conformer in the conformational ensemble. Thereby, a pool with random conformers is generated, and only those were selected fitting the experimental data. The intensity of a set of conformations $I(s)$ with N conformers is given by

$$I(s) = \frac{1}{N}\sum_{n=1}^{N} I_n(s) \quad (5)$$

ASTEROIDS Calculations. *Flexible meccano* calculations were performed as described previously.[13] The ASTEROIDS calculations included two iteration steps where φ and ψ distributions of individual residues were obtained followed by a selection of entire structures.[49] In the iteration steps, 5 × 200 φ and ψ angles were selected for each residue on the basis of C^{α}, C^{β}, and C′ chemical shifts as well as the scalar couplings $^{3}J(H^{N},H^{\alpha})$, $^{1}J(N,C^{\alpha}_{i})$, and $^{2}J(N,C^{\alpha}_{i-1})$. Karplus parametrizations for the scalar couplings were taken from Pardi et al.[50] and Wirmer et al.[32] The chemical shifts of aspartic acids, glutamic acids, and histidines were not included in the calculations because of the influence of the protonation states on the chemical shifts at low pH. The final selection of entire structures was done on the basis of chemical shifts, scalar couplings, and $^{1}H-^{15}N$ RDCs from a pool of structures obtained by combining conformers from both iteration steps (15K conformers in total). RDCs were calculated for each structure in the pool using PALES[38] employing the local alignment window (LAW) approach with a window size of 15 amino acids.[51,52] Seven dummy residues (alanines) were added at each end of the lysozyme sequence to allow calculation of the RDCs for all residues. A baseline corresponding to the length of lysozyme was imposed on the RDCs as described previously.[53] An ensemble of 200 structures was selected on the basis of C^{α}, C^{β}, and C′ chemical shifts (0.1 ppm), $^{3}J(H^{N},H^{\alpha})$ (0.3 Hz), $^{1}J(N,C^{\alpha}_{i})$ (0.15 Hz), and $^{2}J(N,C^{\alpha}_{i-1})$ (0.15 Hz) scalar couplings, and $^{1}H-^{15}N$ RDCs (0.5 Hz), the numbers in brackets indicating the uncertainty in the experimental values used in the ASTEROIDS selection. For regional analysis, the five clusters [1 (residues 7−13), 2 (residues 28−34), 3(N) (residues 53−59), 3(C) (residues 60−66), 5 (residues 107−113), and 6 (residues 120−126)] were analyzed in terms of mean differences and evaluated by −−, −, 0, +, and ++, depending on the size of the deviation. Cluster 3 was split into N-terminal [3(N)] and C-terminal [3(C)] portions.

RESULTS AND DISCUSSION

Backbone Dynamics of WT^{Ala} As Evidenced by ^{15}N Relaxation. To characterize the non-native states of lysozyme in terms of dynamics and nonrandom structure, a comprehensive analysis of dynamics at various time scales is required. We first measured and analyzed heteronuclear ($^{1}H-^{15}N$) R_1 and R_2 relaxation rates, heteronuclear NOEs (hetNOEs) (panels A−C, respectively, of Figure 3), and relaxation dispersion rates (Figure S7 of the Supporting Information) to determine dynamics on the subnanosecond time scale and the chemical exchange contribution to R_2 relaxation rates. R_1 and R_2 relaxation rates and hetNOE data all showed similar sequence dependence: low values were found at the termini of the protein, while the central portion formed a plateau value from which deviations toward higher values are observed. The most pronounced deviations were found around hydrophobic residues, in cluster 3 around W62 and W63 with a shoulder toward I56 and also in cluster 5 around W108 and W111.

dx.doi.org/10.1021/bi300222f | *Biochemistry* 2012, 51, 3361−3372

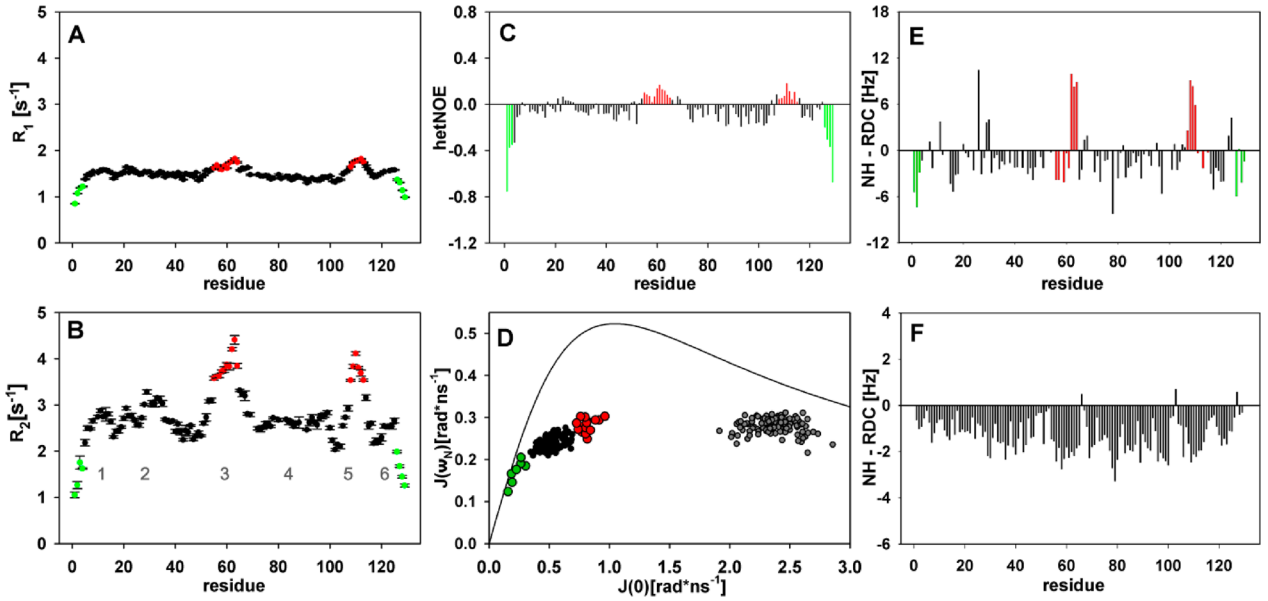


Figure 3. Relaxation and RDC data of WT^{Ala}: (A) ^{15}N R_1 relaxation rates, (B) ^{15}N R_2 relaxation rates, (C) $^{1}H-^{15}N$ heteronuclear NOEs, (D) spectral density mapping, (E) experimental NH RDCs, and (F) predicted NH RDCs from the *flexible meccano* approach. Color coding in panels A–E: red for amino acids with high rates (rigid regions) and green for amino acids with low rates at the N- and C-termini (flexible regions). The black line in panel D represents the spectral density function (dynamics limited to a single motion for which $J(\omega) = J(0)/\{1 + 6.25[\omega J(0)]^2\}$),[55] and gray circles indicate data of native hen egg white lysozyme for comparison. All experiments were performed at a field strength corresponding to a ^{1}H frequency of 600 MHz, using 300 μM ^{15}N-labeled protein at pH 2 and 293 K.

In both clusters, high R_1 relaxation rates and positive hetNOE values were also observed. Large deviations in R_2 relaxation rates and hetNOE values were also found in cluster 2 from residue 24 to 36 around W28. Medium-sized deviations were observed in cluster 1 around A10 and in cluster 6 around W123.

In folded states of proteins, the dynamics of a given N–H or C–H bond can be deduced from analysis of relaxation data using the Lipari–Szabo model because overall and internal motions can be separated.[54] Such an approach is not feasible for unfolded proteins where overall and internal motions occur on similar time scales. However, spectral densities can be extracted,[55] and such analysis (Figure 3D) revealed three different regimes of motion found in unfolded hen lysozyme: (a) very flexible residues at the termini of the protein where motions are governed by internal motions (colored green), (b) residues with substantial contributions of the overall motions to relaxation found in the central part of the protein (colored red), and (c) residues in which motions are determined by both internal and overall motions (colored black).

Heteronuclear relaxation rates reflect both the time scale and spatial restriction of stochastic motion of the active relaxation mechanism of the NH bond vector in addition to fluctuations of the ^{15}N chemical shift anisotropy. While the relaxation behavior of the terminal residues and the majority of the residues of the protein (black) could be predicted on the basis of homopolymer models for chain dynamics, some residues exhibited more rigidity than expected for an unfolded polypeptide (red) and had R_2 values much higher than those of the remainder of the protein. From relaxation dispersion experiments (Figure S7 of the Supporting Information), we found no evidence that these higher rates were caused by slow chemical exchange phenomena. Therefore, we concluded that these values resulted from even more restricted local sampling and report on enhanced conformational order at these sites. In addition, local conformational order at the centers of hydrophobic clusters was transmitted through the polypeptide chain, leading to Gaussian-type clustering of relaxation rates.

Backbone Structure and Dynamics of WT^{Ala} from ^{15}N RDCs. Relaxation rates report only on motional tumbling that is faster than the global correlation time τ_c of the protein, while RDCs are influenced by longer time scale motions as well as local conformational preferences. A comparison of relaxation rates and RDCs therefore allowed us to determine whether motional restriction persists on a time scale longer than τ_c. Experimental RDCs as a function of the protein sequence are shown in Figure 3E. Positive values were found in the regions of hydrophobic clustering (clusters 1–3, 5, and 6), and positive RDC values were observed, indicating a significant presence of structure in these clusters. Interestingly, while RDC values were positive in the C-terminal part of cluster 3, RDCs in the N-terminal part were negative. Positive RDCs indicate turnlike or helical structures, whereas negative RDCs could arise either from extended structures or from disorder. The heterogeneity of cluster 3 showed that high relaxation rates can be caused by different structural preferences of the polypeptide chain. Residual structure on the N-terminal part of cluster 3 thus either was extended or did not persist on the millisecond time scale, while dynamical restrictions on the nanosecond time scale were observed. Residual structure detected by relaxation data in at least five clusters of the protein was also observed in the RDCs; residual structure therefore persists from the subnanosecond time regime to the millisecond time regime.

Predicted RDCs generated by the *flexible meccano*[13] approach are shown in Figure 3F. Predicted RDC values deviated significantly from the experimental data. Apparently, and in line with relaxation rates and chemical shift and J coupling data (vide infra), the non-native states of lysozyme significantly deviated from random-coil prediction.

dx.doi.org/10.1021/bi300222f | *Biochemistry* 2012, 51, 3361–3372

Backbone and Side Chain Residual Structure of WT^{Ala} from Chemical Shifts and Scalar Coupling Constants. To gain more insight into the nature of local polypeptide chain rigidity, chemical shifts and coupling constants were utilized. Secondary structure propensities (SSPs) obtained from C^{α} and C^{β} chemical shifts,[40] $^{1}H^{N}$ and $^{1}H^{\alpha}$ chemical shift differences from random coil, and $^{3}J(H^{N},H^{\alpha})$, $^{2}J(N,C^{\alpha})$, and $^{1}J(N,C^{\alpha})$ coupling constants are shown in Figure 4. The pattern of WT^{Ala}

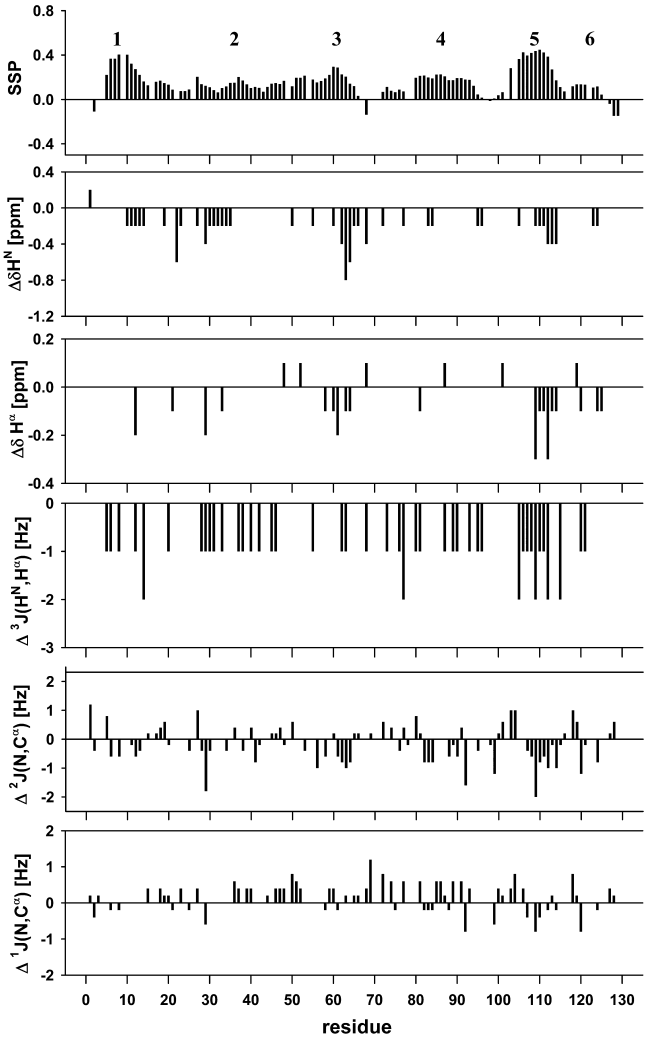


Figure 4. Residual secondary structure in WT^{Ala}. Secondary structure propensities (SSPs), $\Delta\delta H^{N}$, $\Delta\delta H^{\alpha}$, $\Delta^{3}J(H^{N},H^{\alpha})$, $\Delta^{2}J(N,C^{\alpha})$, and $\Delta^{1}J(N,C^{\alpha})$ are denoted with black bars. The values for $\Delta\delta H^{\alpha}$ and $\Delta\delta H^{N}$ were rounded using value 0 ranging from −0.049 to 0.049, value 0.1 ranging from 0.05 to 0.149, value 0.2 ranging from 0.15 to 0.249, etc. $\Delta\delta$ was calculated by $\Delta\delta = \delta_{exp} - \delta_{rc}$. All experiments were performed at 600 MHz with 300 μM protein at pH 2.0 and 293 K.

mostly exhibited positive SSP values, with an average of 14.5%, indicating a slight propensity for sampling of α-helical space in the unfolded protein, which agreed remarkably well with the helical propensity of 14.4% determined previously for the S-methylated variant using CD measurements.[5] In regions of residues 5−12 (cluster 1), 24−36 (cluster 2), 53−67 (cluster 3), 79−96 (cluster 4), and 103−113 (cluster 5) and around residue 120 (cluster 6), helical structure propensities of >0.2 were detected.

Backbone deviations from random-coil structure and modulation of local properties have to be correlated with side chain conformational properties of the polypeptide chain. We therefore tried to assess this correlation by probing side chain conformational preferences from $^{3}J(CO,H^{\beta})$ and $^{3}J(H^{\alpha},H^{\beta})$ coupling constants that report on χ_1.[56] $^{3}J(H^{\alpha},H^{\beta})$ couplings and even more $^{3}J(CO,H^{\beta})$ couplings of WT^{Ala} were on average slightly lower than predicted random-coil coupling constants (Figure 5 and Figure S8 of the Supporting Information). For aspartate (D), deviations in both coupling constants were observable, presumably because of changes in the backbone side chain interactions caused by low pH. Residues F, H, and R exhibited deviations mainly in the $^{3}J(H^{\alpha},H^{\beta})$ couplings, whereas $^{3}J(CO,H^{\beta})$ couplings for residues S and T exhibited larger deviations. The measured coupling constants report on angular distributions of the three rotamers of side chain angle χ_1 and can be modeled by a Pachler-type analysis (Table 1).[57] The random-coil model developed by Schwalbe et al. predicted a preference for the $\chi = -60°$ rotamer for all amino acids but V, S, and T.[58] For long aliphatic amino acids (here E, R, Q, and I), the experimental preference was not as pronounced as predicted: while the random-coil model predicted populations of 56−65% for the $\chi_1 = -60°$ rotamer, we found averaged $\chi_1 = -60°$ populations of 45−48%. Residues of D and N showed large deviations from random-coil behavior: while the model predicted a very slight preference for the −60° conformer, the experimental data showed the 60° rotamer to be the most populated one.

As predicted in the model, the 180° rotamer was preferred in our experimental data for V. This preference was not as pronounced as predicted, an observation that differs from previous measurements of ubiquitin, protein G, and oxidized lysozyme in 8 M urea.[58,59] For amino acids containing hydroxyl groups, S and T, the random-coil model predicted a preference for the $\chi_1 = 60°$ rotamer of 46% for serine and 60% for threonine, most likely caused by polar interactions between the side chain and the backbone. In the data measured here, the preference was similar for the 60° angle, with 43% for S and 70% for T; however, the populations of the 180° and −60° conformers varied significantly. On average, significant deviations for serine and threonine residues were also observed in previous data[58,59] and thus might be a general feature at acidic pH. Values for aromatic residues differed around 20% for almost every rotamer from the random coil. For histidine (H), the −60° and 60° rotamers were equally populated; in the case of phenylalanine (F) and tyrosine (Y), the populations of all three rotamers were similar. Most drastically, in tryptophan (W108 and W63), the 180° rotamer was clearly preferred over the other two rotameric states. Similarly, deviations from the random-coil model were also observed for the tryptophan residues in urea-denatured oxidized HEWL,[58] but not for urea-denatured protein G and ubiquitin[59] with little residual structure.[25,60] Thus, the deviations observed here emphasize the importance of aromatic residues, in particular tryptophans, for the formation of nonrandom residual structure.

Modulation of Nonrandom Structure and Dynamics in Single-Point Mutations. Further insight into long-range interactions within the ensemble of unfolded conformations was obtained by characterization of nonconservative single-point mutations as previously reported for other variants of lysozyme.[7] Here, the following tryptophan-to-glycine mutants were analyzed: W28G^{Ala}, W62G^{Ala}, W108G^{Ala}, W111G^{Ala}, and W123G^{Ala}. The correlation of the H^{N} chemical shifts from WT^{Ala} and its mutants revealed significant differences in the H^{N} chemical shifts around the mutation site because of nearest

dx.doi.org/10.1021/bi300222f | Biochemistry 2012, 51, 3361−3372

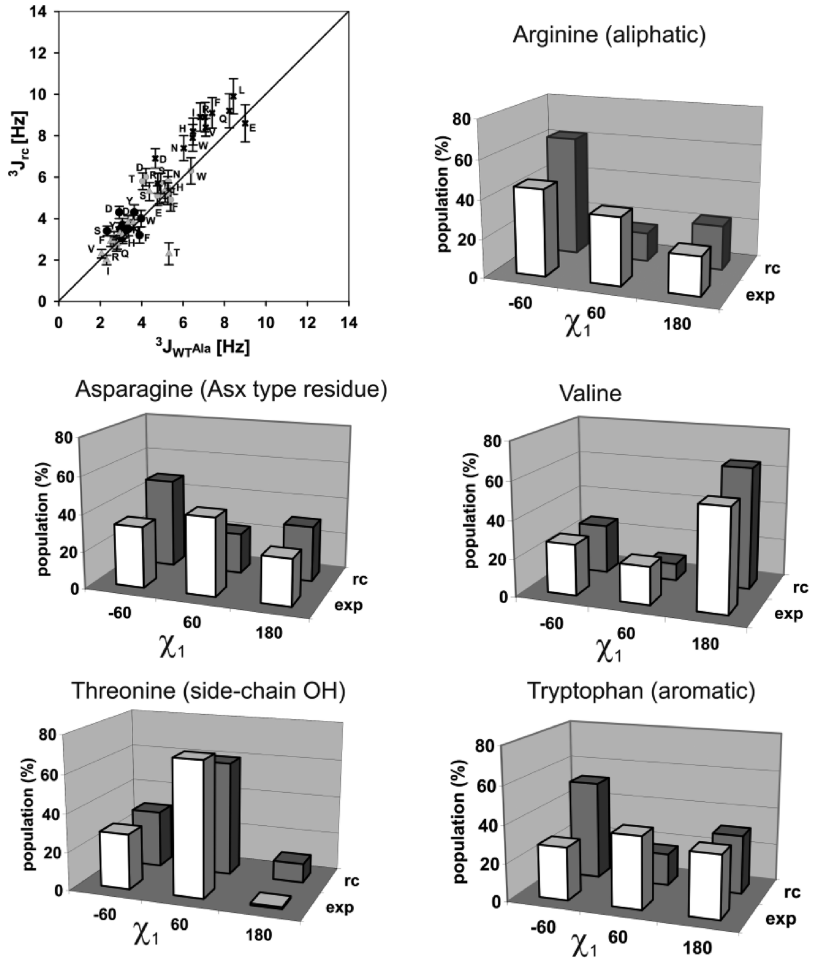


Figure 5. Comparison (top left) of $^3J(CO,H^{\beta})$ and $^3J(H^{\alpha},H^{\beta})$ coupling constants of WT^{Ala} with random-coil predictions: (black circles) $^3J(H^{\alpha},H^{\beta 2})$, (gray triangles) $^3J(H^{\alpha},H^{\beta}{}_3)$, (black crosses) $^3J(CO,H^{\beta 2})$, and (gray circles) $^3J(CO,H^{\beta 3})$. A close-up of this panel is shown in Figure S17 of the Supporting Information. All other panels show experimental values from WT^{Ala} (exp, white bars) and random-coil values (rc, gray bars) for χ_1 populations in different side chains (following ref 58). All experiments were performed at a field strength corresponding to a ^1H frequency of 600 MHz with 300 μM protein at pH 2.0 and 293 K.

Table 1. Populations of Different χ_1 Angles in WT^{Ala} and Random Coil (rc)[a]

	WT^{Ala}			rc			
	−60°	60°	180°	−60°	60°	180°	
E (1)	45	12	42	56	17	26	al
I (3)	47 ± 9	29 ± 8	24 ± 2	64	25	11	al
Q (1)	48	17	36	65	13	22	al
R (4)	45 ± 7	35 ± 11	20 ± 16	62	15	23	al
D (2)	41 ± 12	66 ± 15	−8 ± 5	41	27	32	Asx
N (4)	33 ± 5	42 ± 14	25 ± 10	48	22	30	Asx
S (1)	18	43	38	31	46	23	OH
T (1)	29	70	1	30	60	10	OH
V (4)	27 ± 3	20 ± 2	54 ± 5	26	9	64	Val
H (1)	41	41	18	57	20	23	ar
F (1)	39	30	32	62	13	25	ar
W (2)	28 ± 3	38 ± 3	33 ± 5	52	17	31	ar
Y (1)	31	24	42	60	14	27	ar

[a]Numbers in parentheses in the first column indicate the numbers of amino acids for which the populations have been calculated. Abbreviations: al, aliphatic residues; Asx, Asx-type residues; OH, residues with a side chain OH; Val, valine; ar, aromatic residues.

neighbor effects (Figure S9 of the Supporting Information). In addition, only the $W108G^{Ala}$ mutant showed deviations in H^N chemical shifts at Gly71 and Ser72, indicating long-range interactions of the cluster around W108 with the cluster around G71 and S72. More pronounced changes were observed in the RDC and R_2 profiles of these mutants (Figure 6). We first analyzed the overall behavior of the polypeptide chain by analyzing the distribution of residual dipolar couplings and relaxation rates of WT^{Ala} and its mutants (Figure 6, right panels). The observed distributions for RDCs and R_2 values of all mutants were asymmetric, with a tail toward higher values reflecting residual structure within these mutants. The center of the RDC distribution for all mutants fell within the range of the data obtained for the random-coil ensemble generated by the *flexible meccano* approach (Figure S10 of the Supporting Information). In contrast, R_2 values were significantly higher than the values expected for a random coil, with the center of the distribution outside the range expected for a random coil. This deviation reflected a high degree of residual structure persisting on the subnanosecond time scale, while the degree of residual structure is further averaged on the millisecond time scale. For the $W62G^{Ala}$ mutant, the distributions were rather narrow. Broader RDC and R_2 distributions were observed for all other variants. The degree of residual structure was thus

dx.doi.org/10.1021/bi300222f | *Biochemistry* 2012, 51, 3361−3372

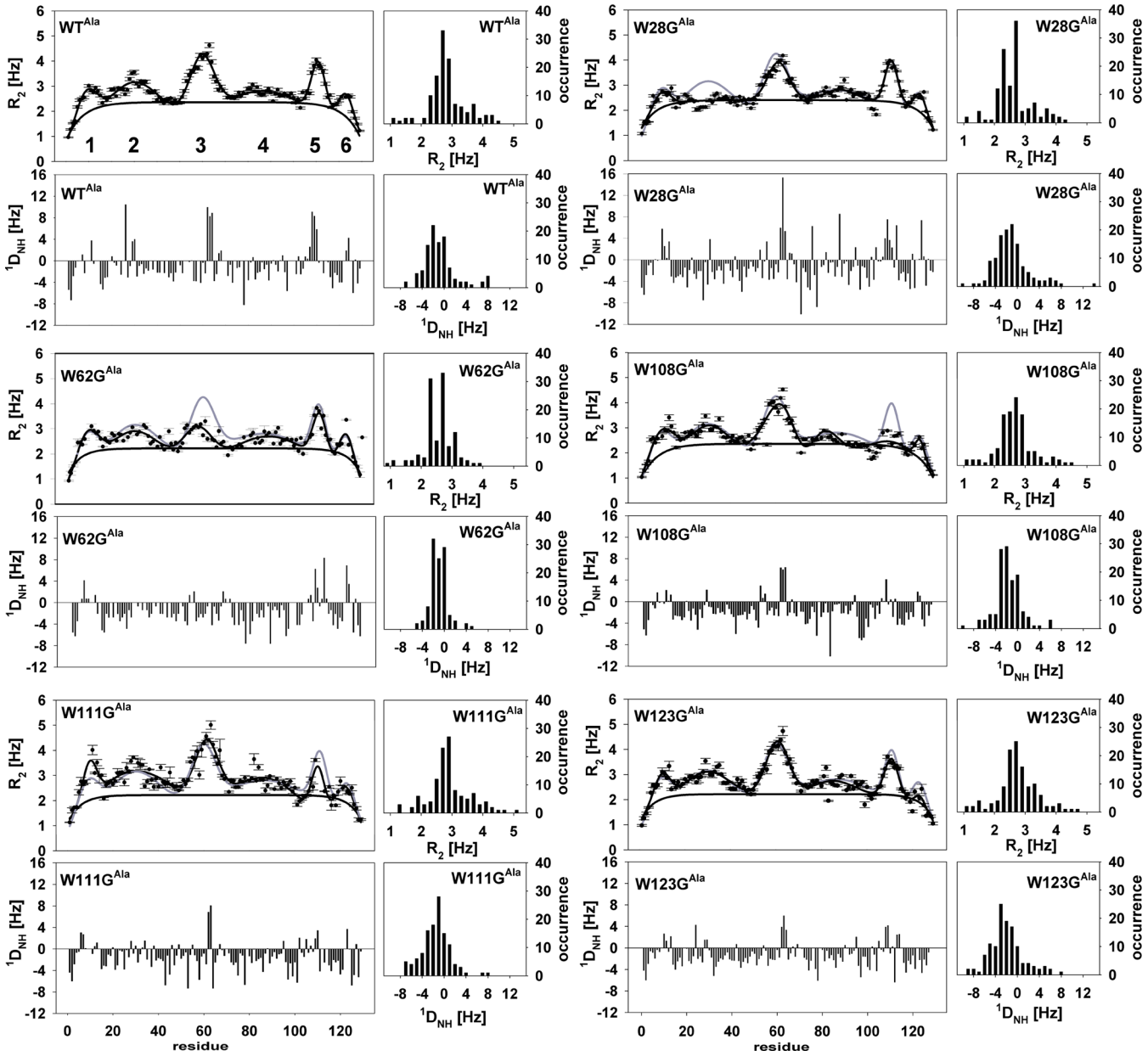


Figure 6. ^{15}N R_2 rates and residual dipolar couplings (RDCs). $^1D(^1H,^{15}N)$ values for WT^{Ala} and mutants W28G^{Ala}, W62G^{Ala}, W108G^{Ala}, W111G^{Ala}, and W123G^{Ala} and their distributions (small panels). The fits of experimental R_2 rates are colored black, and the fit of WT^{Ala} is shown in gray; the baseline is colored black. Distributions derive from *flexible meccano* simulations and R_2 baseline predictions. All experiments were performed at a field strength corresponding to a ^{1}H frequency of 600 MHz, using 300 μM protein at pH 2 and 293 K.

reduced in the W62G^{Ala} mutants, while differences between WT^{Ala} and the other mutants were less pronounced.

Sequence-specific analysis of RDC and R_2 data (Figure 6, left panels) revealed that mutations lead to changes both within the local, perturbed cluster and within remote clusters distant from the mutation site, indicating long-range interactions between the clusters. In W28G^{Ala}, the cluster around the mutation site had vanished; cluster 3 is attenuated, while the other three clusters remained unchanged by the mutation. Only minor changes were observed in W123G^{Ala}: while this cluster becomes less pronounced and minor changes were observed in the neighboring cluster around W108 and W111, the remainder of the protein remained unchanged.

To further complete the description of the unfolded state of lysozyme, additional NMR diffusion measurements[44] and SAXS[61] measurements were performed to obtain the radius of hydration (R_h) as well as the radius of gyration (R_g) (Figure 7). A hydrodynamic radius (R_h) of 31.8 ± 0.5 Å was obtained for WT^{Ala}. The observed value was considerably (14%) smaller than that observed for strongly reduced and methylated hen lysozyme in 8 M urea (R_h = 34.6 Å), as expected for a urea-denatured protein.[44] In WT^{Ala}, an R_g of 35.67 Å was obtained. There was no general correlation between R_g and R_h; however, an empirical linear correlation between $1/R_h$ and $1/R_g$ of 10000 structures from the *flexible meccano* approach was obtained [$R_h^1 = 0.021 + 0.043R_g^{-1}$ (Figure S11 of the Supporting Information)]. The degree of residual structure in W123G^{Ala} was only slightly reduced compared to that in WT^{Ala}. Residual secondary structure and overall compaction as indicated by R_h (31.0 Å) and R_g (38.84 Å) remained similar to those in WT^{Ala}.

dx.doi.org/10.1021/bi300222f | *Biochemistry* 2012, 51, 3361−3372

In W28G^{Ala}, the R_h (30.9 Å) was the same as that of WT^{Ala} while the R_g was significantly larger (41.02 Å).

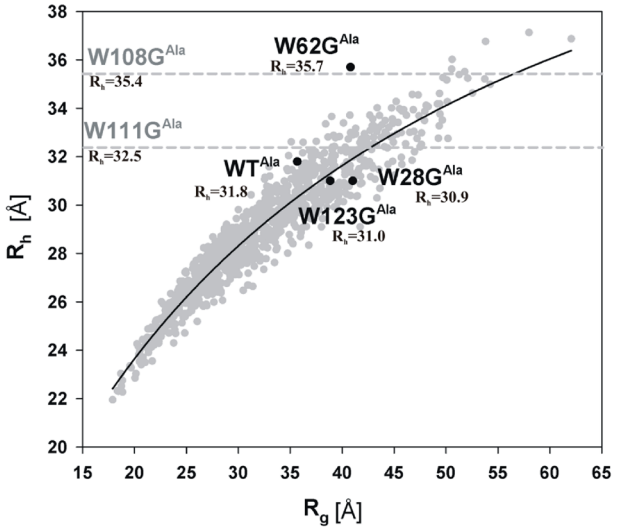


Figure 7. R_h values from NMR diffusion measurements of WT^{Ala} and five mutants (highlighted as black dots) plotted vs their R_g values obtained from SAXS measurements (data fit done with EOM[48]). Gray circles represent an ensemble of 10000 structures of WT^{Ala} from *flexible meccano* simulations; the black line indicates a regression of these simulation data. R_h values were from DOSY measurements.

The residual structure around W28 was thus not important for the overall stabilization of the protein but was stabilized by the clusters around W111G^{Ala} and W108G^{Ala} as seen in the additional mutants investigated. Both mutants (W111G^{Ala} and W108G^{Ala}) were less compact than WT^{Ala} with an R_h of 32.5 ± 1.0 Å for W111G^{Ala} and an R_h of 35.4 ± 0.6 Å for W108G^{Ala}, while the R_g values of both variants could not be obtained because of the apparent aggregation visible in the SAXS results. In both mutants, dynamics increase around the mutation site and in the neighboring cluster. Structure and dynamics on the millisecond time scale as monitored by the RDCs around W28 were significantly altered. Furthermore, in the W108G^{Ala} mutant, residual structure and dynamics in cluster 4 were affected as seen by R_2 rates and analysis of chemical shift differences. The most flexible mutant was the W62G^{Ala} mutant with an R_h of 35.7 ± 0.9 Å and an R_g of 40.80 Å; both RDC and R_2 values were closest to random-coil values.

Combining experimental and computational approaches, to gain more precise insight into which structural preferences exist in the polypeptide chain on a residue-specific level, we performed an ASTEROIDS[49,52,53] selection targeting the C^α, C′, and C^β chemical shifts, coupling constants $^3J(H^N,H^\alpha)$, $^1J(N,C^\alpha_i)$, and $^2J(N,C^\alpha_{i-1})$, and RDCs to select a subensemble in agreement with the experimental data. From a pool of 15000 structures generated by the *flexible meccano* algorithm, a subset of 200 structures was sufficient to reproduce the experimental data. Differences among experimental chemical shifts (C^α, C′,

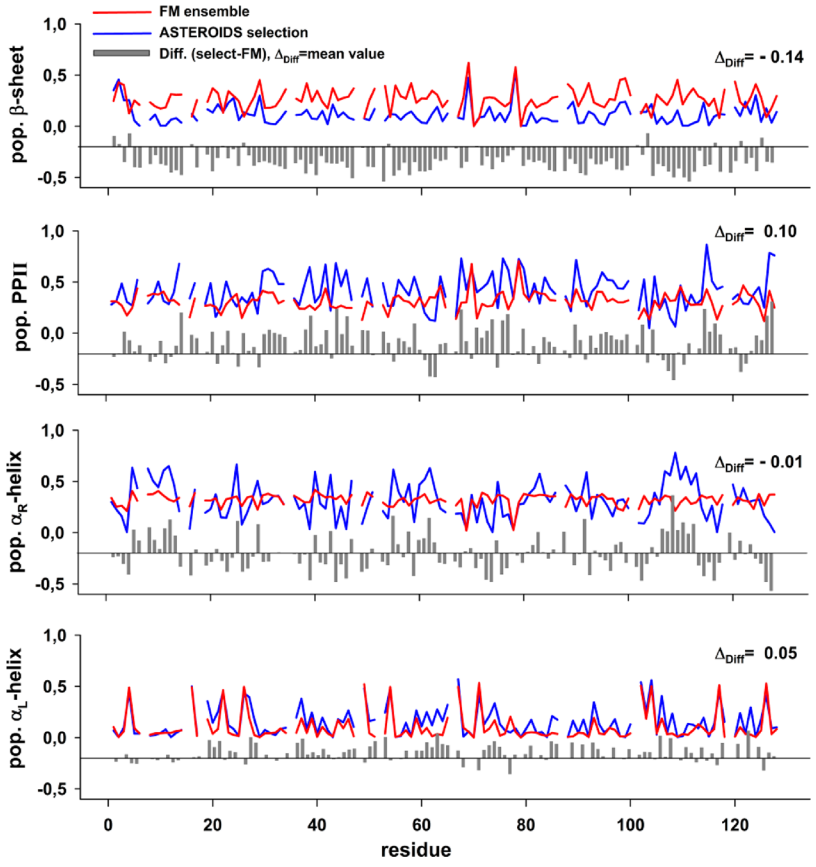


Figure 8. Selected ASTEROIDS populations of secondary structure elements [(A) β-sheet, (B) polyproline II (PP_{II}), (C) α_R-helix, and (D) α_L-helix] of WT^{Ala} (without residues D, E, and H): red for the *flexible meccano*-generated ensemble (15000 structures), blue for ASTEROIDS selection for WT^{Ala} (200 structures), gray bars for the differences between selected structures and the *flexible meccano* ensemble, and Δ_{Diff} for the mean difference values (calculated by $\Delta_{Diff} = Diff_{selected} - Diff_{flecmec}$).

dx.doi.org/10.1021/bi300222f | *Biochemistry* 2012, 51, 3361−3372

and C^{β}), coupling constants [$^{3}J(H^{N},H^{\alpha})$, $^{1}J(N,C^{\alpha}_{i})$, and $^{2}J(N,C^{\alpha}_{i-1})$], RDCs, and calculated data by ASTEROIDS were in a small range (Figures S12 and S13 of the Supporting Information). The H^{α} chemical shifts were not included in the ASTEROIDS selection for the populations (Figure 8), and a comparison between experimental and back-calculated H^{α} chemical shifts offered a reasonable validation of the selected ensemble.

Globally, the ASTEROIDS selection was shifted toward a higher content of PP_{II} ($\Delta_{Diff} = 0.10$) and α_L-helix ($\Delta_{Diff} = 0.05$), was considerably diminished in β-sheet content ($\Delta_{Diff} = -0.14$), and was unchanged in α_R-helical content ($\Delta_{Diff} = -0.01$) compared to the *flexible meccano*-generated ensemble. Although the difference in α_R-helical content was small, substantial regional fluctuations were observed. A closer analysis, where each of the clusters (1−3, 5, and 6) was examined (Table 2), showed that

Table 2. Regional Preferences of the ASTEROIDS-Selected Ensemble[a]

cluster	β-sheet	PP_{II}	α_R-helix	α_L-helix
1	−−	0	++	0
2	−−	++	0	0
3(N)	−−	++	0	−
3(C)	−−	0	0	++
5	−−	−	++	0
6	−	0	0	+

[a]Clusters 1−3, 5, and 6 were analyzed in terms of mean differences and evaluated by −−, −, 0, +, and ++, depending on the size of the deviation. Cluster 3 was split into an N-terminal portion [3(N)] and a C-terminal portion [3(C)].

for each cluster a different subset of populations can be deduced. All clusters revealed a decreased β-sheet population. This decrease in β-sheet population was shifted toward different conformations (α_R, α_L, and PP_{II}) in the different clusters. Clusters 1 and 5 exhibited a higher α_R-helical content; additionally, cluster 5 was also diminished in PP_{II} content. Clusters 2 and 3(N) showed high PP_{II} content, whereas clusters 3(C) and 6 showed high α_L-helical content. This analysis showed that structural preference in hydrophobic clusters can be versatile. The unique combination of chemical shifts, coupling constants, and RDCs combined with *flexible meccano* and ASTEROIDS allowed the detailed description of structural preferences in clusters of non-native states of lysozyme.

The analysis of ensemble compaction of *flexible meccano*-generated ensembles and ASTEROIDS-selected ensembles (Figures S14 and S15 of the Supporting Information) revealed that ASTEROIDS-selected ensembles were shifted toward a higher level of compaction compared to *flexible meccano*-generated ensembles.

CONCLUSIONS

Using WT^{Ala} as a model, we present here a number of novelties in the analysis of unfolded protein ensembles. The use of $^{1}J(N,C^{\alpha})$ and $^{2}J(N,C^{\alpha})$ coupling constants is an additional method for identifying residual secondary structure other than chemical shifts only, because the coupling constants show a dependence on the angle (ψ). Only this dependence allows us to differentiate between negative and positive ψ values.

Scalar side chain couplings in unfolded proteins reveal information about the side chain angle χ_1, which is influenced by the overall compaction of the protein. Further insight into the overall behavior of unfolded proteins can be obtained by the analysis of R_2 and RDC distributions as shown here. The distribution analysis allows identification of changes in the structure and dynamics of protein mutants, even if no chemical shift assignment can be obtained. A more detailed analysis is possible using a $J(\omega_N)/J(0)$ correlation readily allowing differentiation between residues whose motions are governed by internal, overall motions, or both can be made by this analysis in addition to the identification of chemical exchange.

We show that both the analysis of R_2 and RDC data in a joint manner and introduction of single-point mutants can be used to identify regions within the proteins with differential stabilities. To complete these findings, ASTEROIDS calculations further narrow the picture of the degree of residual structure and compactness of unfolded states of hen lysozyme by combining chemical shifts, J couplings, and RDCs. Evidence of helical propensity has been provided through different time scales of averaging; SAXS measurements confirm dynamical characteristics obtained from DOSY diffusion experiments and reveal concentration-dependent aggregation processes in the wild type and single-point mutants.

Given the small energetic differences between ensemble members of the unfolded states of proteins, it is remarkable that NMR spectroscopy can detect very subtle differences in population shifts as well as local and global differences in polypeptide chain dynamics. Modeling approaches have now reached a state where both random-coil properties and conformational biases as small as 10%, presumably corresponding to energetic shifts of the order of kT, can be detected. Providing a comprehensive experimental data set for a well-characterized model protein will allow benchmarking of different algorithms for the prediction of structure and dynamics in unfolded and non-native states of proteins. Such characterization provides precious information about intrinsically unstructured proteins and the effects of single-point mutants and may form the basis of the structural code underlying folding and misfolding of proteins.

ASSOCIATED CONTENT

Supporting Information

Figures S1−S17 and Tables S1 and S2. This material is available free of charge via the Internet at http://pubs.acs.org.

Accession Codes

Associated BMRB entries 18365, 18366, 18367, 18368, 18369, and 18370.

AUTHOR INFORMATION

Corresponding Author

*Institute for Organic Chemistry and Chemical Biology, Center of Biomolecular Magnetic Resonance, Goethe University Frankfurt, Max von-Laue-Strasse 7, 60438 Frankfurt am Main, Germany. E-mail: schwalbe@nmr.uni-frankfurt.de (H.S.) or wirmer@nmr.uni-frankfurt.de (J.W.-B.). Phone: ++49-69- 798-29737. Fax: ++49-69-798-29515.

Author Contributions

F.S. and R.S. contributed equally to this work.

Funding

This work has been supported by the Joint Research Activities program of the EU-funded infrastructure initiative BIO-NMR. We thank TAUSTRUCT-ANR MALZ 2010 (M.B.) and PROTEIN DISORDER-ANR JCJC 2010 (M.R.J.) for financial support. H.S. is a member of the DFG-funded cluster of excellence: macromolecular complexes. F.S. has been supported by the Evonik

Stiftung (Essen, Germany). R.S. was supported by the Stiftung Polytechnische Gesellschaft (Frankfurt am Main, Germany).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We acknowledge computer time provided by the Center for Scientific Computing Frankfurt (CSC).

■ ABBREVIATIONS

ACBP, acyl-CoA-binding protein; CD, circular dichroism; CPMG, Carr−Purcell−Meiboom−Gill; DOSY, diffusion-ordered spectroscopy; EDTA, ethylenediaminetetraacetic acid; IPTG, isopropyl β-D-thiogalactopyranoside; HEWL, hen egg white lysozyme; hetNOE, heteronuclear NOE; HPLC, high-performance liquid chromatography; HSQC, heteronuclear single-quantum coherence; NMR, nuclear magnetic resonance; NOESY, nuclear Overhauser enhancement spectroscopy; PP_{II}, polyproline II helix; RDC, residual dipolar coupling; SAXS, small-angle X-ray scattering; SSP, secondary structure propensity; WT^{Ala}, wild-type lysozyme bearing eight cysteine-to-alanine mutations; W28G, W62G, W108G, W111G, and W123G, tryptophan-to-glycine mutants of WT^{Ala}.

■ REFERENCES

(1) Dobson, C. M. (2003) Protein folding and misfolding. *Nature 426*, 884−890.

(2) Buchner, J., and Kiefhaber, T., Eds. (2005) *Protein folding handbook*, Wiley-VCH, Weinheim, Germany.

(3) Schwalbe, H., Fiebig, K. M., Buck, M., Jones, J. A., Grimshaw, S. B., Spencer, A., Glaser, S. J., Smith, L. J., and Dobson, C. M. (1997) Structural and dynamical properties of a denatured protein. Heteronuclear 3D NMR experiments and theoretical simulations of lysozyme in 8 M urea. *Biochemistry 36*, 8977−8991.

(4) Klein-Seetharaman, J., Oikawa, M., Grimshaw, S. B., Wirmer, J., Duchardt, E., Ueda, T., Imoto, T., Smith, L. J., Dobson, C. M., and Schwalbe, H. (2002) Long-range interactions within a nonnative protein. *Science 295*, 1719−1722.

(5) Wirmer, J., Schlorb, C., Klein-Seetharaman, J., Hirano, R., Ueda, T., Imoto, T., and Schwalbe, H. (2004) Modulation of compactness and long-range interactions of unfolded lysozyme by single point mutations. *Angew. Chem., Int. Ed. 43*, 5780−5785.

(6) Redfield, C., Schulman, B. A., Milhollen, M. A., Kim, P. S., and Dobson, C. M. (1999) α-Lactalbumin forms a compact molten globule in the absence of disulfide bonds. *Nat. Struct. Biol. 6*, 948−952.

(7) Wirmer, J., Berk, H., Ugolini, R., Redfield, C., and Schwalbe, H. (2006) Characterization of the unfolded state of bovine α-lactalbumin and comparison with unfolded states of homologous proteins. *Protein Sci. 15*, 1397−1407.

(8) Higman, V. A., Rosner, H. I., Ugolini, R., Greene, L. H., Redfield, C., and Smith, L. J. (2009) Probing the urea dependence of residual structure in denatured human α-lactalbumin. *J. Biomol. NMR 45*, 121−131.

(9) Obolensky, O. I., Schlepckow, K., Schwalbe, H., and Solov'yov, A. V. (2007) Theoretical framework for NMR residual dipolar couplings in unfolded proteins. *J. Biomol. NMR 39*, 1−16.

(10) Louhivuori, M., Fredriksson, K., Paakkonen, K., Permi, P., and Annila, A. (2004) Alignment of chain-like molecules. *J. Biomol. NMR 29*, 517−524.

(11) Fiebig, K. M., Schwalbe, H., Buck, M., Smith, L. J., and Dobson, C. M. (1996) Toward a Description of the Conformations of Denatured States of Proteins. Comparison of a Random Coil Model with NMR Measurements. *J. Phys. Chem. 100*, 2661−2666.

(12) Smith, L. J., Fiebig, K. M., Schwalbe, H., and Dobson, C. M. (1996) The concept of a random coil: Residual structure in peptides and denatured proteins. *Folding Des. 1*, R95−R106.

(13) Bernado, P., Blanchard, L., Timmins, P., Marion, D., Ruigrok, R. W., and Blackledge, M. (2005) A structural model for unfolded proteins from residual dipolar couplings and small-angle X-ray scattering. *Proc. Natl. Acad. Sci. U.S.A. 102*, 17002−17007.

(14) Mukrasch, M. D., Markwick, P., Biernat, J., Bergen, M., Bernado, P., Griesinger, C., Mandelkow, E., Zweckstetter, M., and Blackledge, M. (2007) Highly populated turn conformations in natively unfolded tau protein identified from residual dipolar couplings and molecular simulation. *J. Am. Chem. Soc. 129*, 5235−5243.

(15) Wells, M., Tidow, H., Rutherford, T. J., Markwick, P., Jensen, M. R., Mylonas, E., Svergun, D. I., Blackledge, M., and Fersht, A. R. (2008) Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain. *Proc. Natl. Acad. Sci. U.S.A. 105*, 5762−5767.

(16) Jensen, M. R., Markwick, P. R., Meier, S., Griesinger, C., Zweckstetter, M., Grzesiek, S., Bernado, P., and Blackledge, M. (2009) Quantitative determination of the conformational properties of partially folded and intrinsically disordered proteins using NMR dipolar couplings. *Structure 17*, 1169−1185.

(17) Alexandrescu, A. T., Jahnke, W., Wiltscheck, R., and Blommers, M. J. (1996) Accretion of structure in staphylococcal nuclease: An ^{15}N NMR relaxation study. *J. Mol. Biol. 260*, 570−587.

(18) Fieber, W., Kristjansdottir, S., and Poulsen, F. M. (2004) Short-range, long-range and transition state interactions in the denatured state of ACBP from residual dipolar couplings. *J. Mol. Biol. 339*, 1191−1199.

(19) Meier, S., Grzesiek, S., and Blackledge, M. (2007) Mapping the conformational landscape of urea-denatured ubiquitin using residual dipolar couplings. *J. Am. Chem. Soc. 129*, 9799−9807.

(20) Modig, K., and Poulsen, F. M. (2008) Model-independent interpretation of NMR relaxation data for unfolded proteins: The acid-denatured state of ACBP. *J. Biomol. NMR 42*, 163−177.

(21) Mohana-Borges, R., Goto, N. K., Kroon, G. J., Dyson, H. J., and Wright, P. E. (2004) Structural characterization of unfolded states of apomyoglobin using residual dipolar couplings. *J. Mol. Biol. 340*, 1131−1142.

(22) Sallum, C. O., Martel, D. M., Fournier, R. S., Matousek, W. M., and Alexandrescu, A. T. (2005) Sensitivity of NMR residual dipolar couplings to perturbations in folded and denatured staphylococcal nuclease. *Biochemistry 44*, 6392−6403.

(23) Yao, J., Chung, J., Eliezer, D., Wright, P. E., and Dyson, H. J. (2001) NMR structural and dynamic characterization of the acid-unfolded state of apomyoglobin provides insights into the early events in protein folding. *Biochemistry 40*, 3561−3571.

(24) Wirmer, J., Peti, W., and Schwalbe, H. (2006) Motional properties of unfolded ubiquitin: A model for a random coil protein. *J. Biomol. NMR 35*, 175−186.

(25) Mishima, T., Ohkuri, T., Monji, A., Imoto, T., and Ueda, T. (2007) A particular hydrophobic cluster in the residual structure of reduced lysozyme drastically affects the amyloid fibrils formation. *Biochem. Biophys. Res. Commun. 356*, 769−772.

(26) Schlörb, C., Ackermann, K., Richter, C., Wirmer, J., and Schwalbe, H. (2005) Heterologous expression of hen egg white lysozyme and resonance assignment of tryptophan side chains in its non-native states. *J. Biomol. NMR 33*, 95−104.

(27) Bodenhausen, G., and Ruben, D. J. (1980) Natural Abundance N-15 NMR by Enhanced Heteronuclear Spectroscopy. *Chem. Phys. Lett. 69*, 185−189.

(28) Palmer, A. G., Cavanagh, J., Wright, P. E., and Rance, M. (1991) Sensitivity Improvement in Proton-Detected 2-Dimensional Heteronuclear Correlation NMR-Spectroscopy. *J. Magn. Reson. 93*, 151−170.

(29) Schleucher, J., Schwendinger, M., Sattler, M., Schmidt, P., Schedletzky, O., Glaser, S. J., Sorensen, O. W., and Griesinger, C. (1994) A general enhancement scheme in heteronuclear multidimensional NMR employing pulsed field gradients. *J. Biomol. NMR 4*, 301−306.

(30) Wüthrich, K. (2005) Optimizing the process of nuclear magnetic resonance spectrum analysis and computer aided resonance

assignment. Ph.D. Dissertation, Swiss Federal Institute of Technology Zurich, Zurich.
(31) Tollinger, M., Skrynnikov, N. R., Mulder, F. A., Forman-Kay, J. D., and Kay, L. E. (2001) Slow dynamics in folded and unfolded states of an SH3 domain. *J. Am. Chem. Soc. 123*, 11341−11352.
(32) Wirmer, J., and Schwalbe, H. (2002) Angular dependence of $^{1}J(N_i,C_{\alpha i})$ and $^{2}J(N_i,C_{a(i-1)})$ coupling constants measured in J-modulated HSQCs. *J. Biomol. NMR 23*, 47−55.
(33) Vuister, G. W., and Bax, A. (1993) Quantitative J Correlation: A New Approach for Measuring Homonuclear 3-Bond J(H(N)H(α)) Coupling-Constants in N-15-Enriched Proteins. *J. Am. Chem. Soc. 115*, 7772−7777.
(34) Lohr, F., and Ruterjans, H. (1996) Novel Pulse Sequences for the Resonance Assignment of Aromatic Side Chains in ^{13}C-Labeled Proteins. *J. Magn. Reson. 112*, 259−268.
(35) Ishii, Y., Markus, M. A., and Tycko, R. (2001) Controlling residual dipolar couplings in high-resolution NMR of proteins by strain induced alignment in a gel. *J. Biomol. NMR 21*, 141−151.
(36) Sass, H. J., Musco, G., Stahl, S. J., Wingfield, P. T., and Grzesiek, S. (2000) Solution NMR of proteins within polyacrylamide gels: Diffusional properties and residual alignment by mechanical stress or embedding of oriented purple membranes. *J. Biomol. NMR 18*, 303−309.
(37) Ottiger, M., Delaglio, F., and Bax, A. (1998) Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. *J. Magn. Reson. 131*, 373−378.
(38) Zweckstetter, M. (2008) NMR: Prediction of molecular alignment from structure using the PALES software. *Nat. Protoc. 3*, 679−690.
(39) Ortega, A., Amorós, D., and García de la Torre, J. (2011) Prediction of Hydrodynamic and Other Solution Properties of Rigid Proteins from Atomic- and Residue-Level Models. *Biophys. J. 101*, 892−898.
(40) Marsh, J. A., Singh, V. K., Jia, Z., and Forman-Kay, J. D. (2006) Sensitivity of secondary structure propensities to sequence differences between α- and γ-synuclein: Implications for fibrillation. *Protein Sci. 15*, 2795−2804.
(41) Wishart, D. S., Sykes, B. D., and Richards, F. M. (1992) The chemical shift index: A fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry 31*, 1647−1651.
(42) Wishart, D. S., and Sykes, B. D. (1994) The ^{13}C Chemical-Shift Index: A simple method for the identification of protein secondary structure using ^{13}C chemical-shift data. *J. Biomol. NMR 4*, 171−180.
(43) Smith, L. J., Bolin, K. A., Schwalbe, H., MacArthur, M. W., Thornton, J. M., and Dobson, C. M. (1996) Analysis of main chain torsion angles in proteins: prediction of NMR coupling constants for native and random coil conformations. *J. Mol. Biol. 255*, 494−506.
(44) Wilkins, D. K., Grimshaw, S. B., Receveur, V., Dobson, C. M., Jones, J. A., and Smith, L. J. (1999) Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. *Biochemistry 38*, 16424−16431.
(45) Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J., and Svergun, D. I. (2003) PRIMUS: A Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr. 36*, 1277−1282.
(46) Bernado, P., Mylonas, E., Petoukhov, M. V., Blackledge, M., and Svergun, D. I. (2007) Structural characterization of flexible proteins using small-angle X-ray scattering. *J. Am. Chem. Soc. 129*, 5656−5664.
(47) Debye, P., Anderson, H. R., and Brumberger, H. (1957) Scattering by an Inhomogeneous Solid. 2. The Correlation Function and Its Application. *J. Appl. Phys. 28*, 679−683.
(48) Porod, G. (1951) Die Röntgenkleinwinkelstreuung von dichtgepackten kolloiden Systemen, I. Teil. *Kolloid Z. Z. Polym. 124*, 31.
(49) Jensen, M. R., Salmon, L., Nodet, G., and Blackledge, M. (2010) Defining conformational ensembles of intrinsically disordered and partially folded proteins directly from chemical shifts. *J. Am. Chem. Soc. 132*, 1270−1272.
(50) Pardi, A., Billeter, M., and Wuthrich, K. (1984) Calibration of the angular dependence of the amide proton-Cα proton coupling constants, $^{3}J_{HN\alpha}$, in a globular protein. Use of $^{3}J_{HN\alpha}$ for identification of helical secondary structure. *J. Mol. Biol. 180*, 741−751.
(51) Marsh, J. A., Baker, J. M., Tollinger, M., and Forman-Kay, J. D. (2008) Calculation of residual dipolar couplings from disordered state ensembles using local alignment. *J. Am. Chem. Soc. 130*, 7804−7805.
(52) Nodet, G., Salmon, L., Ozenne, V., Meier, S., Jensen, M. R., and Blackledge, M. (2009) Quantitative description of backbone conformational sampling of unfolded proteins at amino acid resolution from NMR residual dipolar couplings. *J. Am. Chem. Soc. 131*, 17908−17918.
(53) Salmon, L., Nodet, G., Ozenne, V., Yin, G., Jensen, M. R., Zweckstetter, M., and Blackledge, M. (2010) NMR characterization of long-range order in intrinsically disordered proteins. *J. Am. Chem. Soc. 132*, 8407−8418.
(54) Lipari, G., and Szabo, A. (1982) Analysis of NMR Relaxation Data on Macromolecules Using the Model-Free Approach. *Biophys. J. 37*, A380.
(55) Krizova, H., Zidek, L., Stone, M. J., Novotny, M. V., and Sklenar, V. (2004) Temperature-dependent spectral density analysis applied to monitoring backbone dynamics of major urinary protein-I complexed with the pheromone 2-sec-butyl-4,5-dihydrothiazole. *J. Biomol. NMR 28*, 369−384.
(56) Hähnke, M. J., Richter, C., Heinicke, F., and Schwalbe, H. (2010) The HN(COCA)HAHB NMR experiment for the stereo-specific assignment of Hβ-protons in non-native states of proteins. *J. Am. Chem. Soc. 132*, 918−919.
(57) Pachler, K. G. R. (1963) Nuclear Magnetic Resonance Study of Some α-Amino Acids. 1. Coupling Constants in Alkaline and Acidic Medium. *Spectrochim. Acta 19*, 2085−2092.
(58) Hennig, M., Bermel, W., Spencer, A., Dobson, C. M., Smith, L. J., and Schwalbe, H. (1999) Side-chain conformations in an unfolded protein: χ_1 distributions in denatured hen lysozyme determined by heteronuclear ^{13}C, ^{15}N NMR spectroscopy. *J. Mol. Biol. 288*, 705−723.
(59) Vajpai, N., Gentner, M., Huang, J. R., Blackledge, M., and Grzesiek, S. (2010) Side-chain χ_1 conformations in urea-denatured ubiquitin and protein G from ^{3}J coupling constants and residual dipolar couplings. *J. Am. Chem. Soc. 132*, 3196−3203.
(60) Schwalbe, H., Grimshaw, S. B., Spencer, A., Buck, M., Boyd, J., Dobson, C. M., Redfield, C., and Smith, L. J. (2001) A refined solution structure of hen lysozyme determined using residual dipolar coupling data. *Protein Sci. 10*, 677−688.
(61) Svergun, D. I., and Koch, M. H. (2002) Advances in structure analysis using small-angle scattering in solution. *Curr. Opin. Struct. Biol. 12*, 654−660.
(62) Abraham, D. J., and Leo, A. J. (1987) Extension of the fragment method to calculate amino acid zwitterion and side chain partition coefficients. *Proteins 2*, 130−152.
(63) Frank, M. K., Clore, G. M., and Gronenborn, A. M. (1995) Structural and dynamic characterization of the urea denatured state of the immunoglobulin binding domain of streptococcal protein G by multidimensional heteronuclear NMR spectroscopy. *Protein Sci. 4*, 2605−2615.