Microscopic Analysis of Protein Oxidative Damage: Effect of Carbonylation on Structure, Dynamics, and Aggregability of Villin Headpiece
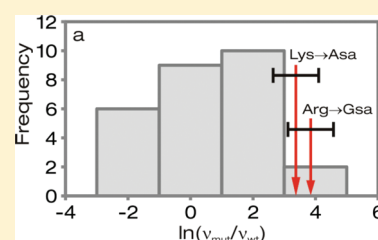
Drazen Petrov and Bojan Zagrovic*

Department of Structural and Computational Biology, Max F. Perutz Laboratories, University of Vienna, Campus Vienna Biocenter 5, Vienna AT-1030, Austria

Mediterranean Institute for Life Sciences, Meštrovićevo šetalište bb, Split HR-21000, Croatia

Department of Physics, Faculty of Science, University of Split, Teslina 12, Split HR-21000, Croatia

S *Supporting Information*

ABSTRACT: One of the most important irreversible oxidative modifications of proteins is carbonylation, the process of introducing a carbonyl group in reaction with reactive oxygen species. Notably, carbonylation increases with the age of cells and is associated with the formation of intracellular protein aggregates and the pathogenesis of age-related disorders such as neurodegenerative diseases and cancer. However, it is still largely unclear how carbonylation affects protein structure, dynamics, and aggregability at the atomic level. Here, we use classical molecular dynamics simulations to study structure and dynamics of the carbonylated headpiece domain of villin, a key actin-organizing protein. We perform an exhaustive set of molecular dynamics simulations of a native villin headpiece together with every possible combination of carbonylated versions of its seven lysine, arginine, and proline residues, quantitatively the most important carbonylable amino acids. Surprisingly, our results suggest that high levels of carbonylation, far above those associated with cell death *in vivo*, may be required to destabilize and unfold protein structure through the disruption of specific stabilizing elements, such as salt bridges or proline kinks, or tampering with the hydrophobic effect. On the other hand, by using thermodynamic integration and molecular hydrophobicity potential approaches, we quantitatively show that carbonylation of hydrophilic lysine and arginine residues is equivalent to introducing hydrophobic, charge-neutral mutations in their place, and, by comparison with experimental results, we demonstrate that this by itself significantly increases the intrinsic aggregation propensity of both structured, native proteins and their unfolded states. Finally, our results provide a foundation for a novel experimental strategy to study the effects of carbonylation on protein structure, dynamics, and aggregability using site-directed mutagenesis.

■ INTRODUCTION

Proteins are frequently modified by different reactions involving reactive oxygen species (ROS), including metal-catalyzed carbonylation, oxidation of aromatic and sulfur-containing amino acid residues, oxidation of the protein backbone, or even protein fragmentation due to backbone breakage.[1−3] One of the most important mechanisms of oxidative damage to proteins is metal-catalyzed carbonylation, where ROS are created in a Fenton-type reaction involving transition metals such as iron or copper.[4] Quantitatively the most important products of carbonylation of amino acids are aminoadipic semialdehyde (Asa), derived from lysine, and glutamic semialdehyde (Gsa), derived from arginine and proline residues (Figure 1a).[5,6] Albeit to a much lower extent, carbonyl groups can also be introduced in proteins by direct carbonylation of threonine residues, a secondary reaction with aldehydes produced during lipid peroxidation, or a secondary reaction with carbonyl derivatives generated in reaction of reducing sugars.[1,2]

As compared to other modifications induced by ROS, carbonylation is relatively difficult to induce, but once proteins get carbonylated, the change is permanent due to the irreversible nature of the reaction.[7,8] For this reason, protein carbonyl content is by far the most widely used marker of protein oxidation.[9] The cellular carbonylation level increases with age, with the concentration of carbonylated proteins rising exponentially during the last third of the life span in a wide range of organisms.[3,10] Furthermore, carbonylation is associated with age-related disorders such as neurodegenerative diseases, cancer, and diabetes, but it should be emphasized that it is still unclear whether carbonylation is a direct cause of aging or just a consequence and a useful reporter on aging.[4] Importantly, the presence of highly carbonylated protein aggregates has been observed in many of these diseases, but so far no clear causal relationship between carbonylation and aggregation has been established.[11,12] Finally, the extreme robustness to ionizing radiation and UV light of some extremophile bacteria such as *Deinococcus radiodurans* has been shown to depend on efficient protection of the proteome against basal and radiation-induced protein carbonylation.[13]
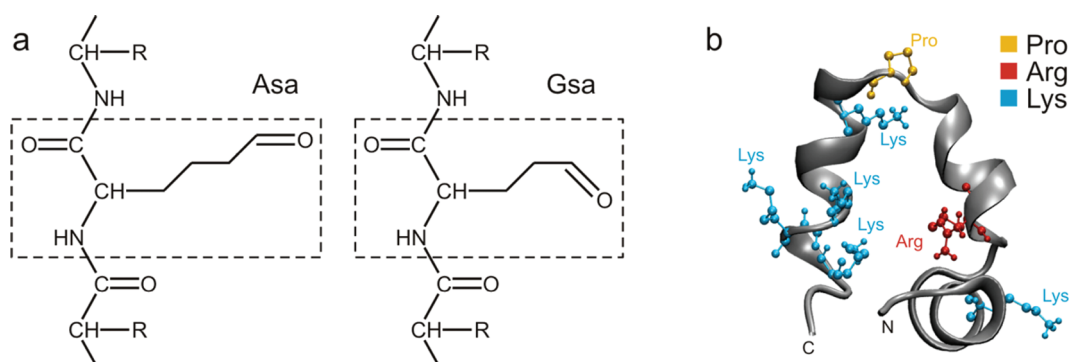
Figure 1. (a) Chemical structures of Asa, derived by carbonylation of Lys, and Gsa, derived by carbonylation of Arg and Pro. (b) Villin structure. The seven quantitatively most important carbonylable sites are colored as follows: Pro, yellow; Arg, red; Lys, blue.

The basal level of carbonylation in cells is approximately 2 nmol of carbonyl per milligram of protein, while the level that appears to correlate with cell and organism death is approximately 6 nmol of carbonyl per milligram of protein, corresponding to about one carbonylated residue per 4000 amino acids on average.[3,10,13] Although these average numbers are actually relatively low, it has been speculated that cellular aging is a direct consequence of the loss of structural stability, unfolding, and exposure of hydrophobic residues of select proteins upon carbonylation.[11] However, to the best of our knowledge, only a few high-resolution studies of structural stability of proteins in the face of oxidative stress have been carried out, and not one focusing explicitly on carbonylation. Most studies focused on the gross functional and structural consequences of oxidation, but with typically very little atomistic details.[14−20] A common denominator of most of these studies is that oxidation of amino acids leads to local disruption of tertiary structure of proteins with a concomitant exposure of hydrophobic amino acids and subsequent aggregation.

Here, we use molecular dynamics (MD)[21] simulation to model the carbonylated villin headpiece molecule. Villin is a tissue-specific actin-binding protein involved in different functions such as cell motility, definition of morphology, and cell death,[22] and it carries out these functions by bundling, nucleating, capping, and severing actin filaments.[23] Our interest in villin was motivated by two principal rationales. First, oxidative stress in non-muscle mammalian cells is known to cause major changes in cellular morphology and structure of the actin cytoskeleton.[24,25] Second, the 36-residue C-terminal headpiece domain of villin is one of the most widely studied and best understood proteins when it comes to folding mechanism and stability.[26−33] Here, we study atomistic-level changes in structure and dynamics of the villin headpiece at different carbonylation levels. In addition to specific effects, such as the disruption of a key salt bridge and a proline kink or alteration of the hydrophobic effect, we ask what overall level of carbonylation can be tolerated without major effects on the molecule's structural and dynamical integrity. Finally, we ask how does carbonylation affects the intrinsic aggregability of the molecule by altering the basic physicochemical properties of the affected amino acids.

■ MATERIALS AND METHODS

We have used MD to study the stability of the carbonylated villin headpiece domain (sequence: MLSDEDF**K**AVFGMT**R**SAFANL**P**LWK-QQNL**KK**E**K**GLF). The seven bolded letters mark the quantitatively most important carbonylable amino acids (K, R, and P) in villin

headpiece (Figure 1b). The simulations were run using the Gromacs 3.3.3 biomolecular simulation package. United-atom GROMOS 45A3 force field,[34,35] SPC explicit water,[36] and a 2 fs integration step were used. For electrostatics calculations, reaction field was employed with a cutoff of r_c = 1.4 nm and the dielectric constant of ε_{rf} = 65. An NMR structure of the villin headpiece domain (PDB code 1VII) was used as the starting structure.[26] After steepest descent minimization was performed in vacuum (500 steps) and subsequently in water (1500 steps), the system was equilibrated by gradually increasing the temperature (from 100 to 300 K) over 100 ps with gradually decreasing position restraints (from 25 000 to 5000 kJ mol^{-1} nm^{-2}) at constant volume and temperature, and finally additionally equilibrated for 20 ps at constant pressure and temperature of 1 bar and 300 K. The temperature and pressure in all production simulations were kept at 300 K and 1 bar using a Berendsen thermostat (τ_T = 0.05 ps) and barostat (τ_p = 1 ps and compressibility = 4.5 × 10^{-5} bar^{-1}).[37]

Since the villin headpiece domain has seven potential carbonylation sites belonging to the quantitatively most important carbonylable residues K, R, and P (see sequence above), the total number of all possible carbonylation combinations of these residues is 2^7 = 128. Altogether, 136 independent 110-ns-long trajectories were generated for a total of 14.96 μs of simulation time (five copies of the native and completely carbonylated villin headpiece plus one copy of every other combination: 5 + 5 + 126 = 136). Coordinates were output every 100 ps, and for the analysis of equilibrium properties, only the last 25 ns of each trajectory was used. Carbonylation modifications were introduced before energy minimization by changing the villin headpiece coordinate file and introducing parameters for the two carbonylated amino acids, Asa and Gsa, in the force field. Details of parametrization for Asa and Gsa are given in the Supporting Information (SI). For analysis, the simulations were divided into four sets: (1) simulations in set K include different levels of carbonylation of Lys residues, (2) those in set KR include Arg15 carbonylation together with different levels of Lys carbonylation, (3) those in set KP include carbonylation of Pro22 together with different levels of Lys carbonylation, and (4) those in set KRP include simultaneous Arg15 and Pro22 carbonylation with different levels of Lys carbonylation. The trajectories were analyzed primarily using Gromacs tools,[38] including calculation of root-mean-square deviation (rmsd), solvent-accessible surface area (SASA), molecular volume, and distances between groups of atoms, except for (1) secondary structure analysis, where DSSP[39] was used, (2) conformational entropy calculations, where quasiharmonic approximation for calculating entropy was used as described in refs 40 and 41, and (3) characterization of hydrophobic properties of protein surface, where the molecular hydrophobicity potential (MHP) calculation was employed, as described by Efremov et al.[42] Throughout the paper, atom-positional rmsd after rotational-translational fitting was calculated with respect to the native NMR villin headpiece structure (PDB code
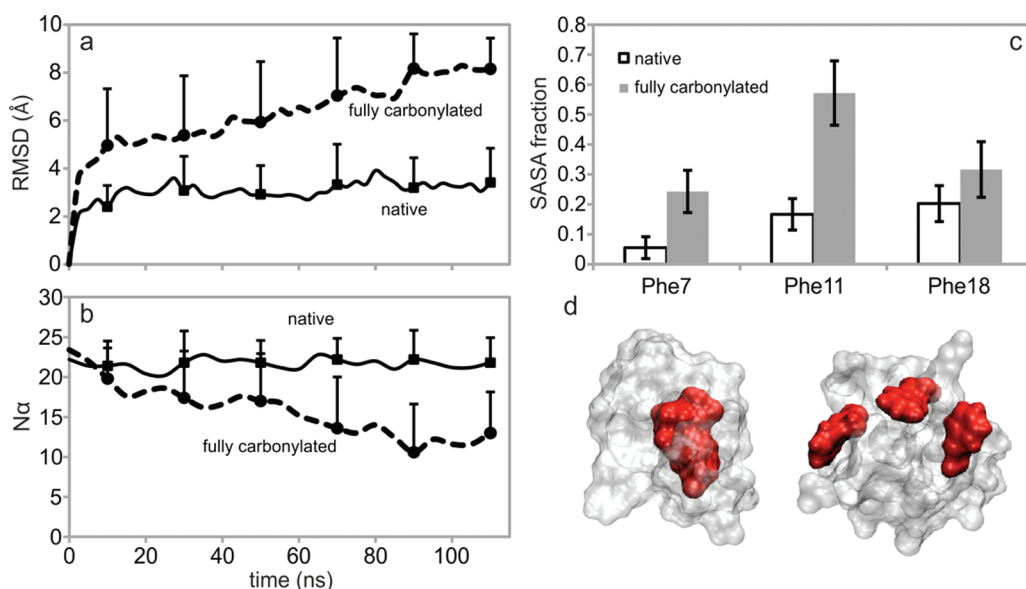
Figure 2. Complete carbonylation leads to unfolding of the villin headpiece. (a) All-atom rmsd from the native villin headpiece structure 1VII (residues 43−74) as a function of time. (b) The number of residues in α-helical conformation as a function of time. In both plots the solid line (■) represents native villin headpiece while the dashed line (●) represents fully carbonylated villin headpiece simulations. (c) Total SASA of the three core Phe residues normalized by the total SASA of fully solvent-exposed Phe. Empty bars represent native villin headpiece while gray bars represent fully carbonylated villin headpiece simulations. All curves and bars in panels a−c were obtained as averages over five independent native/fully carbonylated simulations and are shown with standard deviations. (d) A representative structure of native (left) and completely carbonylated (right) villin headpiece showing the surface of the three core phenylalanines in red and the rest of the protein in white.

Table 1. Types of Trajectories Studied[a]

set	traj no.	carb res	evaluated effects
K	36	0−5 Lys	Δ hydrophobicity
KR	32	0−5 Lys + Arg	salt bridge disruption + Δ hydrophobicity
KP	32	0−5 Lys + Pro	proline kink disruption + Δ hydrophobicity
KRP	36	0−5 Lys + Arg + Pro	all three effects combined

[a] Each set contains a given carbonylation event plus every possible combination of carbonylating the 0−5 Lys residues in the villin headpiece. Set K includes five simulations of the native villin headpiece, i.e., a combination with no Lys residues carbonylated. Similarly, set KRP contains five simulations of the completely carbonylated villin headpiece, i.e., a combination with all five Lys residues carbonylated.

1VII, residues 43−74; the two residues at each end of the peptide were excluded in calculation as they tend to be dynamic). The PDB structure 1VII starts with the residue 41 and ends with the residue 76, since the complete villin headpiece is 76 residues long, while the reported structure captures just the 36-residues-long C-terminal domain. We used 1−36 numbering here, using backbone atoms for alignment and all atoms for rmsd calculation, if not stated otherwise. Exposure of residues to water was estimated by calculating solvent accessibility fraction, calculated as a fraction of SASA calculated for a given side chain in the context of the villin headpiece structure as compared with the SASA of the same side chain when completely exposed to solvent. Distance between given residues was calculated as the distance between centers of masses of their side chains.

We used the formula proposed by Chiti et al.[43] to calculate the change in aggregability upon carbonylation (for structurally destabilized proteins):

$$\ln(\nu_{mut}/\nu_{wt}) = A\Delta\text{Hydr} + B(\Delta\Delta G_{\text{coil-}\alpha} + \Delta\Delta G_{\beta\text{-coil}}) + C\Delta\text{Charge} \quad (1)$$

where ν_{mut} and ν_{wt} are rates of forming aggregates, ΔHydr is the change in hydrophobicity according to the hydrophobicity scale based on water-to-octanol partition,[43] $\Delta\Delta G_{\text{coil-}\alpha} + \Delta\Delta G_{\beta\text{-coil}}$ is the change in propensity to form an α-helix over a β-sheet, ΔCharge is the change in the absolute value of protein net charge, and $A = 0.633$, $B = 0.198$, and $C = -0.491$ are fitted constants. We used only the first term and the last term, with water/octanol partitioning hydrophobicity values for Asa and Gsa obtained from the MHP calculation described by Efremov et al.[42] and the correlation given in Figure 5a (below), but also provided upper and lower bounds derived from the standard deviation of α-helix and β-sheet propensities over the 27 protein studied by Chiti et al.[43] We used the "Zyggregator"[44] model to study the carbonylation-induced change in intrinsic aggregability directly from the native state of the villin headpiece. For this calculation, Asa and Gsa residues were replaced by amino acids that most closely match them in terms of charge and MHP, i.e., Leu and Val, respectively. We could not use Asa and Gsa directly, as not all parameters (such as α and β propensities) needed for prediction are available at this point.

■ RESULTS

Effect of Carbonylation on Protein Structural and Dynamical Integrity. The structure of the villin headpiece domain is principally stabilized by (1) three phenylalanines buried in the
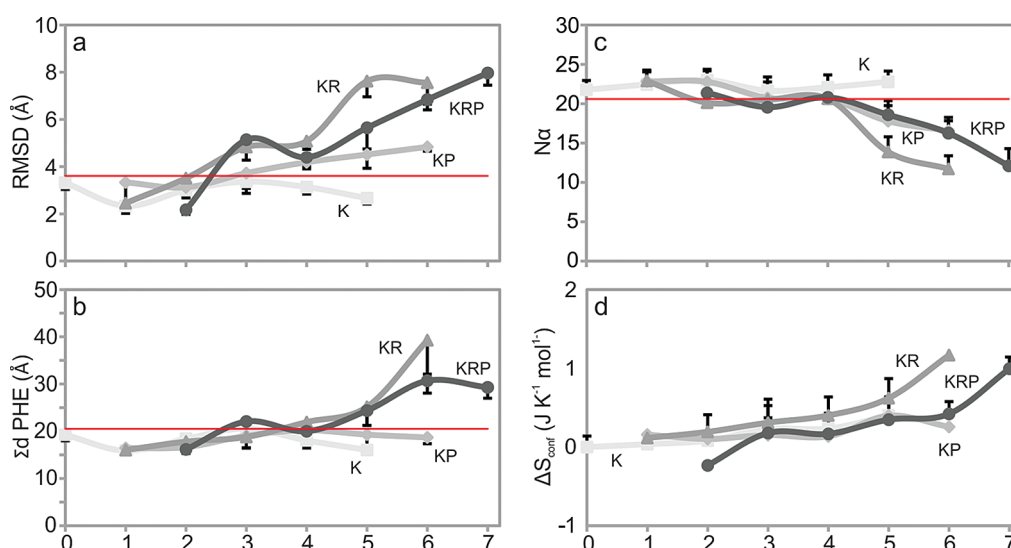
Figure 3. Structural stability of the villin headpiece as a function of carbonylation level: (a) rmsd from the native villin headpiece structure 1VII, (b) sum of the distances between the three core phenylalanines, (c) number of residues in α-helical conformation, and (d) quasi-harmonic conformational entropy normalized by number of atoms, relative to that of native villin headpiece, all shown as a function of the number of carbonylated residues. Each line is obtained from a different subset of simulations: the lightest lines (■), set K; dark gray lines (▲), set KR; light gray lines (◆), set KP; the darkest lines (●), set KRP. All points in the plots were calculated as averages over the last 25 ns in each simulation and all simulations in a given subset with the same number of carbonylated residues and are shown with standard deviations. Red lines represent averages of given observables over the five native simulations, increased by the standard deviations, and were used as a criterion for determining whether a given structure is folded or unfolded.

core of the protein, keeping its α-helices together,[32] and (2) a salt bridge between Asp4 and Arg15 residues.[33] Simultaneous carbonylation of all seven major carbonylable residues in the protein results in the loss of approximately 40% of its α-helical secondary structure and most of its tertiary structure in approximately 100 ns (Figure 2a,b). While the simulated native protein keeps its α-helical content and tertiary structure largely intact over the same period, the carbonylated protein unfolds and starts exploring the accessible areas of the Ramachandran map more freely. Unfolding of the protein upon carbonylation occurs simultaneously with the disintegration of its Phe core (Phe7, Phe11, and Phe18) (Figure 2c,d). Interestingly, the solvent exposure of the Phe core increases multiple-fold as a consequence of carbonylation (Figure 2c), while the total SASA of the protein remains the same although the core residues become exposed to the surface.

What happens if just a subset of different carbonylable residues in villin headpiece are carbonylated, and how does this depend on the type and position of the residues affected? To address this, we have divided our simulations into four distinct sets which were analyzed separately (Table 1). The four sets were organized in such a way to probe different structural effects of carbonylation: surface hydrophobicity change, Asp4-Arg15 salt bridge removal, proline kink removal, or a combination thereof. Interestingly, the structure remains intact when it comes to tertiary fold (as measured by rmsd from the native structure), core compactness (as measured by the sum of distances between the centers of mass of core phenylalanines, ΣdPhe), and secondary structure (as measured by the number of α-helical residues) for all single and double carbonylation hits, regardless of their type or location (Figure 3a−c). Moreover, the structural features remain largely unchanged with regard to all structural measures, even with all five lysine residues carbonylated (K set) (Figure 3a−c). This is striking as Asa, the product of carbonylation of Lys residues, is
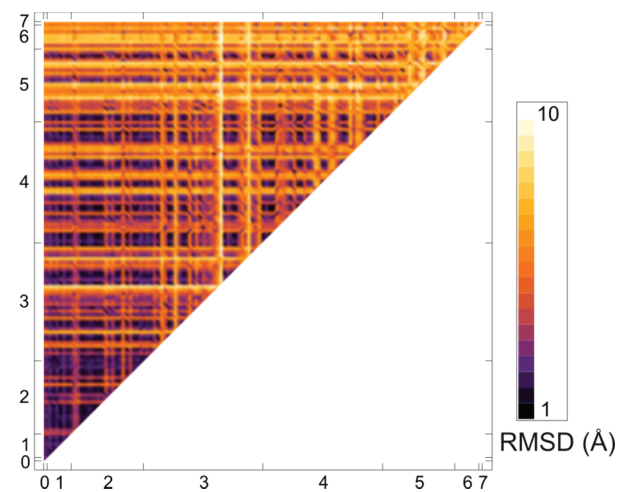

Figure 4. Pairwise rmsd density plot. Each point represents the results of comparison of two ensembles of structures from a given pair of simulations of given combinations of carbonylated residues (linear average of the distribution of all-against-all rmsd evaluations for the two ensembles). Altogether 128 combinations were analyzed, and both *x* and *y* axis range from 0 to 7 (number of carbonylated residues), where 0 represents the native ensemble and 7 represents the fully carbonylated ensemble. Backbone atoms were used for alignment and rmsd calculations.

significantly less hydrophilic than Lys itself (i.e., its hydration free energy is more positive, SI Figure S1), which could lead to a significant disruption of the stabilizing hydrophobic effect. Our thermodynamic integration (TI) calculations suggest that ΔG_{hydr} between them exceeds 10 kJ mol^{-1} in uncharged form and 230 kJ mol^{-1} in charged form (SI Figure S1), with the carbonylation of Arg having a similar effect.
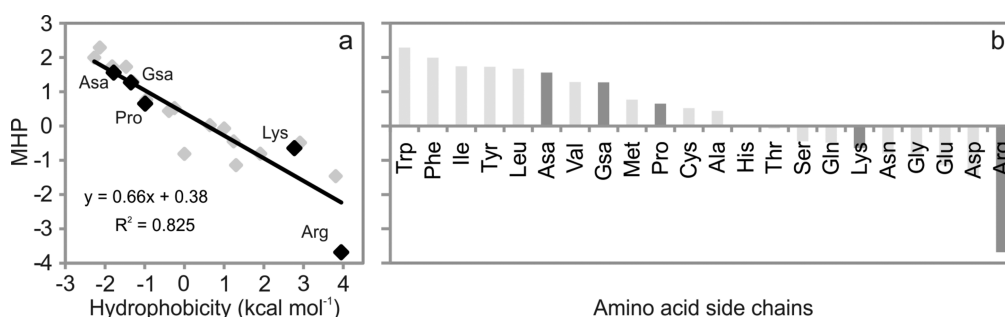
Figure 5. Molecular hydrophobicity potential (MHP) values of native and carbonylated residues. (a) Comparison of the calculated MHP values with the hydrophobicity scale based on free energy of partitioning between aqueous phase and octanol.[43] Large black diamonds represent carbonylable and carbonylated residues Lys, Arg, Pro, Asa, and Gsa, while small gray diamonds represent standard amino acid residues. The regression line (calculated without Asa and Gsa residues) with $R^2 = 0.825$ shows that the calculated values correlate well with hydrophobicity scale for standard amino acid residues. (b) Calculated MHP values for all native and carbonylated residues (black bars, carbonylable and carbonylated residues; gray bars, other native residues).

For all three structural measures of foldedness (rmsd, core compactness, and α-helicity), an ensemble of structures with a given number of carbonylated residues (regardless of their type or position) was defined to be unfolded if the average value of the measure in question, reduced by its standard deviation, was greater that its average over the five native trajectories, increased by its respective standard deviation (e.g., if $\text{rmsd}^{\text{carb}} - \sigma_{\text{rmsd}^{\text{carb}}} > \text{rmsd}^{\text{nat}} + \sigma_{\text{rmsd}^{\text{nat}}}$). According to this definition, significant unfolding (i.e., $\text{rmsd}^{\text{carb}} - \sigma_{\text{rmsd}^{\text{carb}}} > 3.6$ Å, $\Sigma d\text{Phe}^{\text{carb}} - \sigma_{\Sigma d\text{Phe}^{\text{carb}}} > 20.5$ Å, and $N\alpha^{\text{carb}} + \sigma_{N\alpha^{\text{carb}}} < 20.6$) was observed for (1) simulations in the KR set with Arg15 and at least four Lys residues carbonylated according to α-helicity and core compactness, and in simulations with Arg15 and at least two Lys residues carbonylated according to rmsd; (2) simulations in the KP set with Pro22 and at least four Lys residues carbonylated according to α-helicity, and simulations with Pro22 and at least three Lys residues carbonylated according to rmsd; and (3) simulations in the KRP set with Arg15, Pro22, and at least three Lys residues carbonylated according to α-helicity and core compactness, and in simulations with Arg15, Pro22, and at least one Lys residue carbonylated according to rmsd. When it comes to conformational entropy, our results suggest that there is a sizable increase in conformational entropy after complete carbonylation (approximately 1 J K^{-1} mol^{-1} per atom). While this increase is consistent with unfolding of the protein, its absolute magnitude could be affected by the sampling employed. On the whole, our results suggest that globular proteins like villin headpiece likely remain structurally stable upon carbonylation, unless the carbonylation level is high. Moreover, these results suggest that the type and the position of the affected residues do make a difference in terms of the extent of structural damage caused. These findings are further corroborated by the analysis of the average pairwise rmsd between different simulated trajectories exhibiting varying levels and types of carbonylation (Figure 4). According to this analysis, ensembles with five or more carbonylated residues are by-and-large non-native-like, and furthermore their unfolded ensembles are largely mutually different when it comes to structure.

In order to estimate to what extent carbonylation of different residues affects protein stability, we used a simple, two-state model with the difference in free energy between the folded and the unfolded states of villin headpiece represented as $\Delta G_{\text{f-u}}$. We assumed that carbonylation of each residue acts independently and

that carbonylation of each Lys residue has the same thermodynamic effect. Using these assumptions, the total change in the free energy of stabilization of villin headpiece upon carbonylation, $\Delta\Delta G_{\text{carb}}$, can be expressed as

$$\Delta\Delta G_{\text{carb}} = n_K \Delta\Delta G_K + n_R \Delta\Delta G_R + n_P \Delta\Delta G_P \qquad (2)$$

where $\Delta\Delta G_K$, $\Delta\Delta G_R$, and $\Delta\Delta G_P$ are free energy changes upon carbonylation of individual Lys, Arg, and Pro residues, respectively, and n_K, n_R, and n_P are the numbers of carbonylated Lys, Arg, and Pro residues, respectively. According to this,

$$\Delta\Delta G_{\text{carb}} > \Delta G_{\text{f-u}} \qquad (3)$$

when the villin headpiece unfolds, and

$$\Delta\Delta G_{\text{carb}} < \Delta G_{\text{f-u}} \qquad (4)$$

when the molecule remains in the native conformation upon carbonylation. We used our simulated data and traces given in Figure 3a−c, together with the definition of foldedness given above, to determine whether a given ensemble of structures is folded or unfolded. For example, one inequality, derived using core compactness as a reporter of the state of the villin headpiece (folded or unfolded), is

$$5\Delta\Delta G_K + \Delta\Delta G_P < \Delta G_{\text{f-u}} \qquad (5)$$

Using the same approach, we derived 19 additional inequalities using rmsd, core compactness, and α-helicity criteria (see SI for details). This system of inequalities has no unique solution, partly because different structural measures that were used are not necessarily mutually consistent when it comes to the definition of foldedness. However, it was possible to find an approximate solution using numerical approaches (see SI for details). Average values over 1136 variations of $\Delta\Delta G_K$, $\Delta\Delta G_R$, and $\Delta\Delta G_P$ values that fulfilled the maximal 16 out of 20 inequalities are $\Delta\Delta G_K = 0.13\Delta G_{\text{f-u}}$, $\Delta\Delta G_R = 0.54\Delta G_{\text{f-u}}$ and $\Delta\Delta G_P = 0.13\Delta G_{\text{f-u}}$. In other words, carbonylation of Arg15 has by far the largest effect on the protein's stability, while the effect of carbonylating Lys and Pro residues is markedly smaller.

Effect of Carbonylation on Protein Aggregability. Since hydrophobicity is one of the most dominant properties that determines aggregability, we have compared the hydrophobicity (related to aggregability) of native Lys, Arg, and Pro residues and their carbonylated counterparts, Asa and Gsa, using the MHP approach[42] (Figure 5). MHP values correlate
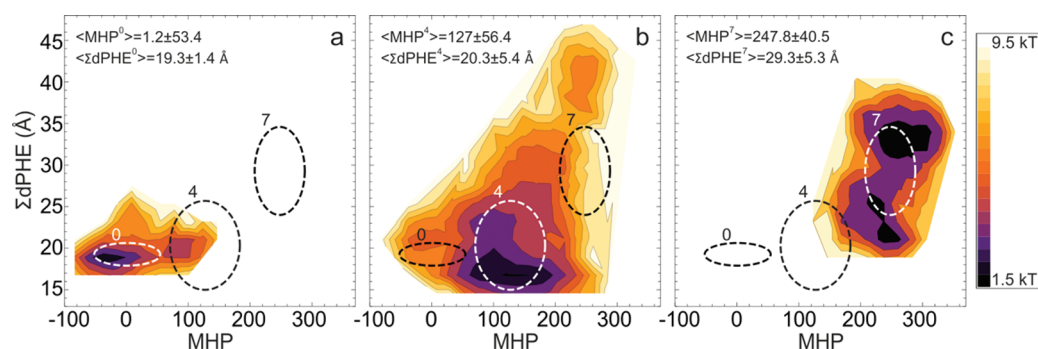
Figure 6. Two-dimensional projections of the free-energy surface as a function of MHP and hydrophobic core compactness (as captured by the sum of the distances between the three core phenylalanines) as calculated for (a) the native ensemble (five simulations), (b) the ensemble containing all combinations with four carbonylations events (35 simulations), and (c) the fully carbonylated ensemble (five simulations). Only the last 25 ns of each simulation were used to calculate the free-energy maps. The ellipses are centered at the average values of MHP and ΣdPhe for the three populations, with the major and minor semiaxes equal to the standard deviations of the distributions. The relative free energies were calculated as $F = -kT \ln P$, where P is the probability of occurrence of a given state, as seen in our simulations.

well with octanol/water partition factors for amino acid side-chain analogues (Figure 5a), a widely used measure of amino acid hydrophobicity. In agreement with our TI result discussed above, the intrinsic hydrophobicity of Lys and Arg residues increases dramatically upon carbonylation. In terms of MHP weights, converting Lys to Asa or Arg to Gsa is similar to mutating them to highly hydrophobic leucine and valine, respectively (Figure 5b). In fact, the similarity between Asa and Leu, and Gsa and Val, respectively, extends to other basic physicochemical properties as well (SI Table S1). The effect of carbonylating Pro to Gsa is less dramatic, albeit still resulting in a net increase in hydrophobicity (Figure 5b). In addition to increasing hydrophobicity, carbonylation also results in charge neutralization when it comes to Lys and Arg residues, and both of these effects potentially increase the intrinsic aggregability of the affected proteins.[43−45]

To analyze this more closely, we have used the MHP approach to characterize the hydrophobicity of a protein surface and compare it with the degree of the protein's structural integrity. In particular, we have studied the projections of the free energy surface for native and fully carbonylated forms of villin head-piece, as well as for an intermediate including all combinations with four carbonylated residues, using MHP, phenylalanine core compactness, and rmsd from the experimental NMR structure as order parameters (Figure 6 and SI Figure S2). Note that free energy projections using MHP as an order parameter were first used by Polyansky and Zagrovic in the context of protein phosphorylation (manuscript in preparation). These free energy maps clearly demonstrate that surface hydrophobicity and the fraction of partially unfolded, aggregation-prone structures increase as the level of carbonylation increases. However, this analysis also shows that compact and structurally native-like, yet significantly hydrophobic structures are present to a high degree in ensembles with moderate levels of carbonylation (Figure 6b). This is a direct consequence of the fact that Asa and Gsa are significantly more hydrophobic compared with Lys and Arg, respectively, which at moderate levels of carbonylation may not be enough to unfold the molecules but is enough to increase their surface hydrophobicity. On the other hand, only unfolded structures with high surface hydrophobicity are populated in the fully carbonylated ensemble (Figure 6c and SI Figure S2). Taken together, these results suggest that carbonylation increases
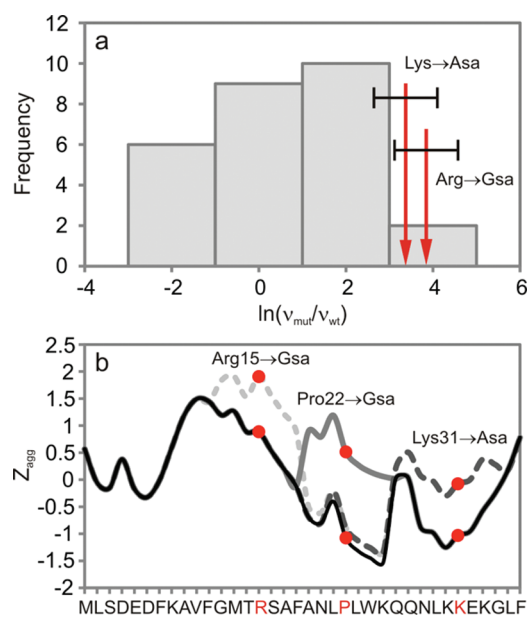

Figure 7. Aggregability versus carbonylation. (a) Distribution of changes in aggregability upon 27 single-point mutations in six known proteins causing amyloidogenic disease and in one model protein. Red arrows represent the estimation of aggregability change upon Lys (left arrow, $\ln(\nu_{mut}/\nu_{wt}) = 3.37 \pm 0.73$) and Arg (right arrow, $\ln(\nu_{mut}/\nu_{wt}) = 3.85 \pm 0.73$) carbonylation, calculated using eq 5. The error bars show standard deviation around zero calculated for the change in α-helix and β-sheet propensities for all proteins used by Chiti et al.[43] (b) Site-specific intrinsic aggregation propensity, Z_{agg}, for the native (solid black line) and three carbonylated villin headpiece molecules (only single carbonylation hits: Arg15 to Gsa, the lightest gray dotted line; Pro22 to Gsa, dark gray solid line; and Lys31 to Asa, the darkest gray dashed line).

surface hydrophobicity through unfolding, but it does so also just by itself by modifying the physicochemical properties of the affected residues, which consequently may lead to an increase in aggregability.

To put this on a quantitative footing, we have used the model of Chiti et al.[43] to estimate the expected change in the aggregation rate of an unstructured protein as a consequence of a single-point "mutation" induced by carbonylation (i.e., Lys to Asa or Arg to Gsa conversion). Changes in aggregability for these two

"mutations" are $\ln(\nu_{mut}/\nu_{wt}) = 3.37 \pm 0.73$ (Lys to Asa) and $\ln(\nu_{mut}/\nu_{wt}) = 3.85 \pm 0.73$ (Arg to Gsa), i.e., more than a 30- to 40-fold increase in the aggregation rate for a single carbonylation event, with the error bounds capturing the expected effects of the change in the intrinsic α-helix and β-sheet propensity, not modeled here. Note that this model is sequence-independent and that these changes in aggregability are applicable not just to the villin headpiece but rather to any given positively charged protein. Carbonylation increases aggregability of negatively charged proteins as well, but to a somewhat lower extent ($\ln(\nu_{mut}/\nu_{wt}) = 2.39 \pm 0.73$ for Lys to Asa, and $\ln(\nu_{mut}/\nu_{wt}) = 2.87 \pm 0.73$ for Arg to Gsa), which is still extremely significant. In Figure 7a, we compare this change in aggregability with the predicted change in aggregation rate for 27 different point mutations in seven different unstructured peptides or intrinsically disordered proteins causing diabetes and different neurodegenerative diseases (amylin, two prion peptides, α-synuclein, amyloid β-peptide, tau protein, and a model protein) examined by Chiti et al.[43] Note that the latter predictions correlate well with experimental measurements (Pearson $r = 0.85$). Strikingly, the effect of carbonylation of a single Lys or Arg residue is comparable in magnitude with the most drastic aggregation-inducing mutations in these proteins, with only 2 out of 27 such mutations having a greater effect on protein aggregability than either one of these carbonylation-induced conversions.

Finally, to probe the effect of carbonylation on protein aggregability directly from the native state, we have used the model of Vendruscolo et al.,[44] replacing Asa and Gsa with amino acids that most closely match them in terms of charge and MHP hydrophobicity, i.e., Leu and Val, respectively. These approximations notwithstanding, it appears that carbonylation of a single residue indeed significantly increases intrinsic aggregability of villin headpiece even in the absence of any major unfolding (Figure 7b). For example, the carbonylation of Pro22 increases its Z_{agg} score, i.e., the aggregability of the region around this residue, from an unfavorable -1 to a highly favorable 0.5. In this model, the Z_{agg} score reflects the combined effect of the intrinsic aggregation propensity of a given sequence and its tendency to be structurally unstable in the native state. Our analysis (Figure 7b and SI Figure S3), showing a dramatic increase in Z_{agg} upon carbonylation throughout the villin sequence, suggests that both of these factors change significantly upon carbonylation in the direction that favors aggregation.

■ DISCUSSION

This study presents the first-ever computer simulation effort to analyze the effects of carbonylation on protein structure. We have shown that the structure of the completely carbonylated villin headpiece domain unfolds in the course of simulated trajectories. Further analysis showed that extent of destabilization is site-specific, and that the disruption of both specific stabilizing elements (salt bridges and proline kinks) together with surface hydrophobicity change upon carbonylation of a large number of residues is required for protein unfolding. Note, however, that the effect of surface hydrophobicity change cannot be disentangled from other specific effects: for example, the carbonylation event, which disrupts salt-bridge formation, invariably also changes the hydrophobicity of the involved residue. Carbonylation of the proline residue is the only case where such a binary effect is avoided, as the native and the carbonylated forms of the residue exhibit similar levels of hydrophobicity.

The level of carbonylation at which the villin headpiece, a marginally stable protein,[46] unfolds in our simulations (ca. 14% of all residues) is ∼3 orders of magnitude greater than the average level of carbonylation which correlates with cellular senescence and death in vivo (ca. 0.025% of all residues).[3,10,13] This suggests that typical cellular levels of carbonylation are not likely to disrupt protein structure, which is not in contradiction with experimental studies.[16,18] If carbonylation does not induce major structural destabilization of a typical protein and thus lead to aggregation, as widely assumed,[17,47] how does it then promote the formation of cytotoxic aggregates? First, several studies suggest that some proteins are more susceptible to carbonylation than others,[9,48−50] which may result in a situation where these proteins are completely carbonylated, while the majority of other proteins are still intact, in agreement with our results. Second, newly synthesized proteins in an old organism might be misfolded or unfolded due to, for example, accumulated DNA mutations and would therefore be more susceptible to carbonylation, as has been shown by Dukan et al.[51] Finally, modifications of specific, functionally important residues may abolish the function of a protein without destabilizing its structure.

We propose two novel possibilities to explain this. First, our results suggest that carbonylation just by itself may increase protein aggregability of unstructured polypeptides (Figure 7a). In this scenario, protein unfolding arises because of either carbonylation or some other environmental factors, but importantly, carbonylation makes the unfolded molecule additionally aggregation-prone. Second, our results suggest that carbonylation may increase protein aggregability even in the absence of major unfolding (Figure 7b). From lysozyme to superoxide dismutase to prolactin, there are numerous examples of proteins which undergo amyloid formation under native conditions and without any major unfolding, simply as a consequence of local thermal fluctuations.[52] Increasing the local propensity to aggregate through carbonylation may speed this process up.

Using TI and MHP approaches, we have shown that the products of carbonylation of Lys and Arg residues are significantly more hydrophobic than these amino acids. Such an increase in hydrophobicity, together with a concomitant charge removal, significantly increases the intrinsic aggregability of proteins.[43−45] On the other hand, the aggregation propensity of proline residues is decreased by precluding β-sheet formation, and their mutations are known to lead to protein aggregation.[53] Finally, recent findings that the three quantitatively most important carbonylable residues (Arg, Lys, and Pro) in many proteins actually serve as key gatekeepers that flank aggregation-prone regions of proteins and prevent aggregation further support our suggestions.[54]

In general, the list of amyloidogenic, deposition diseases shows a strong overlap with the list of diseases in which high levels of protein carbonylation are detected, including Alzheimer's disease, Parkinson's disease, diabetes, rheumatoid arthritis, dementia with Lewy bodies, familial amyotrophic lateral sclerosis, and others.[7,9,52,55] Because of the complex pathophysiology of these diseases, the causal link between carbonylation and aggregation is not simple and is still to be fully explored. In fact, not all aggregation-prone proteins associated with these diseases are at the same time highly carbonylable and vice versa. However, there are a number of important examples where this is precisely so. For example, ubiquitin C-terminal hydrolase L1 (UCH-L1), which features critically in sporadic variants of Alzheimer's and Parkinson's diseases, forms aggregates but is also highly

carbonylable.[7,56] Superoxide dismutase 1 in familial amyotrophic lateral sclerosis,[7,52] human β2-microglobulin in end-stage renal failure,[52,57] and β-actin and α/β-tubulin in multiple sclerosis[58] are further examples of such proteins. Importantly, our results provide a novel explanation that links aggregation and carbonylation in these systems, even at relatively low levels of carbonylation typically seen *in vivo*. One way to test this connection experimentally would be to use Lys to Leu, Arg to Val, and Pro to Val point mutations for high-resolution, site-specific studies of the structural and functional consequences of carbonylation, as these mutations, according to our results, quantitatively match the effects of hydrophobicity increase and charge removal in the course of carbonylation. Such studies, hand in hand with further computational analyses, may provide an atomistic picture behind protein carbonylation and its cellular consequences, protein aggregation, and cell senescence.

■ ASSOCIATED CONTENT

ⓢ Supporting Information. Figures showing TI, MHP vs rmsd free energy surfaces, and Z_{agg} results; table of physicochemical properties of Asa, Gsa, Leu, and Val; parametrization of Asa and Gsa residues; and details of TI calculation. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author
bojan.zagrovic@univie.ac.at

■ ACKNOWLEDGMENT

■ REFERENCES

(1) Stadtman, E. R. *Science* **1992**, *257*, 1220.
(2) Berlett, B. S.; Stadtman, E. R. *J. Biol. Chem.* **1997**, *272*, 20313.
(3) Levine, R. L.; Stadtman, E. R. *Exp. Gerontol.* **2001**, *36*, 1495.
(4) Nystrom, T. *EMBO J.* **2005**, *24*, 1311.
(5) Requena, J. R.; Levine, R. L.; Stadtman, E. R. *Amino Acids* **2003**, *25*, 221.
(6) Maisonneuve, E.; Ducret, A.; Khoueiry, P.; Lignon, S.; Longhi, S.; Talla, E.; Dukan, S. *PLoS One* **2009**, *4*, e7269.
(7) Dalle-Donne, I.; Giustarini, D.; Colombo, R.; Rossi, R.; Milzani, A. *Trends Mol. Med.* **2003**, *9*, 169.
(8) Dalle-Donne, I.; Rossi, R.; Colombo, R.; Giustarini, D.; Milzani, A. *Clin. Chem.* **2006**, *52*, 601.
(9) Dalle-Donne, I.; Aldini, G.; Carini, M.; Colombo, R.; Rossi, R.; Milzani, A. *J. Cell. Mol. Med.* **2006**, *10*, 389.
(10) Levine, R. L. *Free Radical Biol. Med.* **2002**, *32*, 790.
(11) Grune, T.; Jung, T.; Merker, K.; Davies, K. J. *Int. J. Biochem. Cell Biol.* **2004**, *36*, 2519.
(12) Davies, K. J. *Biochimie* **2001**, *83*, 301.
(13) Krisko, A.; Radman, M. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 14373.
(14) Davies, K. J. A.; Delsignore, M. E. *J. Biol. Chem.* **1987**, *262*, 9908.
(15) Gardner, P. R.; Nguyen, D. D. H.; White, C. W. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 12248.
(16) Lasch, P.; Petras, T.; Ullrich, O.; Backmann, J.; Naumann, D.; Grune, T. *J. Biol. Chem.* **2001**, *276*, 9492.
(17) Bota, D. A.; Davies, K. J. A. *Nat. Cell Biol.* **2002**, *4*, 674.
(18) Le, H. T.; Chaffotte, A. F.; Demey-Thomas, E.; Vinh, J.; Friguet, B.; Mary, J. *J. Mol. Biol.* **2009**, *393*, 58.
(19) Perez, V. I.; Buffenstein, R.; Masamsetti, V.; Leonard, S.; Salmon, A. B.; Mele, J.; Andziak, B.; Yang, T.; Edrey, Y.; Friguet, B.; Ward, W.; Richardson, A.; Chaudhuri, A. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 3059.
(20) Wu, W.; Wu, X. J.; Hu, Y. F. *LWT-Food Sci. Technol.* **2010**, *43*, 133.
(21) van Gunsteren, W. F.; Bakowies, D.; Baron, R.; Chandrasekhar, I.; Christen, M.; Daura, X.; Gee, P.; Geerke, D. P.; Glattli, A.; Hunenberger, P. H.; Kastenholz, M. A.; Oostenbrink, C.; Schenk, M.; Trzesniak, D.; van der Vegt, N. F.; Yu, H. B. *Angew. Chem. Int. Ed.* **2006**, *45*, 4064.
(22) Khurana, S.; George, S. P. *FEBS Lett.* **2008**, *582*, 2128.
(23) Friederich, E.; Vancompernolle, K.; Louvard, D.; Vandekerckhove, J. *J. Biol. Chem.* **1999**, *274*, 26751.
(24) Dalle-Donne, I.; Rossi, R.; Giustarini, D.; Gagliano, N.; Lusini, L.; Milzani, A.; Di Simplicio, P.; Colombo, R. *Free Radical Biol. Med.* **2001**, *31*, 1075.
(25) Dalle-Donne, I.; Rossi, R.; Milzani, A.; Di Simplicio, P.; Colombo, R. *Free Radical Biol. Med.* **2001**, *31*, 1624.
(26) McKnight, C. J.; Matsudaira, P. T.; Kim, P. S. *Nat. Struct. Biol.* **1997**, *4*, 180.
(27) Duan, Y.; Kollman, P. A. *Science* **1998**, *282*, 740.
(28) Zagrovic, B.; Snow, C. D.; Shirts, M. R.; Pande, V. S. *J. Mol. Biol.* **2002**, *323*, 927.
(29) Zagrovic, B.; Snow, C. D.; Khaliq, S.; Shirts, M. R.; Pande, V. S. *J. Mol. Biol.* **2002**, *323*, 153.
(30) Meng, J.; Vardar, D.; Wang, Y.; Guo, H. C.; Head, J. F.; McKnight, C. J. *Biochemistry* **2005**, *44*, 11963.
(31) Lei, H.; Wu, C.; Liu, H.; Duan, Y. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 4925.
(32) Cornilescu, G.; Hadley, E. B.; Woll, M. G.; Markley, J. L.; Gellman, S. H.; Cornilescu, C. C. *Protein Sci.* **2007**, *16*, 14.
(33) Gronwald, W.; Hohm, T.; Hoffmann, D. *BMC Bioinformatics* **2008**, *9*, 109.
(34) Schuler, L. D.; Walde, P.; Luisi, P. L.; van Gunsteren, W. F. *Eur. Biophys. J.* **2001**, *30*, 330.
(35) Schuler, L. D.; Daura, X.; van Gunsteren, W. F. *J. Comput. Chem.* **2001**, *22*, 1205.
(36) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. *Interaction models for water in relation to protein hydration*; Reidel: Dordrecht, 1981.
(37) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Dinola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.
(38) Lindahl, E.; Hess, B.; van der Spoel, D. *J. Mol. Modell.* **2001**, *7*, 306.
(39) Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577.
(40) Andricioaei, I.; Karplus, M. *J. Chem. Phys.* **2001**, *115*, 6289.
(41) Levy, R. M.; Srinivasan, A. R.; Olson, W. K.; McCammon, J. A. *Biopolymers* **1984**, *23*, 1099.
(42) Efremov, R. G.; Chugunov, A. O.; Pyrkov, T. V.; Priestle, J. P.; Arseniev, A. S.; Jacoby, E. *Curr. Med. Chem.* **2007**, *14*, 393.
(43) Chiti, F.; Stefani, M.; Taddei, N.; Ramponi, G.; Dobson, C. M. *Nature* **2003**, *424*, 805.
(44) Tartaglia, G. G.; Pawar, A. P.; Campioni, S.; Dobson, C. M.; Chiti, F.; Vendruscolo, M. *J. Mol. Biol.* **2008**, *380*, 425.
(45) Dobson, C. M. *Nature* **2003**, *426*, 884.
(46) Godoy-Ruiz, R.; Henry, E. R.; Kubelka, J.; Hofrichter, J.; Munoz, V.; Sanchez-Ruiz, J. M.; Eaton, W. A. *J. Phys. Chem. B* **2008**, *112*, 5938.

(47) Pan, J. C.; Yu, Z. H.; Hui, E. F.; Zhou, H. M. *Biochem. Cell. Biol.* **2004**, *82*, 361.

(48) Das, N.; Levine, R. L.; Orr, W. C.; Sohal, R. S. *Biochem. J.* **2001**, *360*, 209.

(49) Jana, C. K.; Das, N.; Sohal, R. S. *Arch. Biochem. Biophys.* **2002**, *397*, 433.

(50) Choi, J.; Rees, H. D.; Weintraub, S. T.; Levey, A. I.; Chin, L. S.; Li, L. *J. Biol. Chem.* **2005**, *280*, 11648.

(51) Dukan, S.; Farewell, A.; Ballesteros, M.; Taddei, F.; Radman, M.; Nystrom, T. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 5746.

(52) Chiti, F.; Dobson, C. M. *Nat. Chem. Biol.* **2009**, *5*, 15.

(53) Williams, A. D.; Portelius, E.; Kheterpal, I.; Guo, J. T.; Cook, K. D.; Xu, Y.; Wetzel, R. *J. Mol. Biol.* **2004**, *335*, 833.

(54) Rousseau, F.; Serrano, L.; Schymkowitz, J. W. *J. Mol. Biol.* **2006**, *355*, 1037.

(55) Chiti, F.; Dobson, C. M. *Annu. Rev. Biochem.* **2006**, *75*, 333.

(56) Choi, J.; Levey, A. I.; Weintraub, S. T.; Rees, H. D.; Gearing, M.; Chin, L. S.; Li, L. *J. Biol. Chem.* **2004**, *279*, 13256.

(57) Michelis, R.; Sela, S.; Kristal, B. *Nephrol. Dial. Transplant.* **2005**, *20*, 1963.

(58) Bizzozero, O. A. In *Handbook of neurochemistry and molecular neurobiology. Brain and spinal cord trauma*, 3rd ed.; Banik, N. L., Ray, S. K., Lajtha, A., Eds.; Springer: New York ; London, 2009, Chapter 23.

7024

dx.doi.org/10.1021/ja110577e |*J. Am. Chem. Soc.* 2011, 133, 7016–7024