BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

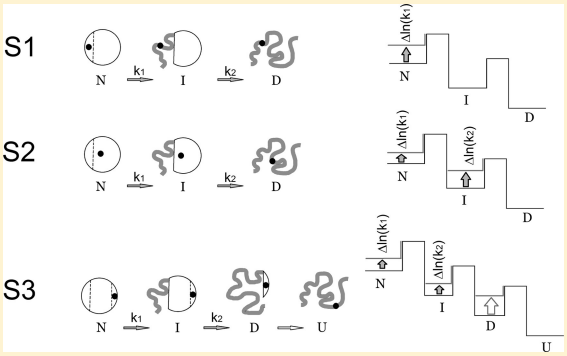
Sequential Melting of Two Hydrophobic Clusters within the Green Fluorescent Protein GFP-cycle3

Tatiana N. Melnik,[†] Tatiana V. Povarnitsyna,[†] Anatoly S. Glukhov,[†] Vladimir N. Uversky,*[,‡] and Bogdan S. Melnik*[,†]

[†]Institute of Protein Research, Russian Academy of Sciences, 142290 Pushchino, Moscow Region, Russia

[‡]Department of Molecular Medicine, College of Medicine, University of South Florida, Tampa, Florida 33612, United States

ABSTRACT: The analysis of the three-dimensional structure of green fluorescent protein (GFP-cycle3) revealed the presence of two well-defined hydrophobic clusters located on the opposite sides of the GFP β-can that might contribute to the formation of partially folded intermediate(s) during GFP unfolding. The microcalorimetric analysis of the nonequilibrium melting of GFP-cycle3 and its two mutants, I14A and I161A, revealed that due to the sequential melting of the mentioned hydrophobic clusters, the temperature-induced denaturation of this protein most likely occurs in three stages. The first and second stages involve melting of a smaller hydrophobic cluster formed around the residue I161, whereas a larger hydrophobic cluster (formed around the residues I14) is melted only at the last GFP-cycle3 denaturation step or remains rather structured even in the denatured state.



The GFP-cycle3 (GFP-c3) mutant is a variant of the green fluorescent protein, which was created by Crameri et al. by random mutagenesis with PCR.[1] The GFP-c3 variant has three mutations, F99S, M153T, and V163A, located on or near the surface of the protein molecule and matures efficiently in *E. coli* at 37 °C.[1] Since this GFP variant has the same chromophore and the same spectral parameters as the wild-type protein, it is now widely used in a variety of applications.[2,3] Furthermore, the GFP-c3 crystal structure and folding behavior are essentially identical to those of the wild-type protein.[4,5] In fact, crystallographic structures of the wild-type GFP, its enhanced mutants (ECFP, 1C4F.ent, and EYFP), and the GFP-c3 variant revealed that these proteins resemble an 11 stranded β-can wrapped around a single central helix, in the middle of which is the chromophore.[6−9] The cylinder has a diameter of about 30 Å and a length of about 40 Å.[7]

At present, there are numerous papers reporting the use of GFP and its numerous variants as unique biosensors for detecting protein−protein interactions, cotranslational protein folding, and localization of different structures within the cell.[2,3] However, information regarding the GFP conformational stability and folding/unfolding mechanisms is quite scarce. GFP is known to possess a high conformational stability under a variety of conditions, including a treatment with detergents,[10,11] proteases,[12] GuHCl,[5,13] and temperature.[13,14] GFP is a rather difficult subject for such unfolding−refolding studies, since the formation and unfolding of its native structure *in vitro* is an intricate sequential process. In fact, some authors suggested that the unfolding of the GFP native structure at pH6 and 25 °C can the half-life of several weeks.[15] Furthermore, GFP unfolding and refolding typically includes several fairly prolonged stages, during which the protein might undergo aggregation. For example, refolding of the pH-unfolded GFP occurs in four slow stages,[16] whereas the GuHCl-induced unfolding occurs in two[5,17] to four stages, with four stages being observed for a more stable GFP variant, superfolder.[18]

Finally, the hydrogen/deuterium (H/D) NMR exchange rates of more than 157 assigned amide protons were used to evaluate the dynamics and stability of GFP at atomic resolution.[15] A complete assignment of the NMR resonances for the backbone (^{13}C, ^{15}N, and ^{1}H) of GFP[19,20] enabled the use of the H/D exchange technique to probe the stability and folding of the protein under equilibrium conditions. This analysis revealed that the slowest exchanging residues in GFP are clustered together primarily on one face of the β-can. However, at one end of the can they do form a ringlike structure. It is likely that these residues exchange so slowly because they remain involved in hydrogen-bonded secondary structure in the intermediate state.[15]

The majority of the earlier studies on the GFP unfolding−refolding were conducted using a set of spectroscopic methods, such as fluorescence of tryptophane residues, fluorescence of GFP chromophore, and circular dichroism. Although this approach is a useful tool for measuring the kinetic parameters of folding/unfolding and for determining the folding/unfolding rate constants, it does not yield sufficient information on the structural peculiarities of intermediate states. In our previous

Received: May 1, 2011
Revised: August 5, 2011
Published: August 9, 2011

paper,[21] temperature-induced denaturation of GFP-c3 was analyzed by differential scanning microcalorimetry. This analysis revealed that, at heating rates exceeding 0.25 K/min, GFP melting is the nonequilibrium process. It was also demonstrated that the GFP heat denaturation can be described by a sequential irreversible model. Fitting experimental data by the two consecutive irreversible denaturation stages model provided reliable kinetic parameters of the GFP-c3 denaturation, such as activation energy and the unfolding rate constants, as well as important energetic parameters, such as melting enthalpy of every stage of GFP-c3 denaturation.[21] The observation of two melting stages suggested the GFP-c3 variant might contain two hydrophobic cores, which are melted sequentially at subsequent denaturation steps.

The intrinsic disorder phenomenon is introduced to describe the biologically active proteins or protein regions which remain unstructured, or incompletely structured, under physiological conditions.[22−52] Since the amino acid compositions, sequence complexity, hydrophobicity, aromaticity, charge, flexibility, etc., of amino acid sequences of these intrinsically disordered proteins (IDP) and intrinsically disordered regions (IDR) are different from sequences of structured globular proteins and their domains, various computational tools were developed for accurate disorder prediction.[53−55] Although these tools are typically used for reliable prediction of IDRs and IDPs, they can also be utilized for the evaluation of structural flexibility of ordered proteins. In fact, in ordered proteins with overall low disorder propensity scores, the predicted local increase in the intrinsic disorder predisposition could be correlated with the increased mobility of a given region.[56−58]

In this study, the calorimetric and computational analyses of the GFP-c3 variant and its mutants further revealed a peculiar unfolding mechanism of GFP and yielded information regarding the sequence of events during the thermal denaturation of this interesting protein.

MATERIALS AND METHODS

Protein Expression and Isolation. The GFP-c3 gene (pBAD-GFP-cycle3 produced by Maxygen) was transferred into the pET-28a vector. Plasmids with the mutant GFP-c3 genes were constructed by a standard PCR technique using appropriate primers and a pET-28a vector as a template. The DNA sequences of all constructs were confirmed by the DNA sequence analysis.

GFP-c3 was expressed and isolated as described elsewhere.[5] The purity of the isolated protein was checked by SDS polyacrylamide gel electrophoresis. Since the ratio of the absorbance values at 395 and 280 nm is equal to 1.1−1.2 for pure GFP-c3, this 395/280 ratio is typically used as a GFP-c3 purity index.[11] The final GFP-c3 sample prepared in this study had a 395/280 ratio of 1.1 or higher. Mutant proteins with substitutions I14A and I161A were isolated and purified using the same techniques. The intensity of chromophore fluorescence in both mutants was similar to that of GFP-c3 protein.

Calorimetry. Calorimetric measurements were performed using a precision scanning microcalorimeter SCAL-1 (Scal Co. Ltd., Russia) with 0.33 mL glass cells, at a scanning rate of 0.25, 0.5, and 1.0 K/min and under the pressure of 2.5 atm.[59] The experiments were performed in 25 mM sodium phosphate buffer at pH 7.2. The protein concentrations in the experiment varied from 0.5 to 1.0 mg/mL. The experimental calorimetric profiles were corrected for the calorimetric baseline, and the molar partial heat capacity functions were calculated using a standard approach. The excess heat capacity was evaluated by subtraction of the linearly extrapolated initial and final heat capacity functions with correction for the difference of these functions by using a sigmoid baseline.[60] A characteristic value of the partial specific volume for typical lobular proteins (0.73 cm^3/g) was chosen arbitrarily, since it does not influence the calculated excess heat capacity.

Estimation of the Rate Constants and Activation Energy from Melting Curves. A detailed explanation and justification of a choice of protein denaturation model described below are given elsewhere.[21] The curves of the dependence of excess heat capacity versus temperature were analyzed using the following irreversible denaturation model:

$$\mathrm{N} \xrightarrow{k_1} \mathrm{I} \xrightarrow{k_2} \mathrm{D} \tag{1}$$

where N, I, and D are native intermediate and denatured states of the protein; k_1 and k_2 are the rate constants of first and second unfolding steps.

The effects of the nonequilibrium experimental conditions on the peculiarities of protein thermal denaturation were described in several studies.[61−65] On the basis of these formalisms, basic regularities determining the position and shape of the heat absorption peak in calorimetric experiments were revealed. When protein melting can be represented as two irreversible sequential steps, the heat absorption peak is described by the following equation:[66]

$$C_p^{\mathrm{ex}} = \frac{\Delta H_1 k_1}{\nu} \exp\left(-\frac{1}{\nu}\int_{T_0}^{T} k_1\, \mathrm{d}T\right) + \frac{\Delta H_2 k_2}{\nu^2} \times \exp\left(-\frac{1}{\nu}\int_{T_0}^{T} k_2\, \mathrm{d}T\right) \times \int_{T_0}^{T}\left[k_1 \exp\left(\frac{1}{\nu}\int_{T_0}^{T}(k_2 - k_1)\, \mathrm{d}T\right)\right]\mathrm{d}T \tag{2}$$

where ΔH_1 and ΔH_2 are changes in enthalpy for the first and second steps; $\nu = \mathrm{d}T/\mathrm{d}t$ is the scanning rate; T and T_0 are the absolute temperature and the temperature of the beginning of the heat absorption curve; k_1 and k_2 are the rate constants for the first and second steps that can be expressed using the activation energies of each stage (ΔE_1, ΔE_2) by the equations

$$k_1 = \exp[-\Delta E_1/R(1/T - 1/T_1^*)] \text{ and}$$
$$k_2 = \exp[-\Delta E_2/R(1/T - 1/T_2^*)] \tag{3}$$

where T_1^* and T_2^* are the temperature parameters equal to the temperature at which the corresponding rate constant is 1 min^{-1}.

Equation 2 is used for approximation of the heat absorption curves obtained experimentally, and the six parameters (ΔH_1, ΔH_2, ΔE_1, ΔE_2, T_1^*, and T_2^*) are chosen so that the calculated heat absorption curve would be perfectly concurrent with the experimental one. Therefore, the fit of experimental data within the frames of this approximation provides important kinetic information, the activation energies, and the temperature parameters T^* of the two steps of the protein thermal denaturation. The fitting of the experimental curves was done using the IgorPro software (WaveMetrics, Inc.)

Table 1. Combined Fitting Parameters of Heat Denaturation Curves of GFP-c3 and Its Mutants (I14A, I161A)[a]

	scan rate, K/min	$\Delta H_1 \pm 5$, kJ/mol	$\Delta H_2 \pm 6$, kJ/mol	$\Delta E_1 \pm 4$, kJ/mol	$\Delta E_2 \pm 2$, kJ/mol	$T_1^* \pm 0.2$, K	$T_2^* \pm 0.2$, K
GFP-c3	1	357	611	311	471	373.4	367.2
	0.5	269	542				
	0.25	278	534				
I14A	1	244	645	311	471	371.1	365.7
	0.5	199	567				
	0.25	201	568				
I161A	1	298	396	311	471	364.7	358.2
	0.5	357	337				
	0.25	372	359				

[a]Fitting errors are given. See experimental errors in Materials and Methods.

Experimental and Fitting Errors. Table 1 provides errors of the combined fitting and shows how well the IgorPro software (WaveMetrics, Inc.) finds the only minimum among all the parameters (for three curves). The experimental error of the melting enthalpy determination is greater than the fitting error, almost equaling ±50 kJ/mol. However, the error of the determination of the rate constants is much lower. This is because the ΔE_1, ΔE_2, T_1^*, and T_2^* parameters are associated only with the shape of the melting curves (eqs 2 and 3). If the height of every melting peak is changed arbitrarily (which can occur, e.g., due to the error in the determination of protein concentration), this would affect the enthalpies (ΔH_1, ΔH_2) but would not affect the shape of the melting curves, i.e., the calculated activation energy values (ΔE_1, ΔE_2), temperature parameters (T_1^*, T_2^*), and rate constants (k_1, k_2). Furthermore, the calculation accuracy of the logarithm of rate constants is ±0.3 for $\ln(k_1)$ and ±0.2 for $\ln(k_2)$ and depends on the noisiness of the melting curves, rather than on the height of melting peaks. The experimental error is about ±20 kJ/mol for the activation energy and ±1 K for the temperature parameters T^*.

Determination of the Amount of Contacts for Amino Acid Residues. To analyze the structure and calculate the amount of contacts for every amino acid residue in the GFP-c3 protein, we used RasWin (v.2.7 Herbert J.Bernstein) and DSSP[67] software. The structure of protein PDB ID 1QYQ was analyzed. When calculating the amount of contacts, two atoms were accepted to be in contact if the distance between them (r) did not exceed 6 Å. The amount of residue−residue and atom−atomic contacts was calculated at r = 5 Å and r = 7 Å, with the conclusions on the choice of amino acid residue for the following substitutions remaining unchanged.

Evaluation of Protein Dynamics via Intrinsic Disorder Calculations. Here, the intrinsic disorder predictor PONDR FIT[68] was used to evaluate the propensities of GFP-c3 for intrinsic disorder, since such propensities reflect the encoding of the dynamic tendencies in the primary structures of studied proteins and can therefore report the dynamic behavior of a given protein in solution.[56−58] PONDRFIT is a meta-predictor[68] that combines six individual predictors, which are PONDR VL-XT,[69] VSL2,[70] VL3,[71] FoldIndex,[72] IUPred,[73] and TopIDP.[74] This meta-predictor is moderately more accurate than each of the component predictors.[68] For an amino acid sequence, PONDRFIT outputs a prediction for the intrinsic disorder propensity in the [0, 1] range per residue. These outputs are then compared to a threshold (we used the default threshold 0.5) and residues with a prediction value greater than the threshold were predicted to be intrinsically disordered. The local increase in the intrinsic disorder tendency for an ordered protein with overall low disorder propensity scores is expected to correlate with the increased mobility of the studied region.

ANCHOR Analysis. Potentially foldable sites in the dynamic regions of GFP-c3 were identified by the ANCHOR algorithm.[75,76] This approach relies on the pairwise energy estimation approach developed for the general disorder prediction method IUPred[73,77] and is based on the hypothesis that long dynamic (or disordered) regions contain localized potential foldable sites, ANCHORs, that cannot form enough favorable intrachain interactions to fold on their own but are likely to gain stabilizing energy by interacting with a protein partner.[75,76]

RESULTS AND DISCUSSION

Analysis of the GFP-c3 Structure. Figure 1 represents the 3D structure of GFP-c3, where the localizations of the Cα-atoms of hydrophobic amino acid residues are shown by balls. Analysis of this figure reveals that hydrophobic amino acid residues are grouped in two clusters. The first group is a large cluster that includes 25 hydrophobic residues. It is centered at the β-sheet shown in red, which is located in the 11−48 region and consists of three β-strands (strands 1−3). In addition, this cluster includes hydrophobic residues located in strands 5, 6 (shown in yellow), and 11 (shown in cyan). The second hydrophobic cluster has smaller dimensions (12 residues), being mainly located within the β-sheet located in the 144−187 region (shown in green). This cluster includes hydrophobic residues from the β-strands 4, 7, 8, and 9. As seen from Figure 1B, these clusters are positioned at the opposite sides of the GFP β-can.

Figure 1D shows residue−residue contact map built based on the GFP-c3 crystal structure (1QYQ). As a rule, the residue−residue contacts in this protein depend on the mutual positions of corresponding β-strands, since most of the dots representing contacts in the Figure 1D are grouped in the extended bands. Such grouping of the residue−residue contacts into dense extended bands is characteristic for the contacts between neighboring β-strands. The residue−residue contacts that belong to the aforementioned large and small hydrophobic clusters are contoured by the red and greed ovals, respectively. The hatched region in the Figure 1D represents the residue−residue contacts that are formed within the central α-helix of the GFP-c3. Figure 1D shows that the hydrophobic clusters are separated by the central helix, that the intrahelical residue−residue contacts within the central helix are uniformly distributed, and that the interactions of all the GFP-c3 β-strands with this α-helix are equally weak.

dx.doi.org/10.1021/bi2006674|Biochemistry 2011, 50, 7735−7744

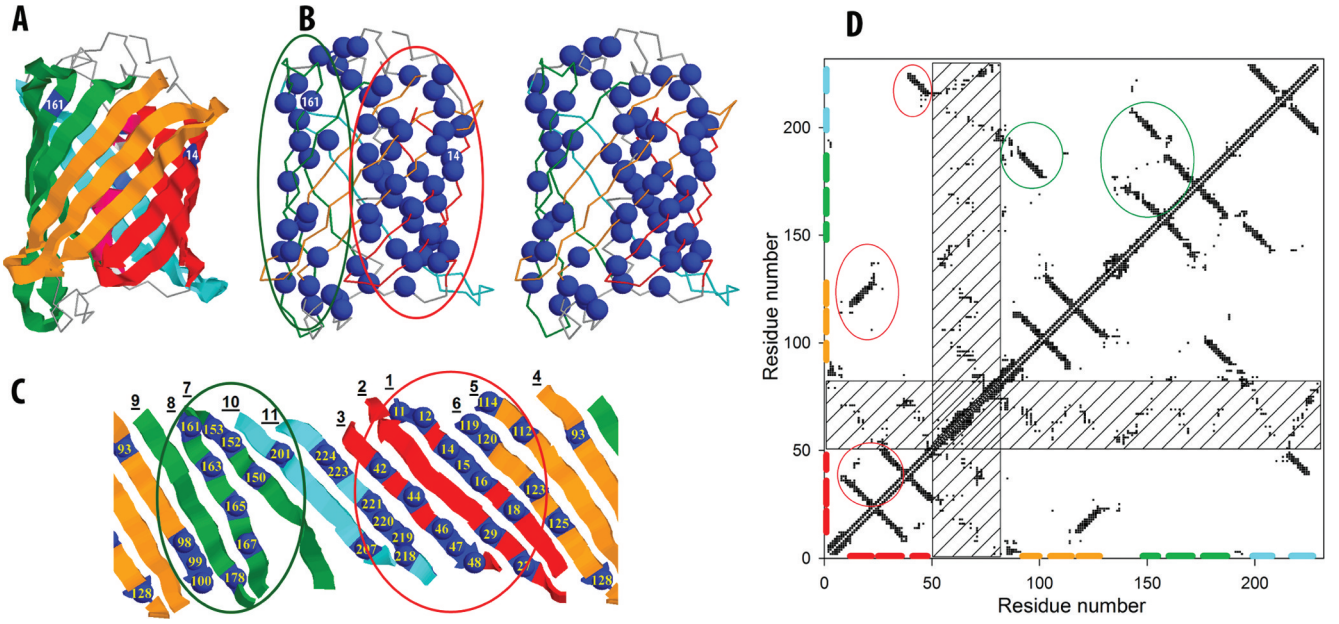


Figure 1. Structure of GFP-c3. Three-dimensional structure is shown with band (A) and wire (B, stereo view) models. (C) Relative position of β-strands (numbered by the underlined digits) of the protein without loops and the central α-helix (as viewed from the center of the protein). Blue balls show hydrophobic amino acid residues. Two hydrophobic clusters well distinguished in (B) and (C) are encircled with a red (larger cluster) and a green (smaller cluster) lines. (D) Residue−residue contact map evaluated based on the crystal structure of the GFP-c3 (PDB ID: 1QYQ). The hatched region contains contacts of the central α-helix of this protein.

Figure 2A represents the evaluation of the expected dynamic behavior of GFP-c3 based on the distribution of the predicted intrinsic disorder propensities within the primary structure of this protein. According to the PONDRFIT algorithm, GFP-c3 is expected to be mostly ordered, with the disorder scores for the majority of its residues being located below the threshold value of 0.5. However, both N- and C-termini of this protein are predicted as regions with increased flexibility. Figure 2A shows that there is generally a strong correlation between the known protein structure and predicted protein mobility. In fact, the majority of disorder spikes in the PONDRFIT plots coincide with known GFP-c3 loop regions linking elements of ordered secondary structure. Analysis of Figure 2A further shows that the residues comprising the large hydrophobic cluster are frequently located in more dynamic regions (see dark red lines in Figure 2A), whereas hydrophobic residues of the second cluster are mostly located in less mobile regions (see dark green lines in Figure 2A). Furthermore, the C-terminal β-strand 11 included in the large hydrophobic cluster is predicted as ANCHOR. In other words, this region of GFP potentially folds only as a result of interaction with the reminder of the protein.

Overall, data shown in Figure 2A suggest that the large hydrophobic cluster of GFP-c3 is more dynamic than the small cluster. This larger conformational flexibility might be used for better spatial adjustment of residues and therefore might contribute to the tighter mutual packing of residues in the large cluster. Figure 2B supports this hypothesis by showing the distribution of the residue−residue contacts within the GFP-c3 amino acid sequence. Here, the mean number of contacts in GFP is 13.3 per residue; the mean number of contacts between the hydrophobic residues is 13.8 per residue; residues in the small hydrophobic cluster (green) have 14.6 contacts per residues, whereas residues in the large hydrophobic cluster (red) are definitely packed tighter, having on average 15.6 contacts per residue.

Therefore, the analysis of the GFP-c3 3D structure, combined with the evaluation of its expected dynamic behavior, reveals that the two opposite sides of this protein have hydrophobic clusters, a large cluster containing 25 hydrophobic residues located in β-strands 1, 2 3, 5, 6, and 11, and a small cluster containing 12 hydrophobic residues located in β-strands 4, 7, 8, and 9. These two clusters are not only different in size but also differ from each other in expected dynamic behavior (with the large cluster predicted to be more mobile) and in packing density (with large cluster having more contacts per residue). Earlier, we found that, calorimetrically, melting of GFP-c3 was described by a sequential irreversible model with two consecutive irreversible denaturation stages.[21] Since there is approximately a 2-fold difference in the enthalpies of GFP-c3 melting at the first and second stages, and since there is 2-fold difference in sizes of described above hydrophobic clusters, we hypothesized that the two-stage melting of GFP-c3 is due to the sequential melting of these two clusters. Since the protein melting enthalpy is mostly due to the disruption of the van der Waals interactions (often between hydrophobic amino acids),[60,78] the first stage of the GFP-c3 structure melting could be due to the disruption of the smaller cluster, whereas the second stage could be determined by melting of the larger cluster.

Choice of Appropriate Mutations. To check the hypothesis that the two-stage melting of GFP-c3 is due to the sequential melting of two hydrophobic clusters, where melting of the small and large hydrophobic clusters corresponds to the first and second stages of the GFP-c3 thermal denaturation respectively, we analyzed the effect of the amino acid substitutions within these clusters on the individual stages of GFP-c3 thermal unfolding. To ensure that the chosen point mutations evoke a comparable destabilization of different states of the protein, the choice of the candidate positions for such substitutions was based on the following criteria: the candidate positions should have identical side chains, and the

dx.doi.org/10.1021/bi2006674 | Biochemistry 2011, 50, 7735–7744

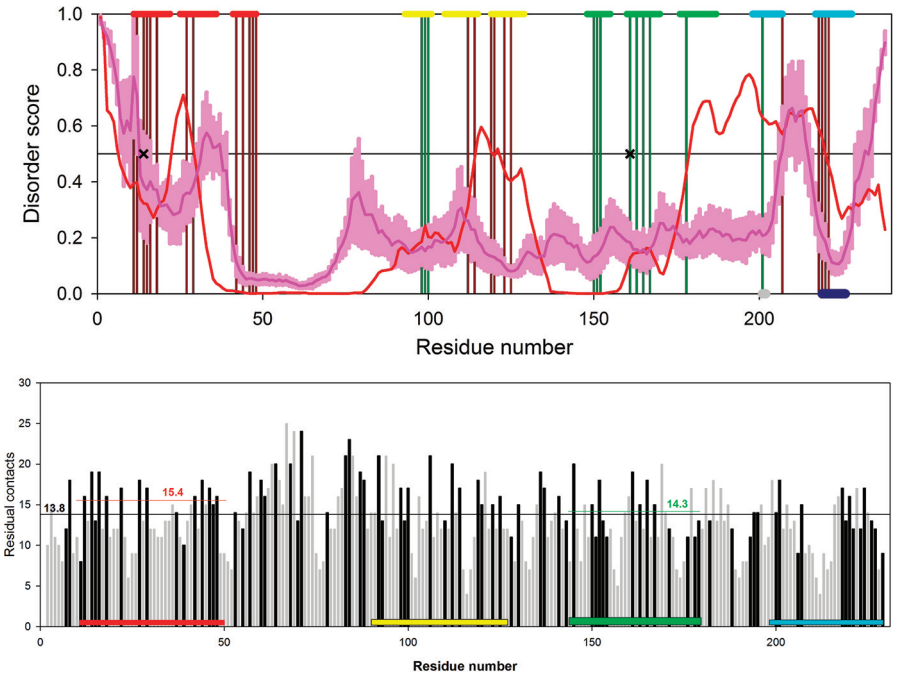


Figure 2. (A) The evaluation of the expected dynamic behavior of GFP-c3 based on the distribution of the predicted intrinsic disorder propensities within the primary structure of this protein. Results of the disorder prediction by PONDR VLXT[69] and PONDR FIT[68] are shown by red and pink lines, respectively. Pink shadow covers the distribution of errors in evaluation of disorder scores by PONDR FIT. Vertical red and green bars show locations of hydrophobic residues involved in the large and small clusters, respectively. Horizontal red, yellow, green, and cyan bars on the top of the plot show localizations of β-strands 1−3, 4−6, 7−9, and 10−11, respectively. Location of mutated residues Ile14 and Ile161 are shown by crosses. Horizontal bars at the bottom of the plot show the locations of the binding motifs predicted by ANCHOR.[75,76] Dark blue bar corresponds to the predicted disordered binding region, whereas gray bar corresponds to the filtered binding motif. (B) The distribution of the residue−residue contacts (light gray bars) within the GFP-c3 amino acid sequence. In GFP-c3, the mean number of contacts between the hydrophobic residues (dark gray bars) is 13.8 per residue; the small hydrophobic cluster (shown by green line) has 14.3 contacts per residues, and the large hydrophobic cluster (shown by red line) has 15.6 contacts per residue.

candidate residues should have a large and similar number of contacts both at residual and atomic levels. After examining the GFP-c3 structure, positions 14 and 161 were selected, since both of them are occupied by isoleucine and both I14 and I161 have a similar (and, more importantly, very large) number of contacts. In fact, at the residual level, each of these residues is involved in contact with 19 other residues. At the atomic level, there are 317 and 313 atom−atom contacts for I14 and I161, respectively. Since the average number of contacts in the protein is 14.3 per residue and the average number of atom−atom contacts is 231, the chosen positions not only have similar residues with a similar number of contacts, but the chosen residues are also highly connected. On the basis of these considerations, we designed two mutant proteins containing single I14A and I161A substitutions.

Calorimetric Studies. Figure 3 represents melting curves for the wild-type GFP-c3 and its two mutant forms at pH 7.2. The melting of each protein was analyzed at three heating rates: 1.0, 0.5, and 0.25 K/min. The fitting of the corresponding curves was done using eq 2. A detailed discussion of the molecular mechanisms of the GFP-c3 thermal melting and the validation of a hypothesis that this process can be described using a formalism of the irreversible two-stage denaturation model are given in our previous paper.[21] In particular, we demonstrated that the GFP-c3 melting process cannot be described either by a two-state irreversible denaturation model or by the Lumry−Eyring two-stage model with the fast equilibrating first step, in which folding and unfolding rate constants of the first step are high in comparison with the rate constant of the second step.[21] Furthermore, the in depth calorimetric analysis of the pH dependence of the GFP-c3 melting allowed us to systematically evaluate the activation energy values for two stages of GFP unfolding and therefore to characterize the two corresponding transition states.[21]

For each protein, a combined fitting of the three curves with similar ΔE_1, ΔE_2, T^*_1, and T^*_2 parameters was performed, which allowed us to reduce errors when determining these parameters. Figure 3 represents experimental data by the symbols and the fitting results by the lines and clearly shows that the fitting curves coincide well with the experimental curves for all three proteins under all the conditions studied. This provides additional support for the correct choice of the model describing the GFP-c3 heat denaturation (eq 1). The values of enthalpies ΔH_1 and ΔH_2, activation energies ΔE_1 and ΔE_2, and the T^* parameters calculated by fitting of the melting curves are listed in Table 1. A comparison of the melting curves of GFP-c3 and its two mutants (Figure 2) revealed that the I14A substitution had a small effect on the position of the melting peaks, whereas the I161A substitution had a rather large effect (the curves were shifted by almost 10 K). At nonequilibrium melting, this shift cannot be directly associated with the change in the protein stability, since it depends on both the protein stability and on the rates of protein unfolding. Parameters calculated from the melting curves (see Table 1) were used to estimate the rate constants of the two unfolding stages for all three proteins. Using eq 3, we plotted the

dx.doi.org/10.1021/bi2006674 | Biochemistry 2011, 50, 7735−7744

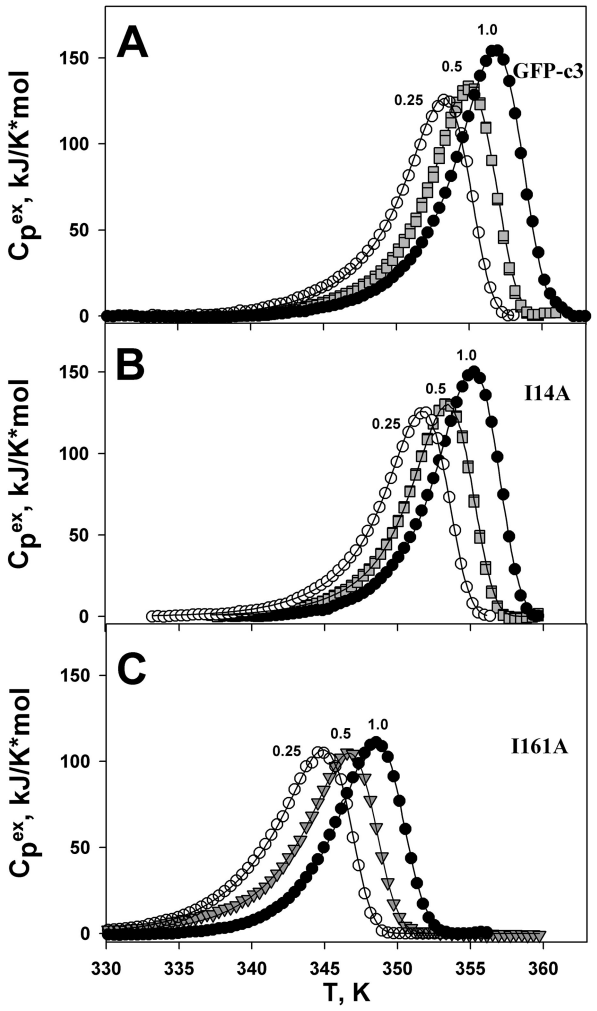


Figure 3. Melting curves of GFP-c3 (A) and its two mutant forms I14A and I161A (B, C) in 25 mM phosphate buffer, pH 7.2. In every plot, symbols show experimentally obtained dependencies of excess heat capacity versus temperature at three heating rates (●, 1 K/min; ▼, 0.5 K/min; ○, 0.25 K/min). Lines show the best fit of experimental data using a two-stage irreversible model of denaturation. Maximal deviation of the fitting curve from the experimental data is 3 kJ/(K mol).

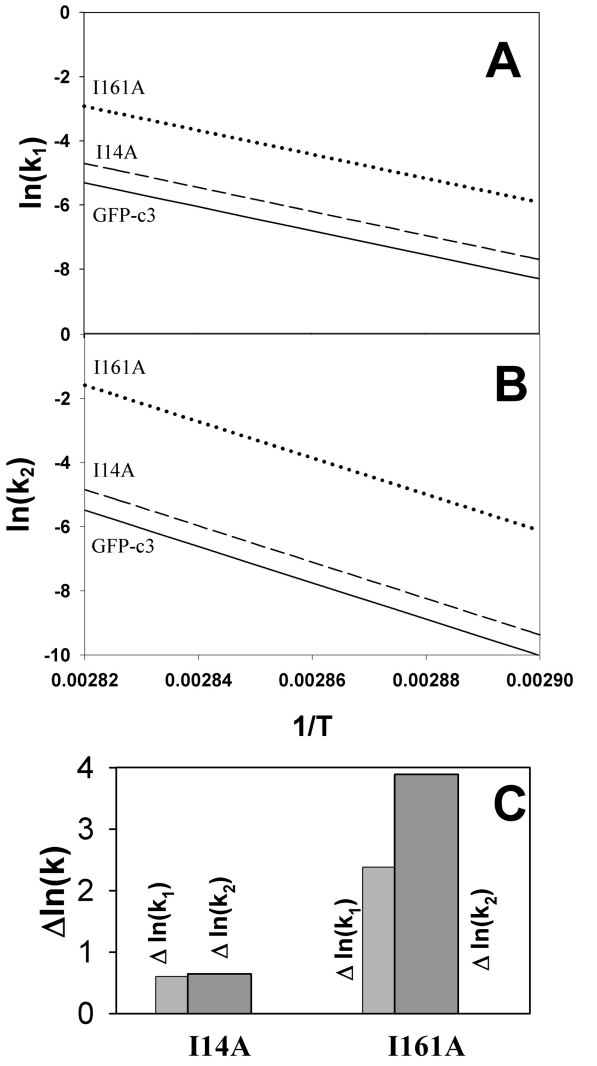


Figure 4. Dependencies of logarithms of rate constants of the first (k_1) and second (k_2) stages of GFP-c3 unfolding (A and B, respectively) versus inverse temperature calculated from melting curves of GFP-c3 (—) and its two mutant forms (I14A, - - -; I161A, ···), pH 7.2. Part C shows the $\Delta \ln(k) = \ln(k_{mutant}) - \ln(k_{GFP-c3})$ value which corresponds to the shift of the lines for mutant proteins in (A) and (B) relative to those for GFP-c3.

dependencies of logarithms of unfolding rate constants on the inverse temperatures (an Arrhenius plot). Figure 4 represents these dependencies for the first (Figure 4A) and the second (Figure 4B) unfolding stages and shows that the dependencies appear to be parallel lines. The slope of these lines is determined by the activation energy ΔE that is not affected by the mutation (see Table 1), and the observed shift between the corresponding lines is dependent on the T^* parameter. Figure 4C demonstrates the change in the logarithm of unfolding rate constants of the mutant proteins relative to the wild-type GFP-c3, calculated as $\Delta \ln(k) = \ln(k_{mutant}) - \ln(k_{GFP-c3})$. The value of the $\Delta \ln(k)$ parameter reflects the effect of each mutation on the unfolding rate of GFP-c3. Data shown in Figure 4 are consistent with the conclusion that both mutations accelerated protein unfolding. The I14A substitution had approximately similar effects on both rate constants k_1 and k_2, whereas the I161A substitution had a stronger effect on the unfolding rate of the second stage, k_2.

On the basis of the analysis of the GFP-c3 3D structure, it was expected that the I161A substitution would affect only the first unfolding step, whereas and the I14A substitution would affect the second unfolding stage to a greater extent. Therefore, the observed fact that both mutations influence both stages of GFP-c3 unfolding was quite surprising and clearly deserved some explanations.

How Can Destabilization of Different States of Protein Affect Its Unfolding Rate? To understand the obtained experimental data, let us consider Figure 5, which schematically shows how mutations in different GFP-c3 parts can influence the stability of its different states and also how such changes in the stability should affect the protein unfolding rates. Since the experimental data were obtained for an irreversible process, the schemes analyzed below also describe the irreversible GFP-c3 unfolding.

Suppose the protein can be conventionally divided into two parts that unfold sequentially and their unfolding rates are k_1

dx.doi.org/10.1021/bi2006674 | Biochemistry 2011, 50, 7735–7744

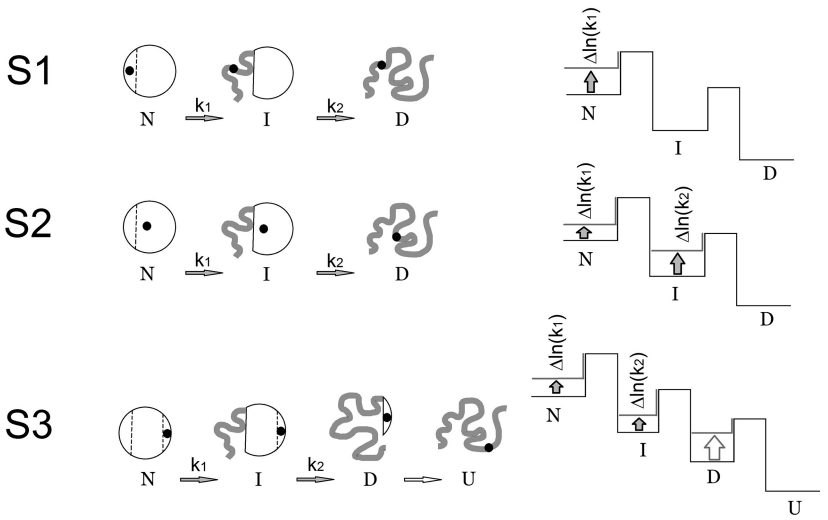


Figure 5. Schematic representation of the sequence of states at unfolding of GFP (on the left) and relative arrangement of energetic levels (on the right) of this protein. N, I, D, and U are native, intermediate, denatured, and unfolded states of the protein, respectively. Substituted amino acid residues (mutations) are shown with black circles. **S1**, the substituted amino acid residue is contained only in the structure of native state. Therefore, destabilization of the native state should lead to acceleration of only the first stage of unfolding (k_1). **S2**, the substituted amino acid residue is contained in the structure of intermediate state. As a result, destabilization of both the intermediate and native states should take place, which in its turn would affect the rates of the two stages of unfolding (k_1 and k_2). **S3**, the substituted amino acid residue is included in the structure of the very last structured state of unfolding (D); hence, it should affect the stability of all states and also the rates of all unfolding stages. There is an intentional inaccuracy in the represented energetic schemes. In fact, if an intermediate state I (for example in **S2**) is destabilized by a mutation, the free energy of the transition state between N and I should change at least by the same value. In this figure, the energetic schemes are represented in assumption that every unfolding stage occurs independently of all the other processes. Since only the unfolding rates are analyzed here, such a simplification of the energetic schemes is quite reasonable and provides an easy way to compare these rates with each other and to relate theoretical considerations with the experimental data.

and k_2. What would happen if we would insert a mutation in that part of protein which is the first to unfold (see Figure 5, scheme S1)? Such a mutation should strongly destabilize the native state, which in its turn should influence the unfolding rate k_1. The intermediate state should not be affected by this mutation at all. It is clearly seen from scheme S1 in Figure 5 that if the amino acid residue is in the unfolded part of the protein after the first stage of unfolding, then its substitution should not affect the stability of the intermediate state I. Furthermore, S1 in Figure 5 (on the right) shows also that changes would occur in the energetic scheme of the protein. Because of the native state destabilization (indicated by an arrow), the height of the first energetic barrier would decrease by the value of $\Delta \ln(k_1)$.

The next scheme (S2 in Figure 5) describes the case when the amino acid residue that stabilized the structure of the intermediate state was substituted. Such a mutation should influence the stability of both the intermediate I and the native N states, since in this case the intermediate is a part of the native state. Furthermore, it is possible to conclude that the native state destabilization would be less pronounced than that of the intermediate, since the intramolecular contacts of the substituted amino acid are far more "important" for the smaller intermediate I than for the larger native state N. In the energetic scheme S2 (see Figure 5), the protein with this substitution would be characterized by the $\Delta \ln(k_1)$ value lower than the $\Delta \ln(k_2)$ value.

Finally, the third scheme (see S3 in Figure 5), in which the substituted amino acid residue would destabilize the denatured state D, was analyzed. This situation could be realized via two potential scenarios. First, the protein unfolding could occur in two stages but not be complete, resulting in the preservation of the structured region in the denatured state D. Second, the unfolding could be complete and occur in three stages, where the partially structured state D is first formed and then unfolds completely at the last unfolding stage (see S3 in Figure 5). In this case, the substitution of the amino acid residue stabilizing state D should affect the rates of all preceding stages (see S3 in Figure 5).

Thus, the analysis of various unfolding schemes shown in Figure 5 suggests that the mutation of the amino acid residue included only in the structure of the native state N would influence only the initial unfolding step. The mutation of the amino acid residue included in the structure of the state close to the unfolded one would influence all the unfolding stages, among which the first unfolding step would be the least affected stage, whereas the last unfolding step would be affected most. These simple considerations are applied below to interpret the experimental data.

Experimental Data Interpretation. Since for the mutation analysis similar amino acid residues (isoleucines) with the same number of contacts were chosen and since these residues were substituted for alanine, it was expected that these mutations would similarly affect the local protein structure. Therefore, the comparison of the unfolding data for these two mutants and the analysis of their differential effects would provide important information on the molecular mechanisms of the GFP-c3 unfolding and would uncover the qualitative and quantitative effects of these mutations on different unfolding stages.

A comparison of the data presented in Figure 4C with the schemes given in Figure 5 suggests that the effects of substitution I161A in GFP-c3 can be perfectly described by scheme S2, where the substitution influenced both unfolding

dx.doi.org/10.1021/bi2006674 | Biochemistry 2011, 50, 7735–7744

stages, with the first unfolding stage being affected less than the second stage (i.e., $\Delta \ln(k_1)$ is lower than $\Delta \ln(k_2)$). This means that the amino acid residue 161 is located within the part of the protein that remains structured in the first intermediate state I (see Figure 5).

However, the effects of the I14A substitution on the GFP-c3 structure and unfolding could not be explained using schemes S1 and S2. In fact, this substitution influenced the rate constants much less than the I161A substitution. Furthermore, the I14A substitution influenced equally both k_1 and k_2 (Figure 4C). These effects of mutation on the GFP-c3 unfolding behavior can only occur if the Ile14 containing part of the protein is melted much later than at the stage of the intermediate state I formation and remains structured later, at the denatured state D (see S3 in Figure 5). Therefore, the effects of the I14A substitution on the GFP-c3 unfolding can be explained in two ways:

(i) The thermal unfolding of GFP-c3 proceeds in two stages, with the state D remaining rather structured (not completely unfolded) and being destabilized by the I14A substitution.

(ii) The thermal unfolding of GFP-c3 proceeds in three stages (as shown by scheme S3 in Figure 5), where the third unfolding stage occurs with no enthalpy change since it is not detected by the microcalorimetry.

In our opinion, the latter scenario is more probable and can take place if state D (see S3 in Figure 5) has the properties of the molten globule state.[79] In fact, the analysis of the melting behavior of many proteins revealed that the unfolding of their early molten-globule-like intermediate states can proceed without enthalpy and therefore cannot be detected by the calorimetry method. Such studies were performed, e.g., in the groups of Prof. Freire[80] and Prof. Privalov.[81]

Obviously, on the basis of analysis of only two mutant forms, it is difficult to make final conclusions on the precise molecular mechanisms of the GFP-c3 thermal unfolding and on the precise number and sequence of stages in this process. However, on the basis of analysis of the GFP-3c 3D structure, energetic schemes, and the obtained experimental data, we can conclude the thermal unfolding of this protein is a complex and multistage process. Although the calorimetric analysis of the GFP-c3 melting detected only two unfolding stages, the studies of two mutant forms of this protein showed that the GFP-c3 unfolding most likely proceeded in three stages (see scheme S3 in Figure 5). The first and second stages involve the unfolding of the protein part around the Ile161 region, i.e., the melting of the small hydrophobic cluster (see Figure 2). At the same time, the region containing the large hydrophobic cluster (Figure 1) remained structured even after the completion of the intermediate state unfolding and was melted only at the very last stage of GFP-c3 denaturation (see scheme S3 in Figure 5). It is likely that both higher packing density and increased intrinsic propensity for being flexible contributed to the higher stability of the large hydrophobic cluster in comparison with the mall hydrophobic cluster. In fact, a larger number of per residue contacts detected in the large cluster would increase the enthalpic component of its Gibbs free energy of unfolding, whereas the corresponding entropic component would be increased due to the higher predicted intrinsic mobility. Furthermore, the higher intrinsic mobility of the regions involved in the large hydrophobic cluster might also determine tighter packing of their residues via an improved spatial adjustment.

AUTHOR INFORMATION

Corresponding Author

*Ph 813-974-5816, Fax 813-974-7357, e-mail vuversky@health.usf.edu (V.N.U.); Ph +(7-495)- 514-02-18, Fax +(7-499)-135-99-84, e-mail bmelnik@phys.protres.ru (B.S.M.).

Funding

This work was supported in part by the MCB program of the Russian Academy of Sciences, a "Scientific Schools" grant, and the Special Federal Program "Research and Scientific-Pedagogical Specialists of Innovative Russia" (P304).

ACKNOWLEDGMENTS

The authors are grateful to Dr. Sergei A. Potekhin for reading the manuscript and fruitful discussion. We are also extremely grateful to Alexey V. Uversky for careful reading and editing this manuscript.

ABBREVIATIONS

GFP, green fluorescent protein; GFP-c3, GFP-cycle 3 mutant; GuHCl, guanidine hydrochloride; IDP, intrinsically disordered protein; IDR, intrinsically disordered region; FP, fluorescent protein.

REFERENCES

(1) Crameri, A., Whitehorn, E. A., Tate, E., and Stemmer, W. P. (1996) Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nature Biotechnol. 14*, 315−319.

(2) Shaner, N. C., Steinbach, P. A., and Tsien, R. Y. (2005) A guide to choosing fluorescent proteins. *Nature Methods 2*, 905−909.

(3) Tsien, R. Y. (1998) The green fluorescent protein. *Annu. Rev. Biochem. 67*, 509−544.

(4) Battistutta, R., Negro, A., and Zanotti, G. (2000) Crystal structure and refolding properties of the mutant F99S/M153T/V163A of the green fluorescent protein. *Proteins 41*, 429−437.

(5) Fukuda, H., Arai, M., and Kuwajima, K. (2000) Folding of green fluorescent protein and the cycle3 mutant. *Biochemistry 39*, 12025−12032.

(6) Wachter, R.M., Elsliger, M. A., Kallio, K., Hanson, G. T., and Remington, S. J. (1998) Structural basis of spectral shifts in the yellow-emission variants of green fluorescent protein. *Structure 6*, 1267−1277.

(7) Yang, F., Moss, L. G., and Phillips, G. N. Jr. (1996) The molecular structure of green fluorescent protein. *Nature Biotechnol. 14*, 1246−1251.

(8) Ormo, M., Cubitt, A. B., Kallio, K., Gross, L. A., Tsien, R. Y., and Remington, S. J. (1996) Crystal structure of the Aequorea victoria green fluorescent protein. *Science 273*, 1392−1395.

(9) Zimmer, M. (2002) Green fluorescent protein (GFP): applications, structure, and related photophysical behavior. *Chem. Rev. 102*, 759−781.

(10) Bokman, S. H., and Ward, W. W. (1981) Renaturation of Aequorea gree-fluorescent protein. *Biochem. Biophys. Res. Commun. 101*, 1372−1380.

(11) Ward, W. W., and Bokman, S. H. (1982) Reversible denaturation of Aequorea green-fluorescent protein: physical separation and characterization of the renatured protein. *Biochemistry 21*, 4535−4540.

(12) Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W., and Prasher, D. C. (1994) Green fluorescent protein as a marker for gene expression. *Science 263*, 802−805.

(13) Verkhusha, V. V., Kuznetsova, I. M., Stepanenko, O. V., Zaraisky, A. G., Shavlovsky, M. M., Turoverov, K. K., and Uversky, V. N. (2003) High stability of Discosoma DsRed as compared to Aequorea EGFP. *Biochemistry 42*, 7879−7884.

(14) Ward, W. W., Prentice, H. J., Roth, A. F., Cody, C. W., and Reeves, S. C. (1982) Spectral perturbations of the Aequorea green-fluorescent protein. *Photochem. Photobiol. 35*, 803−808.

(15) Huang, J. R., Craggs, T. D., Christodoulou, J., and Jackson, S. E. (2007) Stable intermediate states and high energy barriers in the unfolding of GFP. *J. Mol. Biol. 370*, 356−371.

(16) Enoki, S., Saeki, K., Maki, K., and Kuwajima, K. (2004) Acid denaturation and refolding of green fluorescent protein. *Biochemistry 43*, 14238−14248.

(17) Stepanenko, O. V., Verkhusha, V. V., Kazakov, V. I., Shavlovsky, M. M., Kuznetsova, I. M., Uversky, V. N., and Turoverov, K. K. (2004) Comparative studies on the structure and stability of fluorescent proteins EGFP, zFP506, mRFP1, "dimer2", and DsRed1. *Biochemistry 43*, 14913−14923.

(18) Andrews, B. T., Schoenfish, A. R., Roy, M., Waldo, G., and Jennings, P. A. (2007) The rough energy landscape of superfolder GFP is linked to the chromophore. *J. Mol. Biol. 373*, 476−490.

(19) Georgescu, J., Rehm, T., Wiehler, J., Steipe, B., and Holak, T. A. (2003) Backbone H(N), N, C(alpha) and C(beta) assignment of the GFPuv mutant. *J. Biomol. NMR 25*, 161−162.

(20) Khan, F., Stott, K., and Jackson, S. (2003) 1H, 15N and 13C backbone assignment of the green fluorescent protein (GFP). *J. Biomol. NMR 26*, 281−282.

(21) Melnik, T., Povarnitsyna, T., Solonenko, H., and Melnik, B. (2011) Studies of irreversible heat denaturation of green fluorescent protein by differential scanning microcalorimetry. *Thermochim. Acta 512*, 71−75.

(22) Dunker, A. K., Garner, E., Guilliot, S., Romero, P., Albrecht, K., Hart, J., Obradovic, Z., Kissinger, C., and Villafranca, J. E. (1998) Protein disorder and the evolution of molecular recognition: theory, predictions and observations. *Pac. Symp. Biocomput.*, 473−484.

(23) Romero, P., Obradovic, Z., Kissinger, C. R., Villafranca, J. E., Garner, E., Guilliot, S., and Dunker, A. K. (1998) Thousands of proteins likely to have long disordered regions. *Pac. Symp. Biocomput.*, 437−448.

(24) Wright, P. E., and Dyson, H. J. (1999) Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol. 293*, 321−331.

(25) Uversky, V.N., Gillespie, J. R., and Fink, A. L. (2000) Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins 41*, 415−427.

(26) Dunker, A.K., Lawson, J. D., Brown, C. J., Williams, R. M., Romero, P., Oh, J. S., Oldfield, C. J., Campen, A. M., Ratliff, C. M., Hipps, K. W., Ausio, J., Nissen, M. S., Reeves, R., Kang, C., Kissinger, C. R., Bailey, R. W., Griswold, M. D., Chiu, W., Garner, E. C., and Obradovic, Z. (2001) Intrinsically disordered protein. *J. Mol. Graphics Modell. 19*, 26−59.

(27) Dunker, A. K., and Obradovic, Z. (2001) The protein trinity--linking function and disorder. *Nature Biotechnol. 19*, 805−806.

(28) Tompa, P. (2002) Intrinsically unstructured proteins. *Trends Biochem. Sci. 27*, 527−533.

(29) Dunker, A. K., Brown, C. J., Lawson, J. D., Iakoucheva, L. M., and Obradovic, Z. (2002) Intrinsic disorder and protein function. *Biochemistry 41*, 6573−6582.

(30) Dunker, A. K., Brown, C. J., and Obradovic, Z. (2002) Identification and functions of usefully disordered proteins. *Adv. Protein Chem. 62*, 25−49.

(31) Iakoucheva, L. M., Brown, C. J., Lawson, J. D., Obradovic, Z., and Dunker, A. K. (2002) Intrinsic disorder in cell-signaling and cancer-associated proteins. *J. Mol. Biol. 323*, 573−584.

(32) Uversky, V. N. (2002) Natively unfolded proteins: a point where biology waits for physics. *Protein Sci. 11*, 739−756.

(33) Uversky, V. N. (2002) What does it mean to be natively unfolded? *Eur. J. Biochem. 269*, 2−12.

(34) Uversky, V. N. (2003) Protein folding revisited. A polypeptide chain at the folding-misfolding-nonfolding cross-roads: which way to go? *Cell. Mol. Life Sci. 60*, 1852−1871.

(35) Dyson, H. J., and Wright, P. E. (2002) Coupling of folding and binding for unstructured proteins. *Curr. Opin. Struct. Biol. 12*, 54−60.

(36) Fink, A. L. (2005) Natively unfolded proteins. *Curr. Opin. Struct. Biol. 15*, 35−41.

(37) Dyson, H. J., and Wright, P. E. (2005) Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol. 6*, 197−208.

(38) Tompa, P. (2005) The interplay between structure and function in intrinsically unstructured proteins. *FEBS Lett. 579*, 3346−3354.

(39) Dunker, A.K., Cortese, M. S., Romero, P., Iakoucheva, L. M., and Uversky, V. N. (2005) Flexible nets. The roles of intrinsic disorder in protein interaction networks. *FEBS J. 272*, 5129−5148.

(40) Uversky, V. N., Oldfield, C. J., and Dunker, A. K. (2005) Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. *J. Mol. Recognit. 18*, 343−384.

(41) Radivojac, P., Iakoucheva, L. M., Oldfield, C. J., Obradovic, Z., Uversky, V. N., and Dunker, A. K. (2007) Intrinsic disorder and functional proteomics. *Biophys. J. 92*, 1439−1456.

(42) Sickmeier, M., Hamilton, J. A., LeGall, T., Vacic, V., Cortese, M. S., Tantos, A., Szabo, B., Tompa, P., Chen, J., Uversky, V. N., Obradovic, Z., and Dunker, A. K. (2007) DisProt: the Database of Disordered Proteins. *Nucleic Acids Res. 35*, D786−793.

(43) Le Gall, T., Romero, P. R., Cortese, M. S., Uversky, V. N., and Dunker, A. K. (2007) Intrinsic disorder in the Protein Data Bank. *J. Biomol. Struct. Dyn. 24*, 325−342.

(44) Vucetic, S., Xie, H., Iakoucheva, L. M., Oldfield, C. J., Dunker, A. K., Obradovic, Z., and Uversky, V. N. (2007) Functional anthology of intrinsic disorder. 2. Cellular components, domains, technical terms, developmental processes, and coding sequence diversities correlated with long disordered regions. *J. Proteome Res. 6*, 1899−1916.

(45) Xie, H., Vucetic, S., Iakoucheva, L. M., Oldfield, C. J., Dunker, A. K., Obradovic, Z., and Uversky, V. N. (2007) Functional anthology of intrinsic disorder. 3. Ligands, post-translational modifications, and diseases associated with intrinsically disordered proteins. *J. Proteome Res. 6*, 1917−1932.

(46) Xie, H., Vucetic, S., Iakoucheva, L. M., Oldfield, C. J., Dunker, A. K., Uversky, V. N., and Obradovic, Z. (2007) Functional anthology of intrinsic disorder. 1. Biological processes and functions of proteins with long disordered regions. *J. Proteome Res. 6*, 1882−1898.

(47) Cortese, M.S., Uversky, V. N., and Keith Dunker, A. (2008) Intrinsic disorder in scaffold proteins: getting more from less. *Prog. Biophys. Mol. Biol. 98*, 85−106.

(48) Dunker, A. K., Oldfield, C. J., Meng, J., Romero, P., Yang, J. Y., Chen, J. W., Vacic, V., Obradovic, Z., and Uversky, V. N. (2008) The unfoldomics decade: an update on intrinsically disordered proteins. *BMC Genomics 9* (Suppl 2), S1.

(49) Dunker, A. K., Silman, I., Uversky, V. N., and Sussman, J. L. (2008) Function and structure of inherently disordered proteins. *Curr. Opin. Struct. Biol.*

(50) Dunker, A. K., and Uversky, V. N. (2008) Signal transduction via unstructured protein conduits. *Nat. Chem. Biol. 4*, 229−230.

(51) Fuxreiter, M., Tompa, P., Simon, I., Uversky, V. N., Hansen, J. C., and Asturias, F. J. (2008) Malleable machines take shape in eukaryotic transcriptional regulation. *Nat. Chem. Biol. 4*, 728−737.

(52) Uversky, V. N., Oldfield, C. J., and Dunker, A. K. (2008) Intrinsically disordered proteins in human diseases: introducing the D2 concept. *Annu. Rev. Biophys. 37*, 215−246.

(53) Ferron, F., Longhi, S., Canard, B., and Karlin, D. (2006) A practical overview of protein disorder prediction methods. *Proteins 65*, 1−14.

(54) Dosztanyi, Z., Sandor, M., Tompa, P., and Simon, I. (2007) Prediction of protein disorder at the domain level. *Curr. Protein Pept. Sci. 8*, 161−171.

(55) Dosztanyi, Z., and Tompa, P. (2008) Prediction of protein disorder. *Methods Mol. Biol. 426*, 103−115.

(56) Li, L., Uversky, V. N., Dunker, A. K., and Meroueh, S. O. (2007) A computational investigation of allostery in the catabolite activator protein. *J. Am. Chem. Soc. 129*, 15668−15676.

(57) Liang, S., Li, L., Hsu, W.-L., Zhou, Y., Dunker, A. K., Uversky, V. N., and Meroueh, S. O. (2008) Computational grafting of protein interaction sites with a new algorithm that incorporates receptor flexibility. *Biochemistry*.

(58) Kutyshenko, V. P., Prokhorov, D. A., Timchenko, M. A., Kudrevatykh, Y. A., Gushchina, L. V., Khristoforov, V. S., Filimonov, V. V., and Uversky, V. N. (2009) Solution structure and dynamics of the chimeric SH3 domains, SHH- and SHA-"Bergeracs. *Biochim. Biophys. Acta 1794*, 1813−1822.

(59) Senin, A. A., Potekhin, S. A., Tiktopulo, E. I., and Filomonov, V. V. (2000) Differential scanning microcalorimetr SCAL-1. *J. Therm. Anal. Calorim. 62*, 153−160.

(60) Privalov, P.L., and Potekhin, S. A. (1986) Scanning microcalorimetry in studying temperature-induced changes in proteins. *Methods Enzymol. 131*, 4−51.

(61) Kurganov, B. I., Lyubarev, A. E., Sanchez-Ruiz, J. M., and Shnyrov, V. L. (1997) Analysis of differential scanning calorimetry data for proteins. Criteria of validity of one-step mechanism of irreversible protein denaturation. *Biophys. Chem. 69*, 125−135.

(62) Potekhin, S. A., and Kovrigin, E. L. (1998) Folding under inequilibrium conditions as a possible reason for partial irreversibility of heat-denatured proteins: computer simulation study. *Biophys. Chem. 73*, 241−248.

(63) Potekhin, S. A., and Kovrigin, E. L. (1998) [Influence of kinetic factors on heat denaturation and renaturation of biopolymers]. *Biofizika 43*, 223−232.

(64) Sanchez-Ruiz, J. M., Lopez-Lacomba, J. L., Cortijo, M., and Mateo, P. L. (1988) Differential scanning calorimetry of the irreversible thermal denaturation of thermolysin. *Biochemistry 27*, 1648−1652.

(65) Lyubarev, A. E., Kurganov, B. I., Burlakova, A. A., and Orlov, V. N. (1998) Irreversible thermal denaturation of uridine phosphorylase from Escherichia coli K-12. *Biophys. Chem. 70*, 247−257.

(66) Lyubarev, A. E., and Kurganov, B. I. (1998) Modeling of irreversible thermal protein denaturation at varying temperature. I. The model involving two consecutive irreversible steps. *Biochemistry (Moscow) 63*, 434−440.

(67) Kabsch, W., and Sander, C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers 22*, 2577−2637.

(68) Xue, B., Dunbrack, R. L., Williams, R. W., Dunker, A. K., and Uversky, V. N. (2010) PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. *Biochim. Biophys. Acta 1804*, 996−1010.

(69) Romero, P., Obradovic, Z., Li, X., Garner, E. C., Brown, C. J., and Dunker, A. K. (2001) Sequence complexity of disordered protein. *Proteins 42*, 38−48.

(70) Peng, K., Vucetic, S., Radivojac, P., Brown, C. J., Dunker, A. K., and Obradovic, Z. (2005) Optimizing long intrinsic disorder predictors with protein evolutionary information. *J. Bioinform. Comput. Biol. 3*, 35−60.

(71) Peng, K., Radivojac, P., Vucetic, S., Dunker, A. K., and Obradovic, Z. (2006) Length-dependent prediction of protein intrinsic disorder. *BMC Bioinformatics 7*, 208.

(72) Prilusky, J., Felder, C. E., Zeev-Ben-Mordehai, T., Rydberg, E. H., Man, O., Beckmann, J. S., Silman, I., and Sussman, J. L. (2005) FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics 21*, 3435−3438.

(73) Dosztanyi, Z., Csizmok, V., Tompa, P., and Simon, I. (2005) The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. *J. Mol. Biol. 347*, 827−839.

(74) Campen, A., Williams, R. M., Brown, C. J., Meng, J., Uversky, V. N., and Dunker, A. K. (2008) TOP-IDP-scale: a new amino acid scale measuring propensity for intrinsic disorder. *Protein Pept. Lett. 15*, 956−963.

(75) Meszaros, B., Simon, I., and Dosztanyi, Z. (2009) Prediction of protein binding regions in disordered proteins. *PLoS Comput. Biol. 5*, e1000376.

(76) Dosztanyi, Z., Meszaros, B., and Simon, I. (2009) ANCHOR: web server for predicting protein binding regions in disordered proteins. *Bioinformatics 25*, 2745−2746.

(77) Dosztanyi, Z., Csizmok, V., Tompa, P., and Simon, I. (2005) IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics 21*, 3433−3434.

(78) Privalov, P. L. (1979) Stability of proteins: small globular proteins. *Adv. Protein Chem. 33*, 167−241.

(79) Ptitsyn, O.B., Pain, R. H., Semisotnov, G. V., Zerovnik, E., and Razgulyaev, O. I. (1990) Evidence for a molten globule state as a general intermediate in protein folding. *FEBS Lett. 262*, 20−24.

(80) Haynie, D. T., and Freire, E. (1993) Structural energetics of the molten globule state. *Proteins 16*, 115−140.

(81) Griko, Y. V., and Privalov, P. L. (1994) Thermodynamic puzzle of apomyoglobin unfolding. *J. Mol. Biol. 235*, 1318−1325.