THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

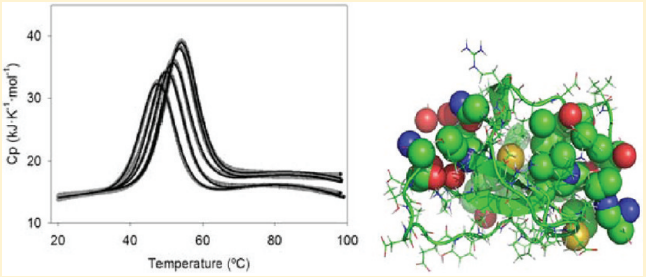
Thermodynamic Impact of Embedded Water Molecules in the Unfolding of Human CD2BP2-GYF Domain

Montserrat Andujar-Sanchez,[†] Eva S. Cobos,[‡] Irene Luque,[‡] and Jose C. Martinez[‡,*]

[†]Department of Physical Chemistry, Biochemistry and Inorganic Chemistry, University of Almeria, Agrifood Campus of International Excellence (ceiA3), Carretera de Sacramento, 04120 Almeria, Spain

[‡]Department of Physical Chemistry and Institute of Biotechnology, Faculty of Sciences, University of Granada, 18071 Granada, Spain

ABSTRACT: GYF domains are small polyproline-recognition modules adopting a structural arrangement consisting of a single α-helix packed against a small β-sheet. Although most families of proline-rich recognition modules have been extensively characterized in terms of function, structure, or conformational flexibility, little is known about GYF domain functionality and folding. We have undertaken the thermodynamic characterization of the unfolding of CD2BP2-GYF domain by combining differential scanning calorimetry and circular dichroism under different pH conditions. The experimental data can be well-described in terms of a two-state equilibrium, although an unusually high heat capacity of the native state reflects a considerable conformational flexibility and dynamics of CD2BP2-GYF domain. In addition, the normalized thermodynamic parameters of unfolding (enthalpy, entropy and heat capacity) are roughly a factor of two greater than expected. In contrast, stability curves reveal an ordinary unfolding behavior of CD2BP2-GYF domain in terms of Gibbs energies, incurring thus unusually strong enthalpy−entropy compensation. This phenomenon, previously described as "thermodynamic homeostasis", has been associated in different examples to the contribution of occluded water (solvent) molecules into the protein structure. By means of CASTp server, we have found seven cavities/pockets scattered throughout of the CD2BP2-GYF structure, each able to harbor at least one water molecule. This structural feature provides rationalization for the atypical enthalpy values observed for CD2BP2-GYF because each water molecule is able to organize an extra amount of hydrogen bonds in the native state. In addition, these bound waters increase the vibrational entropy of the protein, which could also be responsible for an increase in protein flexibility and may thus fully explain the homeostatic behavior experimentally observed.



INTRODUCTION

GYF domains are small (~60 residues) modular domains that belong to the functional class of proline-rich recognition modules.[1] They are present in most eukaryotic species sequenced so far but in contrast with the other types of proline-recognition domains such as SH3 or WW domains, have not experienced the same amplification in metazoan during evolution.[2] They adopt a typical structural arrangement consisting of a single α-helix packed against a small β-sheet. The helix is inclined away from the sheet, allocating some conserved hydrophobic residues in between which form a hydrophobic pocket that accommodates the proline-rich binding motif. As is the case for other domain families, it is expected that all GYF domains would share the same fold, even though structural information for the GYF family of domains is still scarce, being reduced to the solution structures of CD2BP2-GYF (PDB code: 1GYF)[3] and SMY2-GYF (PDB code: 3K3V) domains.

Most families of proline-rich recognition modules have been extensively characterized in terms of function, structure, or conformational flexibility. Nonetheless, little is known about GYF domain functionality and folding. Given the importance of proline-rich sequence recognition in the assembly of intracellular protein−protein interactions it is desirable to create a complete database of structural, conformational, and functional information regarding these domains.

Because the folding properties of GYF domains are fully unknown at present, we have undertaken the thermodynamic characterization of CD2BP2-GYF domain by combining differential scanning calorimetry (DSC) and circular dichroism (CD) under different pH conditions. A multidimensional analysis of the unfolding equilibrium has allowed us to determine reliably the thermodynamic magnitudes. Although the experimental data can be well-described in terms of two-state equilibrium, some anomalous and interesting features related with the unfolding enthalpy, entropy, and heat capacity changes have been found, revealing a highly flexible and dynamic domain, where the extra number of hydrogen bonds formed by some water molecules embedded into the folded

Received: April 11, 2012
Revised: May 18, 2012
Published: May 24, 2012

state will restore the polar versus nonpolar energetic balance of the well-arranged native state of CD2BP2-GYF domain.

EXPERIMENTAL METHODS

Protein Samples. The pTFT74 plasmid encoding an untagged CD2BP2-GYF sequence was a generous gift of Dr. Christian Freund (Protein Engineering Group, Free University and FMP Berlin, Germany). The domain was overexpressed in *Escherichia coli* BL21/DE3 cells and purified as described elsewhere[3] by means of exclusion chromatography using a Superdex-75 resin (GE Healthcare). Experimental samples were prepared by extensive dialysis against a large volume (1 to 2 L) of the appropriate buffer, 50 mM MES (pH 6), 50 mM phosphate (pH 7), 50 mM HEPES (pH 8), 50 mM Bicine (pH 9), and 50 mM CAPS (pH 10). Protein concentrations were measured by absorption at 280 nm using an extinction coefficient of 24 700 $cm^{-1}\cdot M^{-1}$, determined as described by Gill and von Hippel,[4] and a molecular mass of 7450 Da, as determined by mass spectrometry (Mass Spectrometry Service of the CIC, University of Granada).

Differential Scanning Calorimetry. DSC experiments were conducted in a DASM-4 microcalorimeter (Biopribor, Russia) at a scan rate of 2 $K\cdot min^{-1}$ using protein concentrations between 0.5 and 3.2 $mg\cdot mL^{-1}$. The samples were routinely heated twice from 5 to 80−100 °C to check the reversibility of the unfolding process. Each experimental DSC thermogram was corrected for the time response of the calorimeter and from the instrumental baseline obtained with both calorimeter cells filled with the corresponding dialysis buffer. After normalization by protein concentration, the heat capacity curves, $C_p(T)$, were calculated from the resulting thermograms assuming 0.73 $mL\cdot g^{-1}$ for the partial specific volume.[5] These calculations were made using ORIGIN (OriginLab).

According to the two-state model, the DSC traces were fitted using the equations previously described by us.[6] In all strategies described in this work, we have used a linear function to describe the heat capacity of the N state and a quadratic polynomial one to describe that of the U state.[6−8] Such heat capacity functions, together with the unfolding temperature (T_m) and enthalpy (ΔH_m), were the fitting parameters to be obtained from each DSC thermogram. The values found for such parameters were used as initial estimations in our global analysis, carried out as described in the Results and Discussion.

To calculate the van't Hoff enthalpy values (ΔH_{vH}), we have used the following equation

$$\Delta H_{vH} = 4\cdot R\cdot T_m^2 \frac{C_{p,exc,m}}{\Delta H_m} \quad (1)$$

where R is the gas constant, T_m is the temperature corresponding to the maximum heat capacity value in the DSC curve, and $C_{p,exc,m}$ is the value of the excess heat capacity at the T_m. This value can be determined from each experiment as the maximum height of the $C_p(T)$ curve in $kJ\cdot K^{-1}\cdot mol^{-1}$.

Circular Dichroism. CD measurements were taken in a Jasco J-715 spectropolarimeter equipped with a temperature-controlled cell holder. Far-UV CD spectra were recorded from 260 to 200 nm in a 2 mm path length cuvette using sample concentrations of 0.3 $mg\cdot mL^{-1}$. Near-UV CD spectra were measured from 350 to 250 nm in a 2 mm path length cuvette with sample concentrations of 1 $mg\cdot mL^{-1}$. Each CD spectrum was obtained by averaging five accumulations collected at a scan rate of 50 $nm\cdot min^{-1}$. All spectra were corrected for buffer contributions and then normalized as mean-residue molar ellipticity. Temperature scans were conducted at a heating rate of 1 $deg\cdot min^{-1}$ after equilibration of the measuring cell within the cell holder at 2 °C for 10 min. Unfolding reversibility was checked by comparing the shapes and intensities of spectra at 2 °C obtained before and after heating of protein samples inside the instrument.

Thermal unfolding experiments followed by CD obtained at three different wavelengths (220, 230, and 280 nm) and at pH 7 in 50 mM phosphate were fitted individually to the equations previously described by us.[6] Individual fittings of the experimental traces to the two-state model were performed taking the unfolding temperature (T_m), the unfolding enthalpy (ΔH_m), and the pre- and post-transition baselines to be floating parameters. The values obtained for such parameters were used as initial estimations in our global analysis, as it is explained in the Results and Discussion. All fittings and calculations done in this work were performed with SigmaPlot 2000 (Systat Software).

RESULTS AND DISCUSSION

CD2BP2-GYF Domain Behaves As a Two-State Folding Protein. We studied the thermal unfolding of the CD2BP2-GYF domain by DSC over the pH range from 6 to 10 (Figure 1). At moderate ionic strengths, in neutral and basic solutions, the thermal unfolding of this domain was highly reversible (more than 60%) and no appreciable scan rate or protein concentration effects could be detected for the DSC traces (data not shown). Thus, the domain may be considered as being a monomer (in agreement with previous results[5]) that unfolds under equilibrium in these experimental conditions. Dynamic light scattering experiments performed as previously described[9] further support these observations (data not shown). Under acidic pH conditions, the domain is poorly soluble and unfolds in an irreversible way.

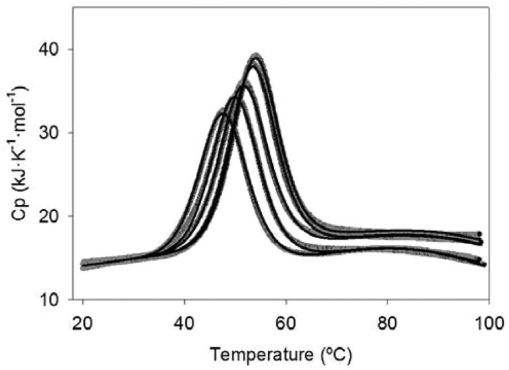


Figure 1. Temperature dependencies of the partial molar heat capacity, C_p, of the CD2BP2-GYF domain. Gray symbols correspond to experimental data, from right to left, to 50 mM MES (pH 6), 50 mM phosphate (pH 7), 50 mM HEPES (pH 8), 50 mM Bicine (pH 9), and 50 mM CAPS (pH 10). Solid black lines are the result of global fitting of the whole set of curves to the two-state model, accepting common $C_{pN}(T)$ and $C_{pU}(T)$ functions, as described in the text.

A global fitting of the DSC thermal unfolding curves was conducted, assuming that DSC traces obtained at the various pH values shared common heat capacity functions for the native, $C_{pN}(T)$, and unfolded, $C_{pU}(T)$, states, which is usual for a two-state analysis,[6−8] whereas the T_m and ΔH_m values were

dx.doi.org/10.1021/jp303495b | J. Phys. Chem. B 2012, 116, 7168−7175

Table 1. Thermodynamic Parameters of CD2BP2-GYF Domain Folding Obtained from the Global Fitting of DSC and CD Thermal Unfolding Experiments under Different pH Conditions[a]

method	pH	T_m (°C)	ΔH_m (kJ·mol^{-1})	ΔH_{vH} (kJ·mol^{-1})	ΔC_{pN-U} (kJ·K^{-1}·mol^{-1})	$\Delta G_{N-U}(298)$ (kJ·mol^{-1})	T_S (°C)	$\Delta G_{max}(T_S)$ (kJ·mol^{-1})
DSC global fitting[b]	6	55.0	285	298	8.36	14.3	23.0	14.4
	7	54.5	280	292		13.8	23.0	13.9
	8	53.0	265	278		12.5	23.0	12.5
	9	51.8	257	262		11.7	23.0	11.8
	10	49.8	242	250		10.4	22.0	10.5
multiple wavelength CD and DSC global fitting[c]	7	53.8	285			14.2		

[a]Errors have been estimated as 5% for T_m, 5% for ΔH_m, and 10% for the rest of parameters represented in the Table. [b]According to the fitting procedure, together with the values of T_m and ΔH_m given, we obtained the numerical values of DSC baselines, which were $C_{pN} = 14.05 + 0.0838\cdot(T - 293)$; $C_{pU} = -5.28 + 0.64\cdot(T - 293) - 5.0 \times 10^{-3}\cdot(T^2 - 293^2)$. [c]According to the fitting procedure, together with the values of T_m and ΔH_m given, we obtained the numerical values of CD and DSC baselines, which were $\Theta_{220,N} = -15.64 + 0.031\cdot(T - 293)$; $\Theta_{220,U} = 3.14 - 0.0186\cdot(T - 293)$; $\Theta_{230,N} = -3.90 + 0.01\cdot(T - 293)$; $\Theta_{230,U} = -2.0 + 0.0016\cdot(T - 293)$; $\Theta_{282,N} = -2.23 + 0.0056\cdot(T - 293)$; $\Theta_{282,U} = 0.69 - 0.0019\cdot(T - 293)$; $C_{pN} = 14.38 + 0.031\cdot(T - 293)$; $C_{pU} = -2.04 + 0.50\cdot(T - 293) - 4.0 \times 10^{-3}\cdot(T^2 - 293^2)$.

considered individually for each pH value. We observed a progressive decrease in the post-transition heat capacity values in parallel with the increase in pH (Figure 1). This phenomenon has been previously described[10] being attributed to the differential contribution of the distinct protonation states of the charged amino acid side chains. Within the pH interval of DSC experiments (pH 6–10), we have five Lys and Arg residues as the most probable candidates. Although the contribution of these protonation effects can be roughly estimated,[10] they have been obtained more precisely as fitting parameters from the two-state approach. As illustrated in Figure 1, this analysis provides a good description of the experimental data, with R and R_{sqr} values of 0.99.

The robustness of the global analysis is stressed by the fact that the values of van't Hoff enthalpies estimated from the shape of the DSC traces (eq 1) are comparable to those derived from the two-state analysis (Table 1) so that the van't Hoff to calorimetric ratio is ~1.05.

To confirm further the two-state behavior, we have studied the thermal unfolding of the CD2BP2-GYF domain by CD at pH 7 (Figure 2). The far-UV CD spectrum of native CD2BP2-GYF is characterized by a small positive peak centered at 230 nm, also observed in SH3 and WW domains, which has been associated with local interactions involving aromatic side chains, mainly those of Trp residues in a hydrophobic microenvironment.[6,7] We have three Trp residues partially exposed to solvent in CD2BP2-GYF. There is also an intensely negative peak at 220 nm, which accounts for the different contributions of secondary structure elements.[11] The near-UV CD spectrum displays an acute minimum at 282 nm also reporting for the tertiary arrangement of aromatic residues, quite abundant in this domain (see the next section). A two-state equilibrium is generally associated with the existence of an isosbestic point. In this respect, interestingly, the far-UV CD spectra obtained for CD2BP2-GYF at different temperatures show two well-defined isodichroic points at around 210 and 225 nm (Figure 2).

Figure 3 shows the thermal denaturation profiles followed by DSC and CD at three wavelengths (220, 230, and 282 nm) in 50 mM phosphate buffer (pH 7). A combined CD-DSC global fitting procedure was performed by combining the equations of both approaches, as we have previously described.[6] As a result, we obtained a single pair of T_m and ΔH_m values for the unique pH condition explored (pH 7), plus the parameters defining the N- and U-state heat capacity functions of the DSC experiment and, last, the respective ones for the N- and U-state

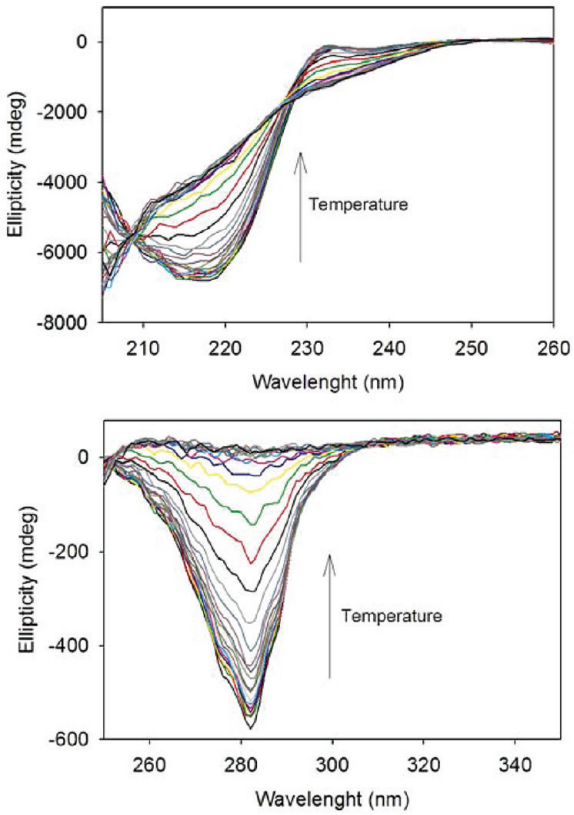


Figure 2. Temperature dependency of the far- (upper panel) and near-UV CD (lower panel) spectra of CD2BP2-GYF in 50 mM phosphate buffer at pH 7. The spectra were recorded at increasing intervals of 2 °C (from bottom to top in both cases). The protein concentration was 40.4 and 134.6 μM for far- and near-UV CD respectively.

baselines of CD measurements at each wavelength. It can be seen that the experimental data can be satisfactorily described by a two-state model, providing a common set of thermodynamic parameters for the four thermal unfolding profiles, also concordant, within errors, with the respective one obtained from the analysis of DSC experiments under different pH conditions (Table 1).

Uncommon Heat Capacities of N- and U-States Are Related to the Abnormal Amino Acids Content of CD2BP2-GYF Domain. The pre- and post-transition heat capacity values, corresponding to the N and U states, can be

dx.doi.org/10.1021/jp303495b | J. Phys. Chem. B 2012, 116, 7168–7175

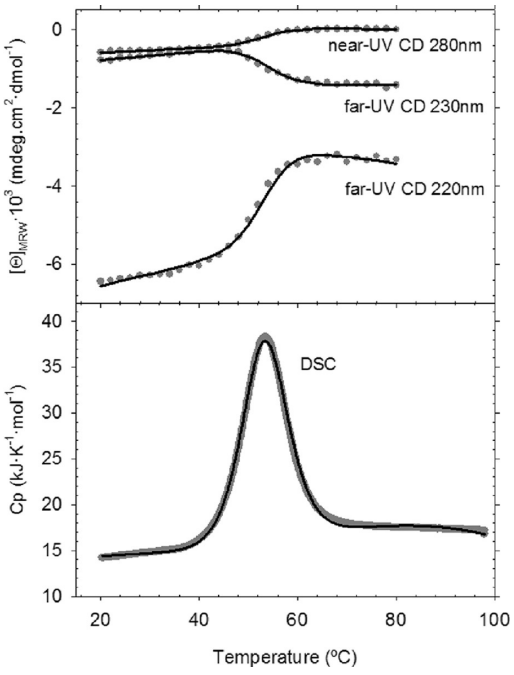


Figure 3. DSC and CD thermal unfolding curves of the CD2BP2-GYF domain followed at 220, 230, and 282 nm in 50 mM phosphate pH 7. Experimental data are represented as gray circles. The solid lines correspond to the simultaneous global fitting to the two-state model, accepting a common temperature function of the heat capacity change for both the DSC and CD data sets.

estimated from the molecular weight of the domain[12] and from tabulated contributions of individual amino acids,[13] respectively. As can be observed in Figure 4, a clear discrepancy

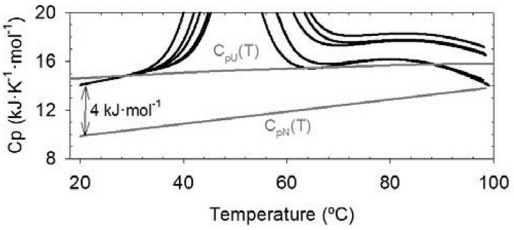


Figure 4. Comparison of the $C_{pN}(T)$ and $C_{pU}(T)$ functions obtained from global fitting of DSC traces (black lines) and from amino acid composition (gray lines).

between the theoretical predictions and the experimentally determined functions was found for the thermal unfolding of CD2BP2-GYF domain. These divergences would derive, in a first approximation, from an abnormal amino acid content of the protein.

We have compared the aromatic, nonpolar, and charged-residue content of CD2BP2-GYF domain with other proline-binding domains whose folding has been previously studied (Table 2). An extremely high content of aromatic residues is observed, as well as a lower than normal number of basic residues (only five Lys and Arg residues). The inspection of the NMR structure (PDB code: 1GYF) reveals that the packing of the single α-helix to the antiparallel β-sheet is achieved by residues G18-P19-F20 and G32-Y33-F34 that organize the binding groove of the domain. The residue W28 lines up along the α-helical axis, allowing a continuous surface patch of aromatic residues.[3] As a result, the surface of the protein in the neighborhood of the binding pocket is fully aromatic in nature (Figure 5). Residues Y6 and M25 appear as the main contributors to the hydrophobic core of the domain, being fully buried from solvent. It has been shown that Ala-scanning mutagenesis of aromatic residues has been proven to destroy the functional and the structural integrity of the domain,[3] highlighting the great importance of these residues for the formation of the protein core and the binding surface, despite the fact that they are partially solvent exposed.

The native state heat capacity function is shifted about 4 kJ·K^{-1}·mol^{-1} with respect to the theoretical values. To understand the dynamic consequences of such upshift, we should bear in mind that $C_{pN}(T)$ functions are closely correlated to enthalpy fluctuations in the macroscopic N state, noncooperative in nature, because they are established between isoenthalpic conformations (ref 6 and references therein). Therefore, the unusually high $C_{pN}(T)$ function of CD2BP2-GYF domain reflects a considerable conformational flexibility and dynamics[6,14] because the isoenthalpic thermal motions are higher than those of the majority of globular proteins studied to date that follow well the $C_{pN}(T)$ function calculated from M_w.[12]

The dispersion observed in the U-state heat capacity can be explained, a priori, by the differential contribution of the distinct protonation states of the charged Lys and Arg residues (not taken into account in the calculations), as explained above. Even so, the predicted shape for the U-state heat capacity is different from the experimental, the latter being a bit more curved (Figure 4). Although the experimental curvature was not predicted from sequence calculations, the different fitting strategies were (Figures 1 and 3). The calorimetric analysis of the temperature dependency of heat capacities of isolated amino acids has revealed that polar groups increase such dependencies with temperature, whereas the opposite is true for nonpolar and aromatic side chains.[15] Therefore, the resulting heat capacity function, $C_{pU}(T)$, calculated from the amino acid composition of any protein example would reflect the resultant of such compensating effects, which usually translates into roughly flat heat capacity functions in the U-state region that decrease with temperature, as shown in Figure 4

Table 2. Sequence Analysis of Some Polyproline Binding Domains

domain	% aromatic [W,Y,F]	% nonpolar [V,L,I,A.P.M][a]	% hydrophobic [nonpolar + aromatic]	% basic [R,K]	% acidic [D,E]	% charged [acidic + basic]
CD2-BP2-GYF (7.4 kDa)	**21.0**	24.2	45.2	**8.0**	16.1	24.1
α-spectrin SH3 (7.1 kDa)	9.7	32.2	41.9	16.1	17.7	33.9
Nedd4-WW4 (5.7 kDa)	8.3	27.1	35.4	10.4	16.7	27.1

[a]According to the hydrophobicity scale published by Mackhatadze and Privalov.[13,15]

dx.doi.org/10.1021/jp303495b | J. Phys. Chem. B 2012, 116, 7168−7175

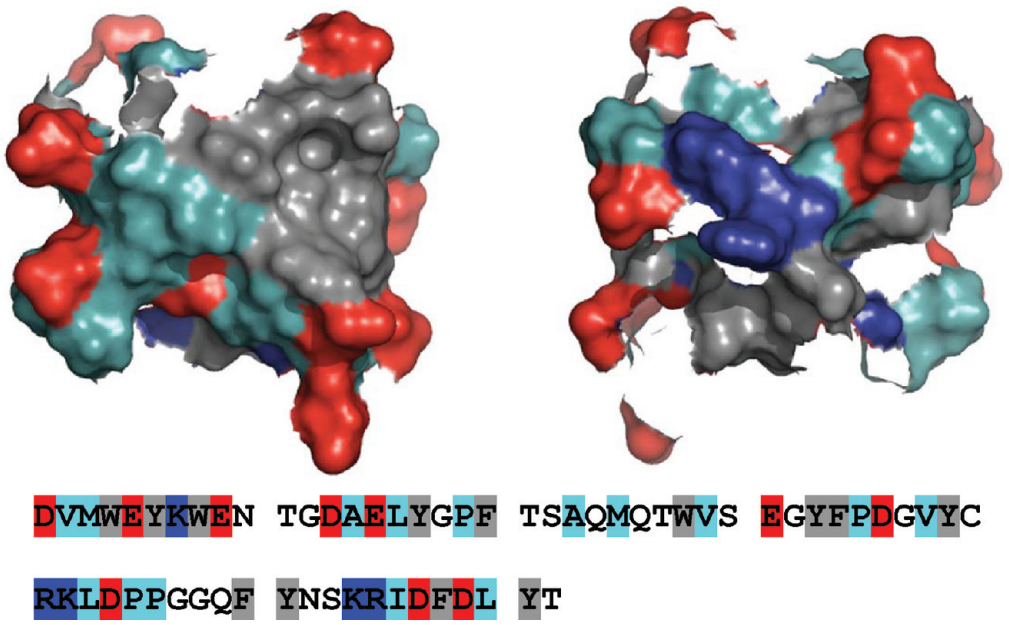


Figure 5. 180° view of the surfaces of CD2BP2-GYF domain (PDB entry 1GYF). Side chains of the aromatic residues are colored gray, nonpolar residues are colored cyan, acidic residues are colored red, and basic residues are colored blue.

and other examples.[15,13,6−8] Nevertheless, the unusual clustering of aromatic residues found in this domain may result in thermal effects not evaluated from the additive contributions of isolated amino acids, which might result in the concave temperature dependency of the U-state heat capacity function.

CD2BP2-GYF Domain Is Characterized by Higher than Average Normalized Thermodynamic Values but Is Not a Very Stable Protein. Using a database of thermodynamic parameters, Robertson and Murphy[16] have found a nice correlation between their magnitudes and the size of the protein, that is, the number of amino acid residues. From such analysis they have proposed that those globular proteins that unfold under equilibrium would have ΔC_{pN-U} = 58 ± 4 J·K^{-1}(mol res)$^{-1}$, ΔH_{N-U}(60 °C) = 2.92 ± 0.35 J·(mol res)$^{-1}$, and ΔS_{N-U}(60 °C) = 9.0 ± 1.4 J·K^{-1}(mol res)$^{-1}$. The reference temperature of 60 °C was chosen because is close to the mean and median T_m values of the data set, which minimizes errors derived from long extrapolations of the experimental values obtained from model analysis at T_m.[16] Because the T_m values of CD2BP2-GYF are relatively close to this reference value (Table 1), the unfolding enthalpy and entropy of CD2BP2-GYF can be easily calculated at 60 °C without requiring long extrapolations from Kirchoff's equations by using a constant ΔC_{pN-U} = 8.4 kJ·K^{-1}·mol^{-1} value obtained from the linear regression of ΔH_m values versus their corresponding T_m ones (Figure 6 and Table 1). The normalized thermodynamic parameters of the 62-residue CD2BP2-GYF are: ΔC_{pN-U} = 135 ± 10 J·K^{-1}(mol res)$^{-1}$, ΔH_{N-U}(60 °C) = 5.45 ± 0.65 J·(mol res)$^{-1}$, and ΔS_{N-U}(60 °C) = 16.5 ± 2.5 J·K^{-1}(mol res)$^{-1}$. These values are very different from the estimated ones by Robertson and Murphy, with discrepancies close to two-fold.

It has also been described that the extrapolated values of specific heats of unfolding [$\Delta h_m = \Delta H_m/M_w$] for globular proteins intersect close to 110 °C, reaching a common value of 54 J·g^{-1}.[17] The extrapolation to 110 °C of the regression analysis given in Figure 6 gives rise to a value of 100 J·g^{-1}, which appears to be the highest Δh_m(110 °C) of all globular proteins studied so far. Other exceptions are either very small

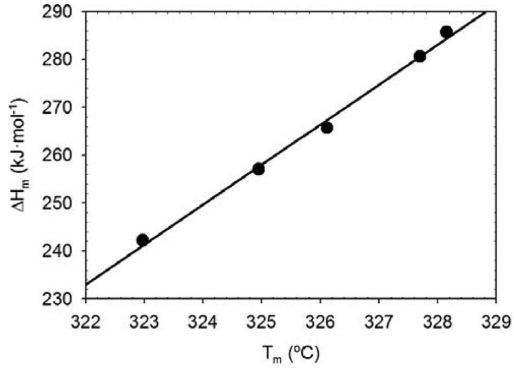


Figure 6. Unfolding enthalpy values plotted versus denaturation temperature obtained from DSC multidimensional analysis (Table 1). The solid line is the linear regression of data.

proteins (like Nedd4-WW4 domain, 45 J·g^{-16}), extensively cross-linked by S−S bridges, or those with an elongated rather than globular structure.[8] All of these cases appear to display lower than expected values.

As far as we know, only another protein, Barnase, a ribonuclease of 12.4 kDa that also follows the two-state regime, has been described to present a specific enthalpy of denaturation well above the enthalpic convergence, with a value of 75 J·g^{-1}. This protein is also characterized by an unusually higher value of normalized thermodynamic enthalpy than predicted. To justify such abnormal values, it has been postulated that the temperature-induced unfolded state of barnase has much less residual structure in the U-state than that of other proteins.[8] If this were the case, then a cooperative melting of all structural elements would be accompanied by higher than average enthalpy and entropy changes. This might also be the case for CD2BP2-GYF domain, and, in fact, ellipticity disappears within the near-UV region after denaturation (Figure 2). Nevertheless, the far-UV CD spectra (Figure 2) show some residual ellipticity at high temperatures, which will report for some remaining local arrangements still present in

dx.doi.org/10.1021/jp303495b | J. Phys. Chem. B 2012, 116, 7168−7175

the unfolded state of CD2BP2-GYF. This is not a strange phenomenon because the thermally unfolded state of proteins has been traditionally described as roughly compact when compared, for example, with the chemically unfolded one.[18]

Opposite to enthalpies, entropies, and heat capacities, stability curves reveal an ordinary unfolding behavior of CD2BP2-GYF domain. We calculated the temperature dependence of the Gibbs energy functions at the different pH values (Figure 7). It appears that CD2BP2-GYF is only marginally

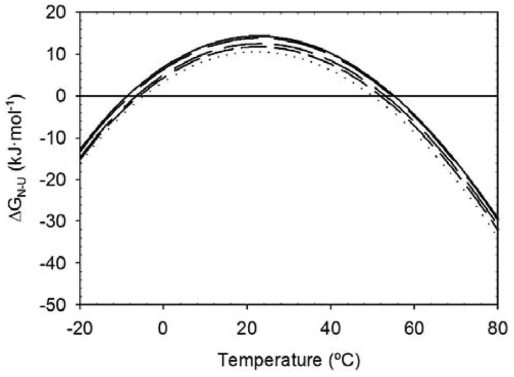


Figure 7. Gibbs energy changes upon unfolding of the CD2BP2-GYF domain found by curve fitting at several pH values. Lines represent the temperature functions at pH 6 (solid), pH 7 (long-dashed), pH 8 (medium-dashed), pH 9 (short-dashed), and pH 10 (dotted line).

stable, the Gibbs energy values being as much as ~15 kJ·mol^{-1} under maximum stability conditions (Table 1). Thus, the stability of this domain, in terms of its $\Delta G_{N-U}(298)$ values, is a bit lower than the respective of α-spectrin SH3 domain (7.1 kDa), roughly of the same size as CD2BP2-GYF domain.[7] Gibbs energies are also comparable to the ones of Nedd4-WW4, another proline-recognition domain with a lower M_w (5.7 kDa).[6] The temperature behavior of the stability curve is quite normal, with maximum stability above 20 °C as usual[19] and cold denaturation below 0 °C.[6−8]

In summary, the thermodynamic behavior of CD2BP2-GYF unfolding is characterized by unusually high enthalpy, entropy, and heat capacity changes, but not Gibbs energies, which reveal a strong enthalpy−entropy compensation effect. The database of thermodynamic parameters by Robertson and Murphy[16] has allowed us to confirm that the ΔH_{N-U}(60 °C) and the ΔS_{N-U}(60 °C) values are in the order of the respective to proteins of higher M_w than CD2BP2-GYF. Therefore, this domain behaves like a protein with a M_w ~50% higher than the current 7.5 kDa.

Embedded Water Molecules Can Explain the Abnormal Unfolding Thermodynamics of CD2BP2-GYF Domain. The compensation between enthalpy and entropy has been described as "thermodynamic homeostasis"[20] and is also associated with an anomalously high value of the heat capacity change. This phenomenon has been well-described previously as being a consequence of the implication of solvent molecules in the dynamics of a variety of biochemical systems, being described both by Gibbs energy calculations[21] and by experiments combined with molecular dynamics simulations.[22−24]

We hypothesize that solvent sequestration might occur upon folding of CD2BP2-GYF domain because the thermodynamic analysis shows a clear "homeostatic" behavior as well as anomalous heat-capacity values. Supporting our hypothesis, some energy calculations on barnase showed that the unfolding enthalpy change in vacuum, normalized per residue, is very similar to that of other globular proteins. Therefore, the larger experimental normalized enthalpy must be a result of a higher than average solvation contribution, but the authors were unable to find out the molecular origins of such extra-solvation contribution.[18] In addition, some water molecules occluded in the hydrophobic core have been described in the X-ray structure of barnase.[21]

Buried water molecules have been observed by NMR, X-ray, and other experimental methods, but the ability to measure or identify them is still a challenge because sometimes such molecules are mixed up with the own hydration shield of the proteins, and some other times they are not properly modeled during structure refinement. Nevertheless, cavities large enough to contain at least one water molecule are found in nearly all globular proteins, and the buried waters commonly occupy such cavities when observed (ref 21 and references therein). We have used The Computed Atlas of Surface Topography (CASTp) with a probe radius of 1.4 Å to analyze the cavities and pockets present in CD2BP2-GYF domain, as it has been previously done with other examples.[25,26] CASTp outputs the set of atoms (identified into their respective amino acids) surrounding each cavity/pocket as well as the number of simultaneous exits from the pocket (mouth size), being a cavity a pocket with zero exits. It also discards shallow pockets.

By using the single NMR structure of the isolated domain deposited to date at the PDB (PDB code: 1GYF), we found seven cavities/pockets, able to allocate at least a water molecule each, being such packing defects widespread through the whole molecule (Figure 8). Surprisingly, when we performed a CASTp analysis to other protein examples, we found only three cavities/pockets with proteins of similar size and whose two-state folding has been previously studied by us, like AS-48, a α-helical and circular protein,[27] or the α-spectrin SH3 domain,[7] of β-sheet topology. Furthermore, we analyzed a number of globular proteins from the database of Robertson and Murphy[16] observing that proteins of greater sizes like Barstar (PDB codes: 1BTA and 1BTB), of ~10 kDa, may display a roughly similar number of cavities/pockets as CD2BP2-GYF. It is interesting to point out also that Barnase (12.4 kDa) possess as many as 16 cavities/pockets, also higher than that displayed by proteins of similar size like Ribonuclease T1 (11.2 kDa; 10 cavities/pockets), and roughly concordant with proteins of higher size like Ribonuclease A (13.7 kDa; ~14 cavities/pockets), Lysozime (14.3 kDa; ~16 cavities/pockets), or Arc-repressor (13 kDa; 14 cavities/pockets).

Thermodynamic Homeostasis As the Consequence of Embedded Waters upon CD2BP2-GYF Unfolding. In a previous work, we have identified a similar thermodynamic behavior for the binding of peptidic ligands to Abl-SH3, another polyproline-binding domain. It is also characterized by unusually high enthalpy and entropy changes (but not Gibbs energies), which cannot be explained by a current hydrophobic interaction between the essentially hydrophobic residues directly involved in binding. We have found some solvent (water) molecules intercalated into the hydrophobic binding interface of the domain-ligand complex,[22] which seem to be universal to the whole family.[24] According to molecular dynamics simulations, these water molecules organize a dynamic network of hydrogen bonds along the binding interface; the most clear thermodynamic consequence is a strong increase in the binding enthalpy as a result of the

dx.doi.org/10.1021/jp303495b | J. Phys. Chem. B 2012, 116, 7168−7175

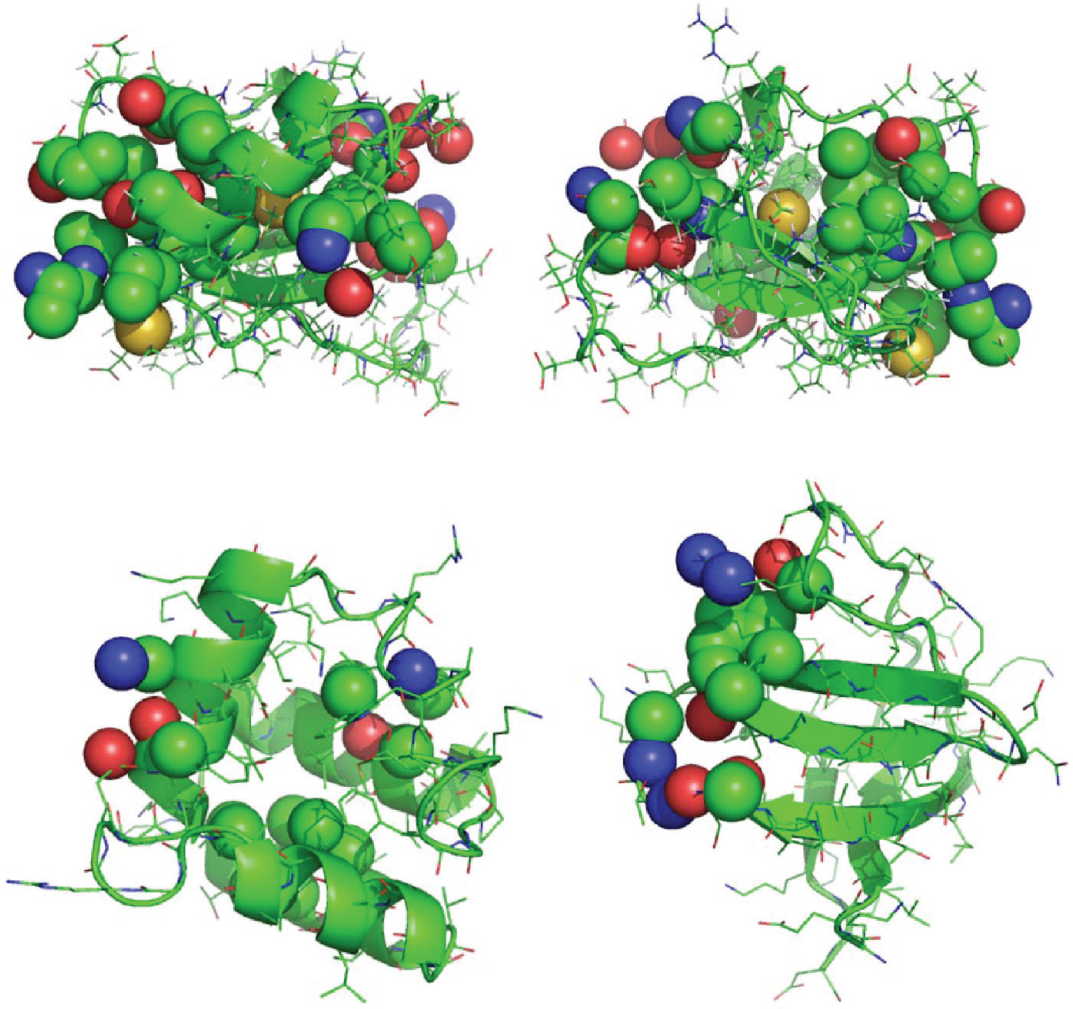


Figure 8. Upper panels show a 180° view of the analysis of cavities and pockets formed into CD2BP2-GYF domain by CASTp with a probe radius of 1.4 Å. The lower panels show the analysis of similar size proteins like AS-48 (left panel), an α-helical and circular protein, and α-spectrin SH3 domain (right panel), of β-sheet topology. The set of atoms surrounding each cavity/pocket is represented as spheres, colored according to the type of atom.

contribution to the binding energetics of several hydrogen bonds established by such water molecules. Conversely, the incomplete desolvation of the ligand−protein interface would imply an additional entropic penalty that would oppose the favorable enthalpic effects, resulting in no more than modest changes in the Gibbs energy.[22,23]

Similarly, the immediate thermodynamic consequences of embedded waters upon protein unfolding would be a strong enthalpy−entropy compensation, giving the surprising result that there is no Gibbs energy cost in transferring a water molecule from the hydration shield of the protein to an environment that can be roughly hydrophobic. Because the proteins analyzed here by CASTp are considered to be classical examples from the point of view of folding/unfolding studies, these occluded waters may thus explain the atypical enthalpy values found in Barnase and CD2BP2-GYF because an extra-amount of hydrogen bonds can be established in the N-state, up to four by each water molecule.[21]

An additional feature that emerges from this picture is that overall hydration increases the flexibility of the protein (refs 21 and 26 and references therein), which has been also detected here through the anomalously high $C_{pN}(T)$ values found in CD2BP2-GYF unfolding. Therefore, a fraction of the significant entropy change would be due to an increase in conformational freedom because the hydrogen bonds between protein atoms at the inner surface of the cavity get longer due to the presence of an intercalated buried water, which explains why the protein gets more flexible as the cavity gets hydrated.[21] The effect could be as if water acts as a lubricant into protein packing. In other words, these bound waters increase the vibrational entropy of the protein, which could also be thought as an increase in protein flexibility.[28] These dynamic fluctuations arise from the exchange with the solvent, when water molecules escape or penetrate the protein. These features have been seen both inside the proteins[29] and upon ligand binding.[23]

The dynamics of water inside the protein core will also affect protein thermostability. In a very recent work, water content has been denoted as the main means of distinguishing between mesophilic and psychrophilic (cold-adapted) protein homologues, the latter having larger average cavity sizes of probe radii of 1.4 Å than the former ones.[26] In the case of CD2BP2-GYF, the T_m values are lower (among 40−55 °C; Table 1) than the averaged 60 °C obtained by Robertson and Murphy from their data set,[16] or even, than the ones of other proline-binding domains (Figure 9). The low thermostability is closely related to the high ΔC_{pN-U} value (8.4 kJ·K^{-1}·mol^{-1}), responsible for

dx.doi.org/10.1021/jp303495b | J. Phys. Chem. B 2012, 116, 7168−7175

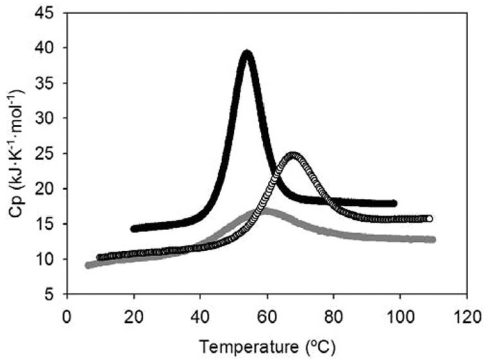


Figure 9. Comparison of the DSC traces of CD2BP2-GYF (black symbols), α-spectrin SH3 (empty symbols),[7] and hNedd4-WW4 (gray symbols)[6] domains obtained under maximal stability (pH 6, pH 4, and pH 7 respectively).

the strong curvature of the Gibbs energy function that narrows the stability interval of the domain (Figure 7). This anomalously high increase in heat capacity can be justified also by the additional degrees of freedom attained by the water molecules, which can soak up the thermal energy and thus increase the ability of the domain to absorb heat energy without any change in temperature.[20]

CONCLUSIONS

According to the most classical view of protein folding, hydrophobic interactions are considered to be highly unspecific, driving to a fluid and slippery packing, where the "uniqueness" of the native structure is achieved mainly by hydrogen bonds and other specific interactions between polar residues.[30] Therefore, it might occur that the higher than normal aromatic (hydrophobic) content of CD2BP2-GYF may not be satisfied solely by the bonding potential of polar groups, which are in a regular proportion in this domain. A plausible energetic explanation of the well-arranged N-state of this domain, despite being highly flexible and dynamic, would be the extra amount of hydrogen bonds formed by some water molecules embedded into the folded state of CD2BP2-GYF domain that will restore the polar versus nonpolar energetic balance.

AUTHOR INFORMATION

Corresponding Author

*Tel: (34)958242370. Fax: (34)958272879. E-mail: jcmh@ugr.es.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was financed by grants BIO2009-13261-C02-01/02 from the Spanish Ministry of Science and Technology, grants CVI-5915 and CVI-5063 from the Andalucian Regional Government, FEDER and Plan E. We also acknowledge Dr. Javier Ruiz-Sanz for his helpful comments.

ABBREVIATIONS

CD2BP2, CD2 binding protein 2; DSC, differential scanning calorimetry; CD, circular dichroism; N, native state; U, unfolded state; M_w, molecular weight; PDB, protein data bank; SH3, Src-homology 3 domain

REFERENCES

(1) Kofler, M.; Motzny, K.; Freund, C. *Mol. Cell. Proteomics* **2005**, *4*, 1797.
(2) Kofler, M. M.; Freund, C. *FEBS J.* **2006**, *273*, 245.
(3) Freund, C.; Dotsch, V.; Nishizawa, K.; Reinherz, E. L.; Wagner, G. *Nat. Struct. Biol.* **1999**, *6*, 656.
(4) Gill, S. C.; von Hippel, P. H. *Anal. Biochem.* **1989**, *182*, 319.
(5) Freund, C.; Kuhne, R.; Park, S.; Thiemke, K.; Reinherz, E. L.; Wagner, G. *J. Biomol. NMR* **2003**, *27*, 143.
(6) Cobos, E. S.; Iglesias-Bexiga, M.; Ruiz-Sanz, J.; Mateo, P. L.; Luque, I.; Martinez, J. C. *Biochemistry* **2009**, *48*, 8712.
(7) Viguera, A. R.; Martinez, J. C.; Filimonov, V. V.; Mateo, P. L.; Serrano, L. *Biochemistry* **1994**, *33*, 2142.
(8) Martinez, J. C.; el Harrous, M.; Filimonov, V. V.; Mateo, P. L.; Fersht, A. R. *Biochemistry* **1994**, *33*, 3919.
(9) Murciano-Calles, J.; Cobos, E. S.; Mateo, P. L.; Camara-Artigas, A.; Martinez, J. C. *Biophys. J.* **2010**, *99*, 263.
(10) Gomez, J.; Hilser, V. J.; Xie, D.; Freire, E. *Proteins* **1995**, *22*, 404.
(11) Woody, R. W. *Eur. Biophys. J.* **1994**, *23*, 253.
(12) Freire, E. Differential Scanning Calorimetry. In *Protein Stability and Folding: Theory and Practice*; Shirley, B. A., Ed.; Humana Press, Inc.: Totowa, NJ, 1995; Vol. *40*, p 191.
(13) Privalov, P. L.; Makhatadze, G. I. *J. Mol. Biol.* **1990**, *213*, 385.
(14) Cooper, A. *Proc. Natl. Acad. Sci. U. S. A.* **1976**, *73*, 2740.
(15) Makhatadze, G. I.; Privalov, P. L. *J. Mol. Biol.* **1990**, *213*, 375.
(16) Robertson, A. D.; Murphy, K. P. *Chem Rev* **1997**, *97*, 1251.
(17) Fu, L.; Freire, E. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89*, 9335.
(18) Lazaridis, T.; Archontis, G.; Karplus, M. *Adv. Protein Chem.* **1995**, *47*, 231.
(19) Felitsky, D. J.; Record, M. T., Jr. *Biochemistry* **2003**, *42*, 2202.
(20) Cooper, A. *Biophys. Chem.* **2005**, *115*, 89.
(21) Olano, L. R.; Rick, S. W. *J. Am. Chem. Soc.* **2004**, *126*, 7991.
(22) Palencia, A.; Cobos, E. S.; Mateo, P. L.; Martinez, J. C.; Luque, I. *J. Mol. Biol.* **2004**, *336*, 527.
(23) Palencia, A.; Camara-Artigas, A.; Pisabarro, M. T.; Martinez, J. C.; Luque, I. *J. Biol. Chem.* **2010**, *285*, 2823.
(24) Martin-Garcia, J. M.; Ruiz-Sanz, J.; Luque, I. *Biochem. J.* **2012**, *442*, 443.
(25) Dundas, J.; Ouyang, Z.; Tseng, J.; Binkowski, A.; Turpaz, Y.; Liang, J. *Nucleic Acids Res.* **2006**, *34*, W116.
(26) Paredes, D. I.; Watters, K.; Pitman, D. J.; Bystroff, C.; Dordick, J. S. *BMC Struct. Biol.* **2011**, *11*, 42.
(27) Cobos, E. S.; Filimonov, V. V.; Galvez, A.; Maqueda, M.; Valdivia, E.; Martinez, J. C.; Mateo, P. L. *FEBS Lett.* **2001**, *505*, 379.
(28) Fischer, S.; Verma, C.; Hubbard, R. E. *J. Phys. Chem. B* **1998**, *102*, 1797.
(29) Garcia, A. E.; Hummer, G. *Proteins* **2000**, *38*, 261.
(30) Dill, K. A. *Biochemistry* **1990**, *29*, 7133.