Dehydration Propensity of Order–Disorder Intermediate Regions in Soluble Proteins

Article *in* Journal of Proteome Research · October 2007

DOI: 10.1021/pr070208k · Source: PubMed

3 authors, including:

Ariel Fernandez
Rice University

research articles Journal of proteome research

Dehydration Propensity of Order−Disorder Intermediate Regions in Soluble Proteins

Natalia Pietrosemoli,† Alejandro Crespo,† and Ariel Fernández*

Department of Bioengineering, Rice University, Houston, Texas 77005

Received April 12, 2007

Soluble folded proteins maintain their structural integrity by properly shielding most backbone amides and carbonyls from full hydration. This structure "wrapping" entails a proper packing of the intramolecular hydrogen bonds. Thus, a poorly wrapped hydrogen bond constitutes an identifiable packing defect. Such defects are promoters of protein associations since they favor the removal of hydrating molecules. In this work we show that large clusters of packing defects generate the most significant dehydration hot spots on the protein surface, inducing a strong dielectric modulation that is reflected by a local quenching of the dielectric permittivity. The PDB-reported proteins with the largest clusters of packing defects are found to be three cancer-related transcription factors, four highly interactive proteins related to cell signaling and cytoskeleton, and a cellular prion protein. A large concentration of packing defects in a soluble protein constitutes a structural singularity that is intermediate between order and disorder. The functional implications of this singularity are investigated to delineate diverse interrelated roles. The presence of these large clusters signals a structural vulnerability, a pronounced dehydration propensity, and a strong electrostatic enhancement.

Keywords: packing defects • hydrogen bond • native disorder • interfacial water

Introduction

The structural integrity of a soluble protein is contingent on its capacity to exclude water from its amide-carbonyl hydrogen bonds.[1,2] Water-exposed intramolecular hydrogen bonds in native folds constitute structural singularities representing packing defects that have been recently characterized.[2−7] In turn, these defects favor the removal of surrounding water as a means to strengthen and stabilize the underlying electrostatic interaction, and thus are implicated in protein associations and macromolecular recognition.[2−5,8−13] The strength and stability of deficiently packed hydrogen bonds—the so-called *dehydrons*[5,6]—may be modulated by an external agent. More precisely, intramolecular hydrogen bonds which are not "wrapped" by a sufficient number of nonpolar groups may become stabilized and strengthened by the attachment of a ligand or binding partner that further contributes to their dehydration.[2−11] The net gain in Coulomb energy associated with wrapping a dehydron has been experimentally determined to be ∼4 kJ/mol.[3] The adhesive force exerted by a dehydron on a hydrophobe at 6 Å distance is ∼7.8 pN, a magnitude comparable to the hydrophobic attraction between two nonpolar moieties that frame unfavorable interfaces with water.[3] Furthermore, dehydrons are decisive factors driving association in 38% of the PDB complexes and constitute significant factors (interfacial dehydron density larger than average on individual partners) in about 95% of all complexes reported in the PDB.[2,4]

In this work we identify functional indicators of PDB-reported soluble proteins with large clusters of dehydrons. Such regions, rich in structural vulnerabilities, may be characterized as belonging to an "intermediate zone" between order and native disorder.[4] This characterization is already suggested by a strong correlation between wrapping of intramolecular hydrogen bonds (ρ, Methods) and propensity for structural disorder (f_d),[4] as shown in Figure 1. This correlation reveals that the inability to exclude water intramolecularly from preformed hydrogen bonds is causative of the loss of structural integrity. The disorder propensity is accurately quantified by a sequence-based score generated by the program PONDR-VLXT,[14−16] a predictor of native disorder that takes into account residue attributes such as hydrophilicity, aromaticity, and their distribution within the window interrogated. The disorder score ($0 \le f_d \le 1$) is assigned to each residue within a sliding window, representing the predicted propensity of the residue to be in a disordered region ($f_d = 1$, certainty of disorder; $f_d = 0$, certainty of order). Only 6% of 1100 nonhomologous PDB proteins gave false positive predictions of disorder in sequence windows of 40 amino acids. The strong correlation (over 2806 nonredundant nonhomologous PDB domains, Figure 1) between disorder score of a residue and extent of wrapping of the hydrogen bond engaging the residue (if any) implies that clusters of dehydrons correspond to extended regions of structural vulnerability within a structured domain. Hence, the characterization of

* To whom correspondence should be addressed. Phone, 713 348 3681; Fax, 713 348 3699; E-mail, arifer@rice.edu.

† These authors contributed equally to this work and should be regarded as first authors.

10.1021/pr070208k CCC: $37.00

Published on Web 08/02/2007

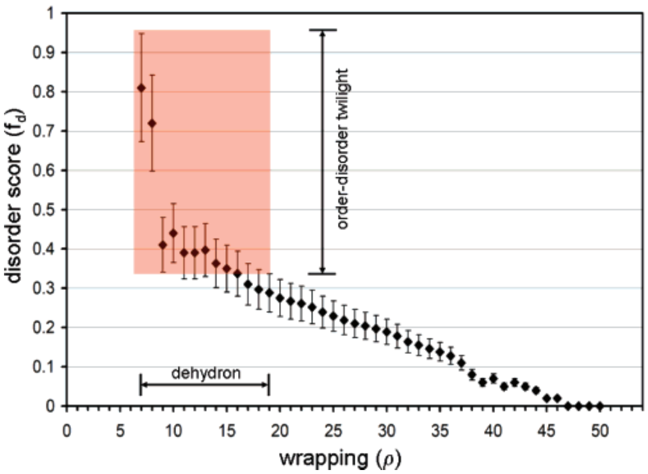

Figure 1. Correlation between the disorder score of a residue and the extent of wrapping (ρ) of the backbone hydrogen bond engaging that particular residue (if any). The disorder score on each individual residue was obtained for 2806 nonredundant nonhomologous PDB domains. Residues have been independently grouped in 45 bins of 400 residues each, according to the extent of wrapping ($7 \le \rho \le 52$). The average score has been determined for each bin (square), and the error bars represent the dispersion of disorder scores within each bin. The strong correlation between the disorder score and the extent of wrapping and the dispersions obtained imply that dehydrons can be safely inferred in regions where the disorder score is $f_d > 0.35$. The red rectangle represents the region of order−disorder intermediate region where the existence of dehydrons ($7 \le \rho \le 19$, for desolvation radius 6 Å) may be inferred from disorder score with absolute certainty. No hydrogen bond in monomeric domains reported in PDB was found to possess less than 7 wrappers, implying a threshold for structural sustainability in soluble proteins.

dehydron clusters as belonging to an order−disorder intermediate zone appears to be warranted.

Both order and native disorder are well characterized structural attributes of protein chains.[16] However, the highly vulnerable regions in a soluble fold defined in this work appear to belong to a novel category because their structural integrity is contingent on binding partnerships.[2,8]

In this work, we report on the functional role of regions with the largest dehydron clusters across the PDB. A cluster is defined as the maximal set of dehydrons with intersecting desolvation domains (Methods). These extensive regions between order and disorder are shown in this work to be strong dielectric modulators, i.e., quenchers of the local dielectric permittivity, and thus providing a local enhancement of the Coulomb fields. This property arises since clusters of packing defects promote a local dehydration due to a significant reduction in hydrogen-bonding partnerships for solvating water molecules, absent in any other region of the protein interface.

These discrete effects relating to local water structuring around packing defects cannot be captured properly by existing continuous models of the interfacial electrostatics.[17] Such models are based on mean-force potential approximations to the solvation interactions where solvent degrees of freedom are averaged out. Thus, continuous models should be adapted to deal with local dielectric modulations promoted by the interfacial regions investigated in this work.

A curated PDB database was scanned to search for proteins with the largest clusters of packing defects. These clusters, yielding the most extreme dielectric modulation, were found in three cancer-related transcription factors, in three highly interactive proteins and in a promoter for aberrant protein aggregation. The functional implications of such statistics are discussed.

Methods

Identification of Packing Defects in Soluble Proteins. The extent of hydrogen-bond wrapping, ρ, may be quantified by determining the number of nonpolar groups contained within a desolvation domain typically defined as two intersecting balls of fixed radius (~thickness of three water layers) centered at the α-carbons of the residues paired by the amide-carbonyl hydrogen bond. In structures of soluble proteins at least two-thirds of the backbone hydrogen bonds are wrapped on average by $\rho = 26.6 \pm 7.5$ nonpolar groups for a desolvation ball radius 6 Å. Dehydrons lie in the tails of the distribution, i.e., their microenvironment contains 19 or fewer nonpolar groups, so their ρ-value is below the mean ($\rho = 26.6$) minus one standard deviation ($\sigma = 7.5$).[4,6]

Hydrogen-Bonding Partnerships for Interfacial Water. The thermal average, $<\Gamma>$, of the average number of hydrogen-bond partnerships involving water molecules hydrating the p53 DNA-binding domain were obtained from classical trajectories generated by 5 ns molecular dynamics (MD) simulations. The Γ-values were determined for each water molecule within a 6 Å-radius spherical domain centered at the α-carbon of the residues. The adopted starting geometry is the monomeric structure from PDB entry 2GEQ. Similar computations were conducted for three additional domain folds, thus covering all topology classes (all β-stranded, α/β and all α-helical): the SH3 domain ($N = 55$, 2 dehydrons, PDB 1SRL); ubiquitin ($N = 76$, 16 dehydrons, PDB 1UBI), and λ-repressor ($N = 86$, 26 dehydrons, PDB 1LMB). In all cases, the starting conformation was embedded in a pre-equilibrated cell of explicitly represented water molecules and counterions.[18,19] The entire system was equilibrated for 5 ns. Computations were performed by integration of Newton's equations of motion with time step 2 fs using the GROMACS program[20] in the NPT ensemble with box size $8 \times 8 \times 8$ nm^3 and periodic boundary conditions. The box size was calibrated so that the solvation shell extended at least 12 Å from the protein surface at all times. The long-range electrostatics were treated using the Particle Mesh Ewald (PME) summation method.[21] A Nose-Hoover thermostat[22] was used to maintain the temperature at 300 K, and a Tip3P water model with OPLS (Optimized Potential for Liquid Simulations) force field[18,19] was adopted.

Results and Discussion

Packing Defects and Discrete Solvent Effects. The incompletely wrapped backbone hydrogen bonds in a soluble protein ($\rho \le 19$, for desolvation radius 6 Å) are partially exposed to solvent. These bonds favor the removal of hydrating molecules in order to enhance the polar-pair electrostatics.[2,3,5,6] Furthermore, the resulting bond stabilization overcomes the amount of work needed to remove the solvating water molecules.[3,5] To describe this dehydration propensity, we first compute the extent of constraint of interfacial water molecules. This parameter is identified with the thermal average of Γ, the average number of hydrogen-bond partnerships involving water molecules within the desolvation domain (Methods) of each residue along the chain ($0 \le \Gamma \le 4$, Figure 2). As an illustration we

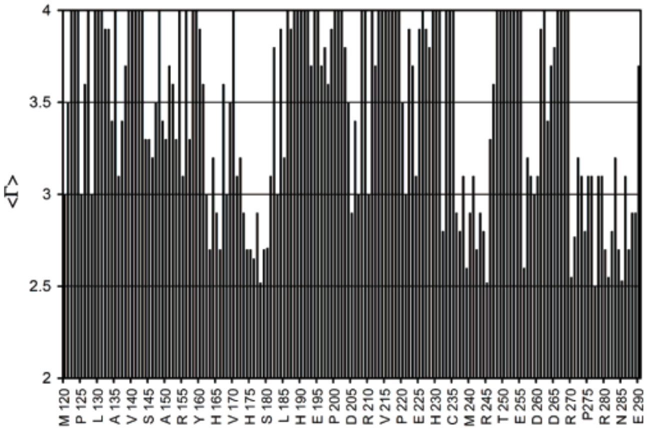

Figure 2. Thermal average of the average number of hydrogen-bond partnerships, $<\Gamma>$, for water molecules within the desolvation domain for each residue in the DNA-binding domain of p53. If no water is found in the desolvation domain (buried residue), the bulk water value $\Gamma = 4$ is adopted.

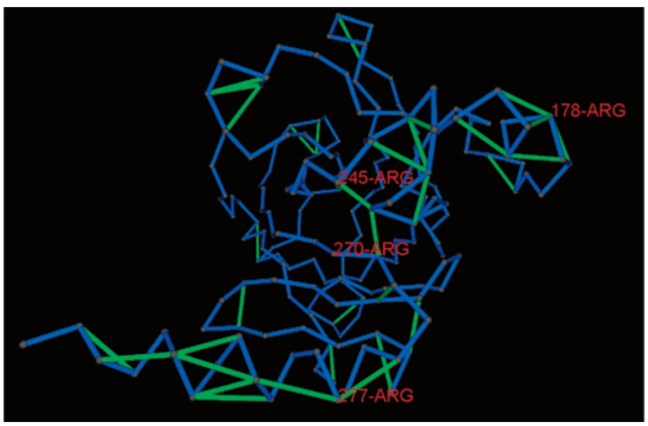

Figure 3. Dehydrons for p53 DNA-binding domain. The backbone is indicated by blue virtual bonds joining α-carbons and dehydrons are shown as green segments joining the α-carbons of residues paired by backbone hydrogen bonds.

focus on the DNA-binding domain of antitumor gene p53.[23] This domain was selected because it contains three of the largest dehydron clusters to be found in PDB (Figure 3). The functional significance of these clusters and their interplay with the electrostatics of DNA recognition by this transcription factor are reported in this work.

Figure 2 shows the thermal average, $<\Gamma>$, of the average number of hydrogen-bond partnerships involving water molecules within the desolvation domain for each residue. Three dehydration hot spots are apparent, comprising residues 171–181, 236–246, and 270–289, respectively. The location of these hot spots corresponds to the three major dehydron clusters shown in Figure 3.

As shown below, this computation was repeated to include representatives of the three major fold topologies: all-β (SH3 domain, $N = 55$, 2 dehydrons, PDB 1SRL), all-α (λ-repressor, $N = 86$, 26 dehydrons, PDB 1LMB) and α/β (ubiquitin, $N = 76$, 16 dehydrons, PDB 1UBI). Consistency with the p53 calculation was obtained, as dehydrons proved to become the dehydration hot spots on the protein interface.

The solvating-water confinement induced by a packing defect is illustrated in Figure 4. The figure displays a selected snapshot of a water molecule within the desolvation domain

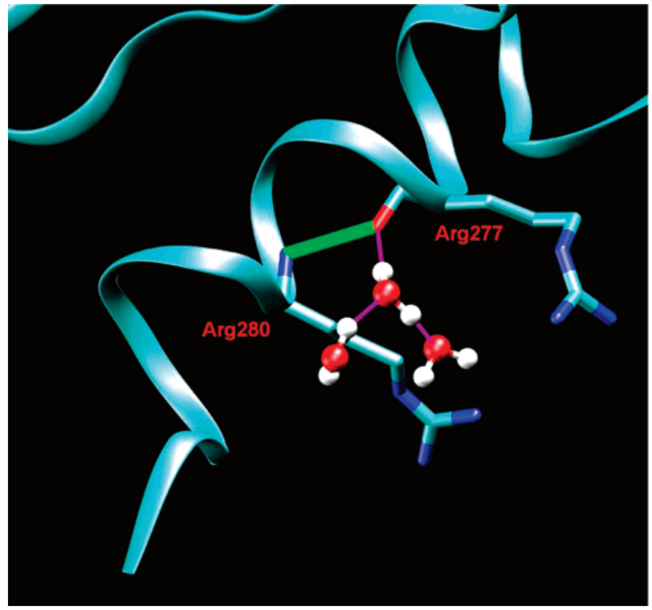

Figure 4. Selected snapshot (after 1 ns of MD) of a solvating water molecule and its hydrogen bond partnerships (purple bonds) within the desolvation domain of Arg277 in the DNA-binding domain of tumor antigen p53 (ribbon representation, fragment). The backbone amide–carbonyl dehydron Arg277–Arg280 is shown in green.

of Arg277, a residue paired by a dehydron to Arg280 (Figure 2, 3). This water molecule has 3 hydrogen-bond partners, two with vicinal water and one with the Arg277 backbone carbonyl. A 3.6 Å-threshold is adopted for hydrogen bond distance between the heavy atoms. The snapshot was extracted after 1 ns of MD simulations equilibrating the protein chain with surrounding water (Methods). Due to the incomplete wrapping of dehydron (Arg277, Arg280) in the p53 domain, the closest water molecule is found at a distance 2.8 Å between carbonyl and water oxygen atoms. While electrostatically engaged with the Arg277 backbone carbonyl, this water molecule is deprived of one hydrogen bond partnership when compared with bulk water (Figures 3 and 4).

To determine a generic relation between ρ and Γ, three additional single-domain folds representative of protein topologies were also analyzed: SH3-domain; ubiquitin, and λ-repressor, as indicated above.

The correlation between wrapping and dehydration propensity (Figure 5) has the following characteristics: (a) dehydrons ($\rho \leq 19$) generate Γ-values in the range $2 \leq \Gamma \leq 3.6$; (b) no PDB hydrogen bond is wrapped by $\rho < 7$; (c) the upper wrapping bound, $\rho = 28$ corresponds to bulk-like water ($\Gamma = 4$) in the desolvation domain; and (d) all solvating water is excluded from the desolvation domain for $\rho > 28$.

Defective Packing and Dielectric Modulation. In this section, we show how dielectric modulation is promoted by discrete solvent effects arising from an incomplete packing of the protein backbone. These discrete effects cannot be captured properly by conventional continuous models, which should be adapted to deal with local dielectric modulations. As demonstrated in this section, the dielectric modulation, i.e., the quenching in the local dielectric permittivity is caused by the local reduction in hydrogen-bond partnerships of solvating water molecules.[24] To quantify this effect, we compute the polarizability associated with restricted interfacial water under

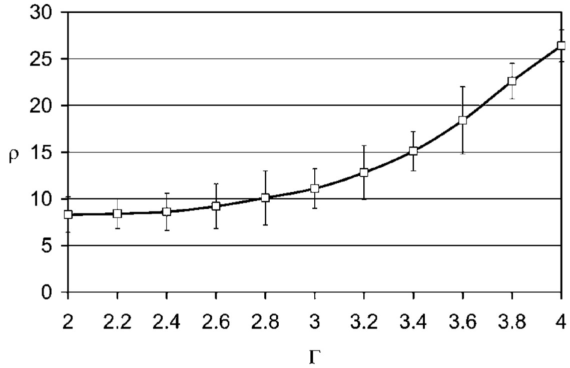

Figure 5. Correlation between hydrogen-bond wrapping ρ and wetting parameter Γ. Each residue is assigned a ρ-value averaged over all backbone hydrogen bonds in which it is engaged. The data was extracted from the wetting computation on the p53 DNA-binding domain and three additional folds: the SH3 domain (N = 55, 2 dehydrons, PDB 1SRL); ubiquitin (N = 76, 16 dehydrons, PDB 1UBI), and λ-repressor (N = 86, 26 dehydrons, PDB 1LMB).

the influence of a nonzero net internal electrostatic field **E**. By contrast, in bulk water, the average net internal field is zero as there is no preferred dipole orientation and therefore, no collective net effect. Thus, we first define the entropy loss[25] of a water molecule associated to the reduction of hydrogen-bond exchange possibilities from 4 possibilities in bulk water to Γ possibilities at the interface: $\Delta s(\Gamma) = -k \ln(\Gamma/4)$ (k = Boltzmann constant), and the related dimensionless parameter $\lambda(\Gamma) = T\Delta s(\Gamma)/(Ld)$, where d = dipole moment of a single water molecule, L = Lorentz field $= \eta d/3\epsilon_o$, with η = bulk water density and ϵ_o = vacuum permittivity. The thermal average of the dipole moment vector **d** is given by: $<\mathbf{d}> = d<\mu>$, where $\mu = \mu(\mathbf{E})$ is the dipole moment unit vector. The thermal average of the dipole projection μ_E onto the net internal field may be determined from the Boltzmann average which gives:[26]

$$<\mu_E> = [\coth(\beta Ed) - (\beta Ed)^{-1}] \qquad (1)$$

where $\beta = (kT)^{-1}$, $E = ||\mathbf{E}||$. The scattering $\sigma(\mathbf{E})$ of the net field produced by dipole interactions within volume V is the real part of

$$S = V - \int \exp\,[-i\nu(\mathbf{r})\rho<\mu>]dr \qquad (2)$$

where $\nu(\mathbf{r}) = d/(4\pi\epsilon_o r^3)$ is the dipole interaction field and ρ = net field unit vector. Integration in eq 2 is carried out in the region of correlation: $r_w \leq ||\mathbf{r}|| \leq r_\Gamma$, with $r_\Gamma = r_w[1 - \lambda(\Gamma)]^{1/3}$, where r_w is the bulk dipole−dipole correlation distance. Thus, the field scattering is given as:

$$\sigma(E) = \mathrm{Re}S = \lambda(\Gamma)L^2(|<\mu_E>|^2)/18 = \eta kT[27\epsilon_o]^{-1}\ln(4/\Gamma)[\coth(\beta Ed) - (\beta Ed)^{-1}]^2 \qquad (3)$$

The probability distribution of net internal fields is then

$$P_\Gamma(\mathbf{E}) = [\pi\sigma^2(E)]^{-3/2}\exp[-E^2/\sigma^2(E)] \qquad (4)$$

satisfying:

$$\lim_{\Gamma 4} P_\Gamma(\mathbf{E}) = \delta(\mathbf{E}) \qquad (5)$$

in accord with the fact that no net internal field is expected in bulk water. The dielectric susceptibility χ along a preferential

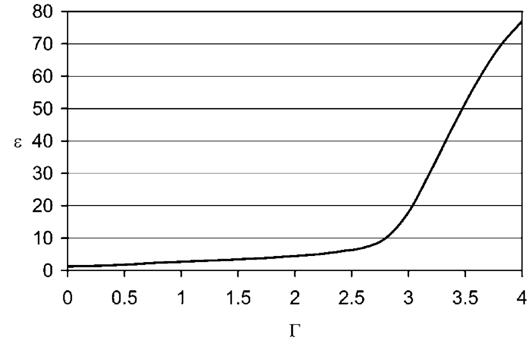

Figure 6. Analytically dependence of the dielectric permittivity ϵ on the parameter Γ.

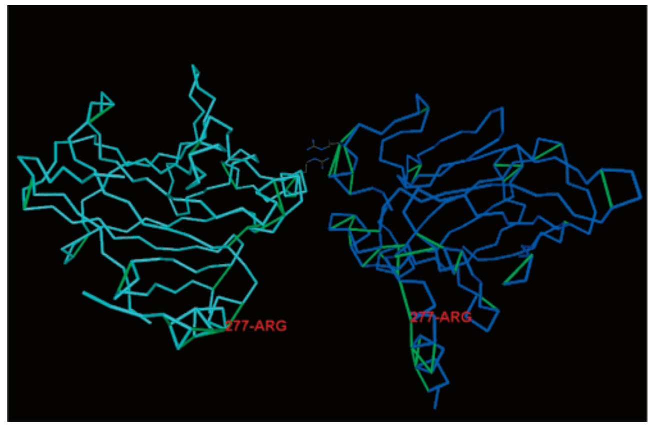

Figure 7. Backbone/dehydron representation of the dimmer interface for the DNA-binding domain of p53 (PDB 2GEQ). The side chains of the Arg178 of each monomer involved in a resonance pair are shown.

direction z is directly computed as function of Γ by introducing a perturbation Δ_z of the net internal field $\tilde{E} = \mathbf{E} + \Delta_z$.

By definition:

$$\chi_z = \lim_{||\Delta|| \to 0} \partial \Xi_z / \partial \Delta_z \qquad (6)$$

where $\Xi_z = \eta d \int [\coth(\beta\tilde{E}_z d) - (\beta\tilde{E}_z d)^{-1}]P_\Gamma(\mathbf{E})\, d(\mathbf{E})$ is the polarizability along the z-direction. This gives:

$$\chi_z = \chi_z(\Gamma) = \eta d^2\beta/3\epsilon_o - \eta d^2\beta/3\epsilon_o \int [\coth^2(\beta E_z d) - (\beta E_z d)^{-2}]P_\Gamma(\mathbf{E})\, d(\mathbf{E}) \qquad (7)$$

where $\eta d^2\beta/3\epsilon_o = \chi_{zw}$ is the bulk water susceptibility. The bulk limit is obtained substituting eq 5 in eq 7 in the limit $\Gamma \to 4$:

$$\lim_{\Gamma \to 4} \chi_z = \eta d^2\beta/3\epsilon_o - \eta d^2\beta/3\epsilon_o \int [\coth^2(\beta E_z d) - (\beta E_z d)^{-2}]\delta(\mathbf{E})\, d(\mathbf{E}) = \eta d^2\beta/3\epsilon_o = \chi_{zw} \qquad (8)$$

since $\lim_{\xi \to 0} [\coth^2(\xi) - (\xi)^{-2}] = 0$.

Figure 6 displays the derived Γ-dependence of the dielectric permittivity $\epsilon = 1 + \chi(\Gamma)$ (the subindex z denoting a generic direction has been dropped). The dielectric quenching is extreme upon moderately small losses in hydrogen-bond partnerships. Thus, the most dramatic decrease is marked by a drop in ϵ-values from 50 to 7 as Γ is reduced from 3.5 to 2.5.

By combining Figures 5 and 6, we may conclude that *clusters of packing defects act as dramatic enhancers of the electric fields generated at the protein interface.* The typical loss in hydrogen-

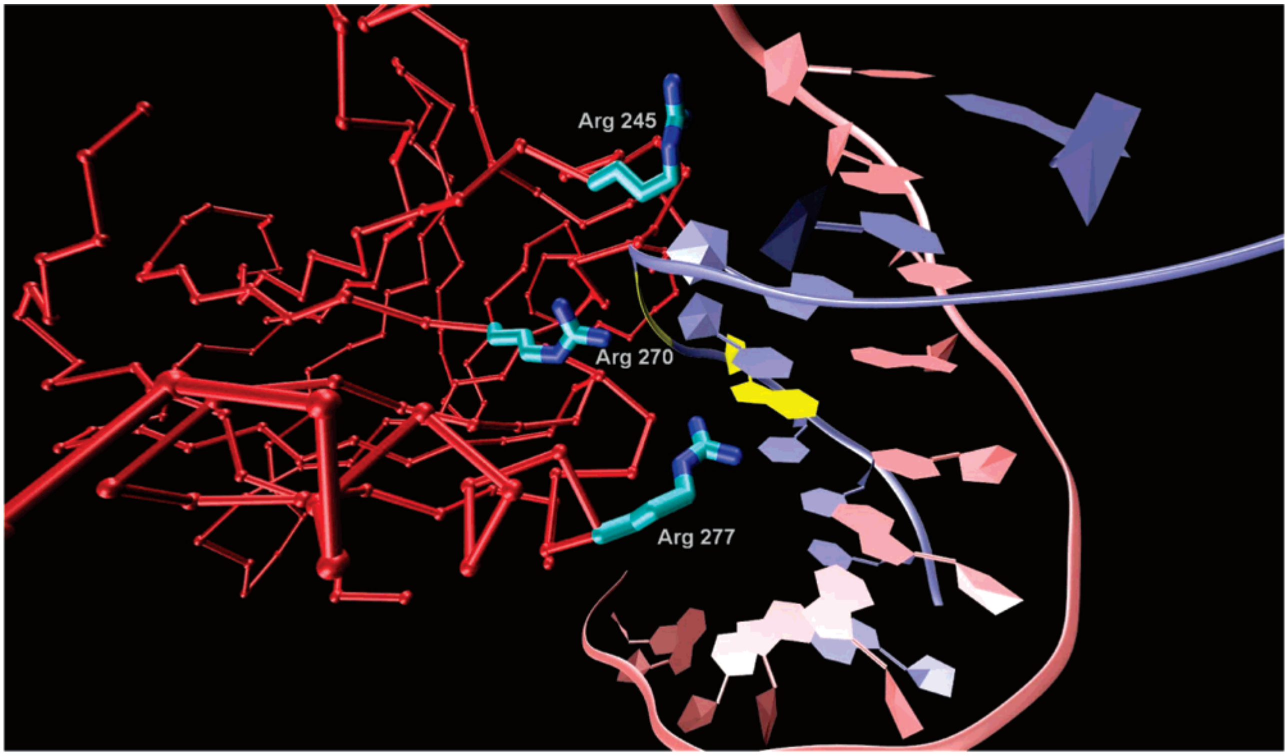

Figure 8. Protein–DNA complex of the DNA-binding domain of p53 (PDB 2GEQ). Side chains of the key residues directly implicated in DNA recognition, Arg245, Arg270, and Arg277 are shown. The pyridine base recognized by Arg277 is shown in yellow, whereas the individual DNA strands are shown in lilac and light magenta.

bonding partnerships associated with dehydrons solvation places Γ in the range $2 \le \Gamma \le 3.6$. This interval contains the region of most dramatic dielectric quenching, decreasing the permittivity by an order of magnitude with respect to bulk water. In turn, this effect translates in an order of magnitude increase in electrostatic interactions.

Discrete Dielectric Quenching in the p53 DNA-Binding Domain: A Study Case. The functional significance of the three dielectric modulators in the DNA-binding domain of p53 may be understood by examining its dimeric state and its role as transcription factor. Thus, a major cluster involving the 5 dehydrons (173, 176), (174, 178), (175, 178), (176, 179), and (178, 180) is found at the dimer interface (Figures 3 and 7). This cluster fosters dimerization in accord with the dehydration propensity of dehydrons and their role as promoters of protein associations.[2,5,10−13] The dimerization involves a resonant pairing of the Arg178 from each monomer (Figure 7) likely, to promote supramolecular charge delocalization with distal charge separation at all times. Significantly, $\langle\Gamma\rangle$ reaches a minimum precisely at Arg178 (Figure 2), in accordance with the low dehydration penalty for the guanidinium ion[27] and with the fact that the Arg–Arg resonant association requires guanidinium dehydration (cf. Figure 7).

Besides Arg 178, there are three additional minima in $\langle\Gamma\rangle$, corresponding to residues Arg245, Arg270, and Arg277 (Figure 2). The latter three arginines play a pivotal role in DNA recognition,[23] as discussed below. Residue Arg245 is engaged in the dielectric quenching region 236–246 and is part of the dehydron cluster involving pairs (236, 239), (237, 245), (237, 271), (239, 242), (239, 244), and (240, 242) (Figure 3). Residues 270 and 277 lie within the dielectric quenching region 270–289 and are part of the dehydron cluster involving pairs (237, 271), (274, 277), (277, 281), (280, 284), (281, 285), (282, 285), and (285, 288) (Figures 3 and 8). The latter cluster is one of the largest to be found in a PDB-reported soluble protein, as discussed below.

Direct examination of the protein-DNA complex (PDB 2GEQ), reveals that the three residues directly implicated in DNA recognition are precisely Arg245, Arg270, and Arg277 (Figure 8). Residue Arg277 acts as intra-base intercalator, while the other two interact with the negatively charged backbone phosphates (Figure 8). The electrostatics of protein-DNA recognition is not merely the result of matching charges along the geometrically compatible interfaces, otherwise ion pairs would prevail in water. Rather, electrostatic recognition of the DNA polyelectrolyte requires a device to promote dehydration at the protein-nucleic acid interface. The large dehydron clusters surrounding the three arginines directly implicated in the protein-DNA association (Figures 3 and 8) provide such an expedient, as they quench the local dielectric (Figure 2), thus enhancing the electrostatic recognition. Thus, the fact that the three arginines involved in DNA recognition are also dehydration hot spots is not adventitious, but a functional necessity for the transcription factor.

The Most Defectively Packed Protein Domains. A curated PDB-derived database was constructed free of redundancy for single-domain proteins with contour length $N > 50$. The database is composed of 2982 entries with <50% pairwise identity in aligned sequences. Proteins were grouped according to n, the size of the dehydron cluster in the structure. The groups intersect to a considerable extent, and their respective populations decrease monotonically with an approximate power law $n^{-1/2}$ (Figure 9, inset).[28]

Each n-group has been lumped into five nondisjoint functional categories: biosynthesis, enzymology, cell signaling, cytoskeleton, and cancer. The contribution of each category

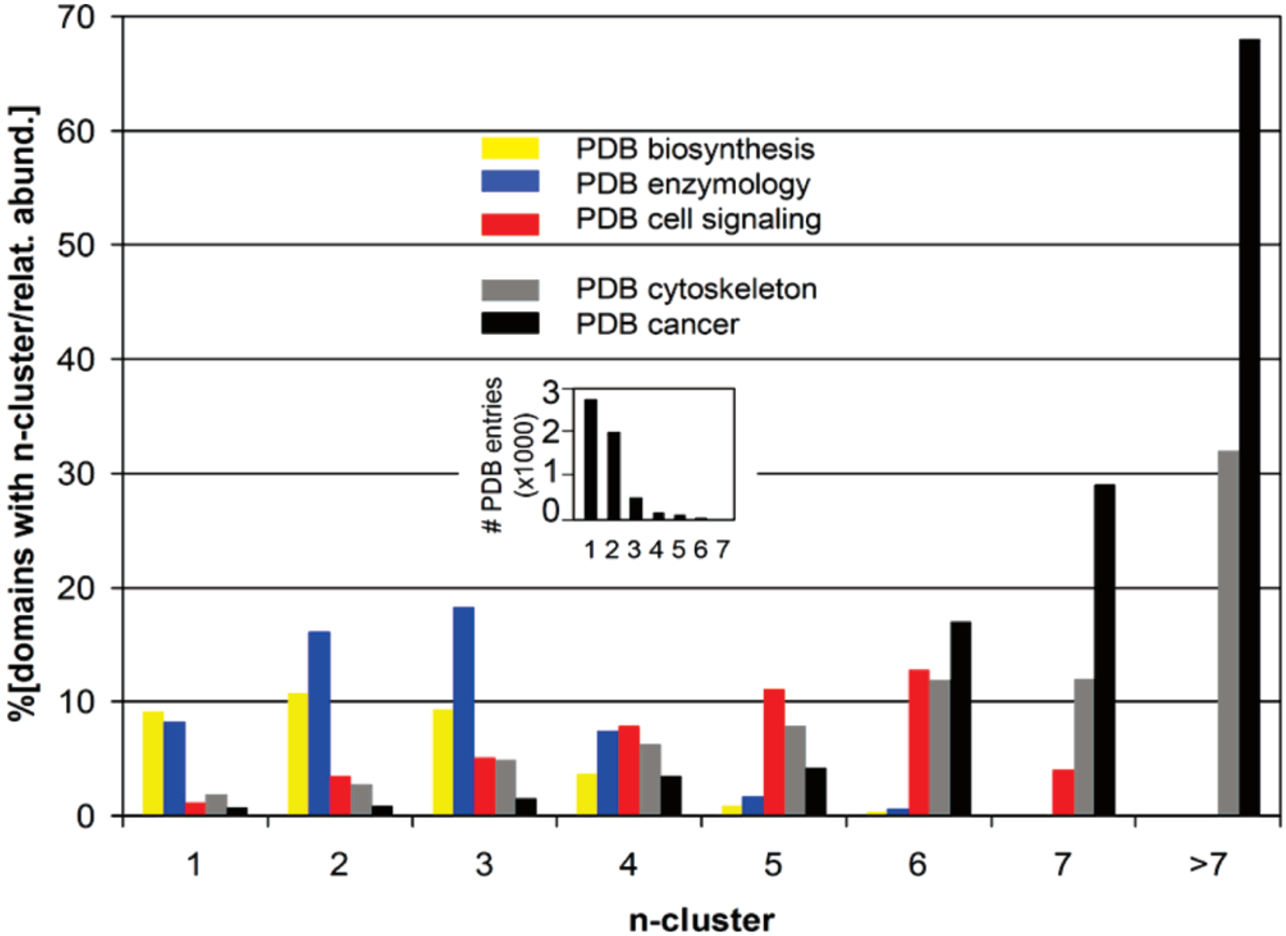

Figure 9. Percentages of PDB-domains in functional categories binned into groups determined by dehydron-cluster size *n*. Each cluster-size group is divided into five nondisjoint functional categories: biosynthesis, enzymology, cell signaling, cytoskeleton, and cancer. The number of PDB domains in each group is normalized to the relative abundance of the functional category. Thus, the number of PDB-domains in a cluster-size group and functional category is divided by the total number of PDB domains in the category. (Inset) Number of domains in each cluster-size group.

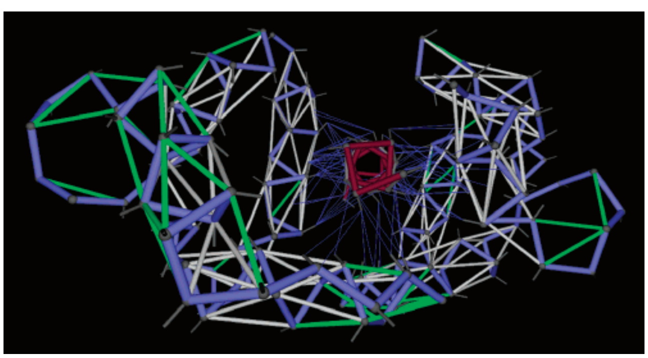

Figure 10. Dehydron pattern of calmodulin (PDB 1CDM) complexed with the calmodulin-binding domain of calmodulin-dependent protein kinase II. The backbone is indicated by blue virtual bonds joining α-carbons and dehydrons are shown in green. The intermolecular wrapping of calmodulin hydrogen bonds by the ligand side chains is indicated by thin blue lines. The extremities of these lines are the β-carbon of the residue contributing to the desolvation shell of the hydrogen bond and the midpoint of the hydrogen bond that is being wrapped intermolecularly.

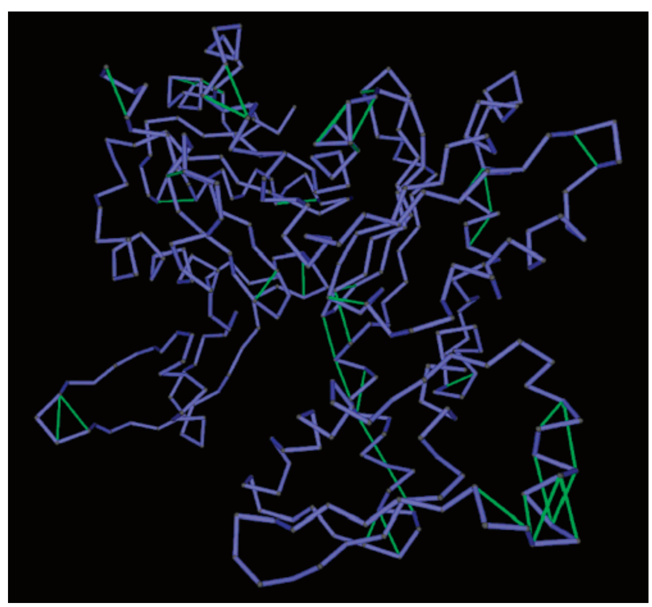

Figure 11. Dehydron pattern of actin (PDB 1ATN).

to each *n*-group is quantified through suitable normalization, taking into account the relative abundance of each category in the curated database. Thus, the number of PDB-domains within an *n*-cluster in a functional category is divided by the total number of PDB domains in the category. After normaliza-

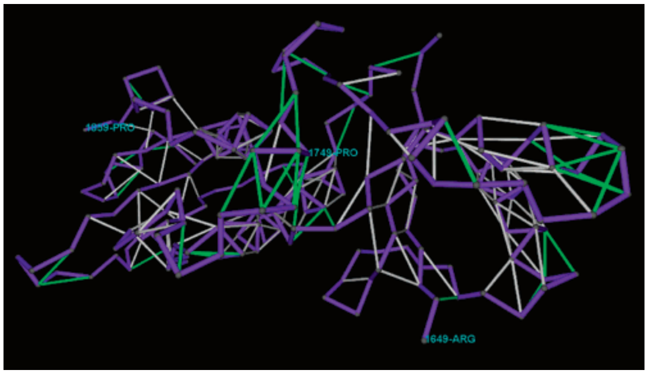

Figure 12. Dehydron pattern of BRCT, the terminal repeat domain of breast cancer gene BRAC1 (PDB 1JNX).

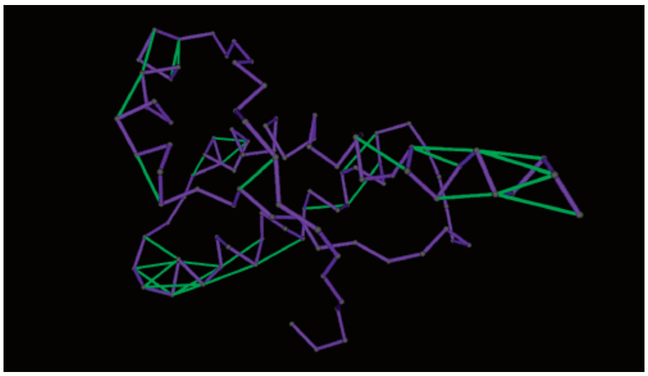

Figure 13. Dehydron pattern of the cellular prion protein (PDB 1QM0).

tion, percentages indicating representativity from each functional category are computed for each n-group (Figure 9).

The distribution of dehydron clusters becomes a marker of the functional category. Thus, biosynthesis peaks at 2, enzymology at 3, cell signaling at 6, whereas cytoskeleton and cancer are monotonically increasing with dehydron cluster size. *Intriguingly, the dominance of the cancer category becomes apparent for the poorest packed proteins.* The dearth of domains for $n \geq 7$ facilitates individual identification and assessment. Thus, there are only 5 domains with $n = 7$: calmodulin (cell signaling, Figure 10),[29] actin (cytoskeleton, Figure 11),[30] p53 DNA-binding domain (cancer, Figures 3 and 7),[23] BRCT, the terminal repeat domain of breast cancer gene BRAC1 (cancer, Figure 12),[31] and the cellular prion protein (not categorized, Figure 13).[32] The group with $n > 7$ only contains 3 members: severin (cytoskeleton, Figure 14)[33] and two oncogenic transcription factors with DNA-stabilizing induced fit, jun/fos (Figure 15)[34] and myc/max (Figure 16).[35]

The eight protein domains with unusually large dehydron clusters possess a high level of network centrality since they are highly interactive proteins.[36,37] In spite of their functional diversity, a common functional motif is discerned: as soluble proteins, they all possess a major weakness in the hydration shell. Thus, the dehydration hot spot plays manifold interrelated roles: (a) promoter of protein associations (calmodulin, actin, severin), (b) dielectric modulator enhancing intermolecular electrostatic interactions (cancer-related transcription factors), and (c) a structural weakness promoting water attack on backbone hydrogen bonds with concurrent refolding leading to aggregation (cellular prion protein).

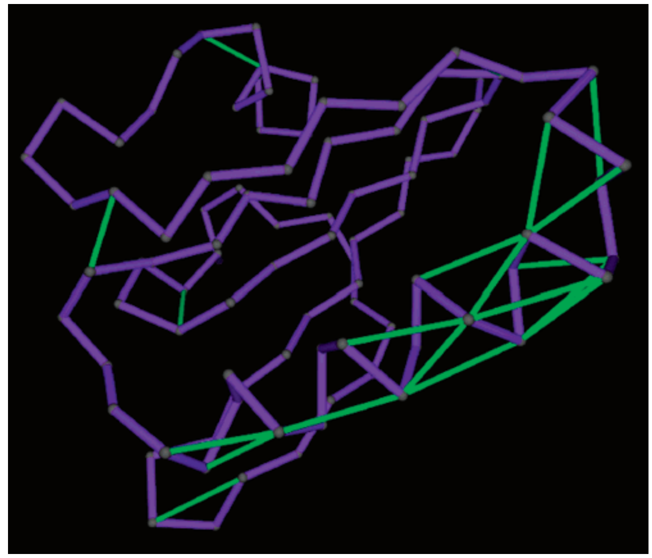

Figure 14. Dehydron pattern of severin (PDB 1SVR).

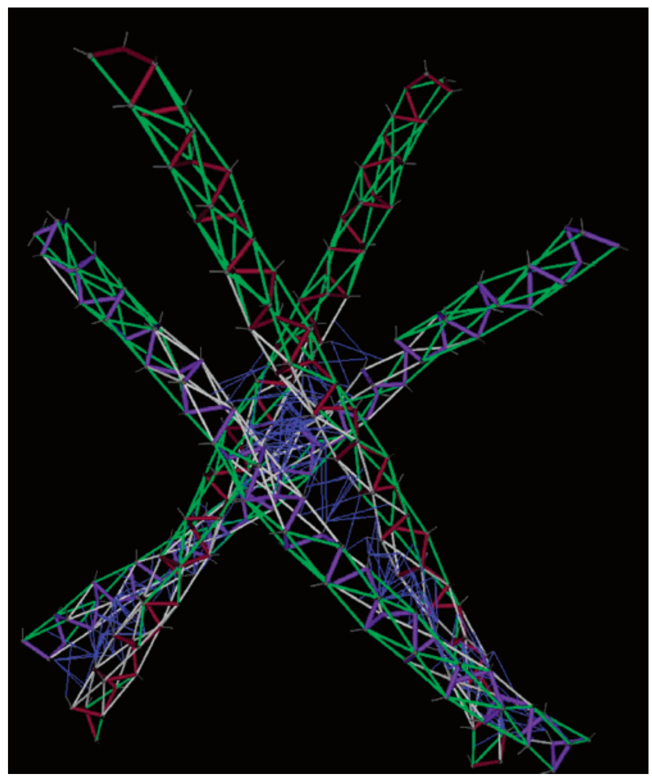

Figure 15. Dehydron pattern of oncogenic transcription factor jun/fos (tetramer, PDB 1FOS). Intermolecular wrapping is displayed as in Figure 10. Monomeric *jun* chains are in blue, *fos* chains, in dark red.

Conclusions

This work identifies a discrete solvent effect associated with defective packing in soluble folded proteins. Our analysis required a description of interfacial water that cannot be properly captured by conventional continuous models, where all solvent degrees of freedom are averaged out.

Soluble proteins are typically well protected from water attack, possessing a tight hydration shell and shielded intramolecular hydrogen bonds. By correlating packing defects with the level of confinement of hydrating water, we showed that

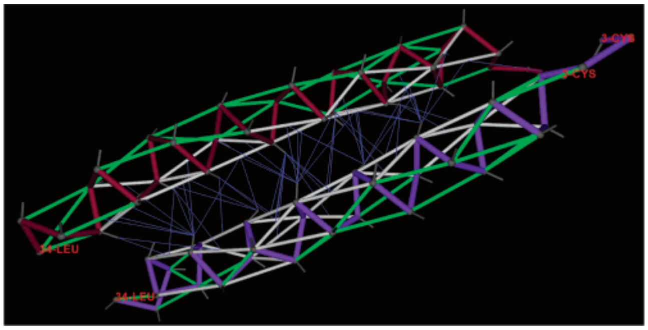

Figure 16. Dehydron pattern of oncogenic transcription factor myc/max (PDB 1A93). Intermolecular wrapping is displayed as in Figure 10. Monomeric *myc* chain is in blue, *max* chain, in dark red.

proteins with deficiently packed hydrogen bonds (known as dehydrons) possess severe weaknesses in the hydration shell and are consequently capable of a local and discrete dielectric quenching. This property becomes most apparent for proteins possessing the largest clusters of packing defects in the PDB (7 or more dehydrons). We found that there are eight such proteins reported in PDB, possessing structural regions intermediate between order and disorder. These regions belong to cancer-related (oncogenic or antitumor) proteins and also to highly interactive proteins and to a cellular prion, promoting misfolding and aberrant aggregation.

Several attributes characterize these order−disorder intermediate regions: high structural vulnerability, local dehydration propensity, and strong electrostatic enhancement on the protein surface. Although these attributes may serve as functional indicators, the dearth of PDB-reported proteins precludes any definite statement linking such attributes with cancer-related dysfunctionality at this time.

Acknowledgment. The research of A.F. is supported by NIH grant 1R01 GM072614 (NIGMS) and by a grant from the John and Ann Doerr Fund for Computational Biomedicine (Program GC4R 2005). We thank Xi Zhang and Jianping Chen for their valuable help with graphics and wrapping computations.

References

(1) Fernández, A.; Sosnick, T. R.; Colubri, A. *J. Mol. Biol.* **2002**, *321*, 659−675.
(2) Fernández, A.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 113−118.
(3) Fernández, A.; Scott, L. R. *Phys. Rev. Lett.* **2003**, *91*, 018102.
(4) Fernández, A.; Berry, R. S. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 13460−13465.
(5) Fernández, A.; Scott, R. L. *Biophys. J.* **2003**, *85*, 1914−1928.
(6) Fernández, A. *Nat. Biotechnol.* **2004**, *22*, 1081−1084.
(7) Fernández, A.; Scott, R. L.; Berry, R. S. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 323−328.
(8) Fernández, A.; Kardos, J.; Scott, R.; Goto, Y.; Berry, R. S. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 6446−6451.
(9) Fernández, A.; Berry, R. S. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 2391−2396.
(10) Deremble, C.; Lavery, R. *Curr. Opin. Struc. Biol.* **2005**, *15*, 171−175.
(11) Ma, B.; Elkayam, T.; Wolfson, H.; Nussinov, R. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 5772−5777.
(12) Ma, B.; Pan, Y.; Gunasekaran, K.; Venkataraghavan, R. B.; Levine, A. J.; Nussinov, R. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 3988−3993.
(13) Rajamani, D.; Thiel, S.; Vajda, S.; Camacho, C. J. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 11287−11292.
(14) Dunker, A. K.; Cortese, M. S.; Romero, P.; Iakoucheva, L. M.; Uversky, V. N. *FEBS J.* **2005**, *272*, 5129−5148.
(15) Iakoucheva, L. M.; Dunker, A. K. *Structure (London)* **2003**, *11*, 1316−1317.
(16) Dunker, A. K.; Obradovic, Z. *Nat. Biotechnol.* **2001**, *19*, 805−806.
(17) Wade, R. C.; Gabdoulline, R. R.; Lüdemann, S. K.; Lounnas, V. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 5942−5949.
(18) Rizzo, R. C.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1999**, *121*, 4827−4836.
(19) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.
(20) Lindahl, E.; Hess, B.; Van der Spoel, D. *J. Mol. Model.* **2001**, *7*, 302−317.
(21) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(22) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695−1697.
(23) Ho, W. C.; Fitzgerald, M. X.; Marmorstein, R. *J. Biol. Chem.* **2006**, *281*, 20494−20502.
(24) Fernández, A. *J. Chem. Phys.* **2003**, *119*, 5315−5319.
(25) Ben-Naim, A. *Hydrophobic Interactions*; Plenum Press: New York, 1980.
(26) Debye, P. *Polar Molecules*; Dover Publications: New York, 1929.
(27) Mason, P. E.; Neilson, G. W.; Dempsey, C. E.; Barnes, A. C.; Cruickshank, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 4557−4561.
(28) Barabási, A-L.; Albert, R. *Science* **1999**, *286*, 509−512.
(29) Kabsch, W.; Mannherz, H. G.; Suck, D.; Pai, E. F.; Holmes, K. C. *Nature* **1990**, *347*, 37−44.
(30) Meador, W. E.; Means, A. R.; Quiocho, F. A. *Science* **1993**, *262*, 1718−1721.
(31) Williams, R. S.; Green, R.; Glover, J. N. *Nat. Struct. Biol.* **2001**, *8*, 838−842.
(32) Zahn, R.; Liu, A.; Luhrs, T.; Riek, R.; von Schroetter, C.; Lopez Garcia, F.; Billeter, M.; Calzolai, L.; Wider, G.; Wuthrich, K. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 145−150.
(33) Schnuchel, A.; Wiltscheck, R.; Eichinger, L.; Schleicher, M.; Holak, T. A. *J. Mol. Biol.* **1995**, *247*, 21−27.
(34) Glover, J. N.; Harrison, S. C. *Nature* **1995**, *373*, 257−261.
(35) Lavigne, P.; Crump, M. P.; Gagne, S. M.; Hodges, R. S.; Kay, C. M.; Sykes, B. D. *J. Mol. Biol.* **1998**, *281*, 165−181.
(36) Han, J. D.; Bertin, N.; Hao, T.; Goldberg, D.; Berriz, G. F.; Zhang, L. V.; Dupuy, D.; Walhout, A. J.; Cusick, M.; Roth, F. P.; Vidal, M. *Nature* **2004**, *430*, 88−93.
(37) Jeong, H.; Mason, S. P.; Barabási, A. L.; Oltvai, Z. N. *Nature* **2001**, *411*, 41−42.

PR070208K