10030 *Biochemistry* **2009,** *48,* 10030–10037
DOI: 10.1021/bi900760n

BIOCHEMISTRY
including biophysical chemistry & molecular biology
Article

Structure-Based Stability Analysis of an Extremely Stable Dimeric DNA Binding Protein from *Sulfolobus islandicus*[†]

Ulrich Weininger,[§,∥] Markus Zeeb,[⊥,#] Piotr Neumann,[▽] Christian Löw,[§,@] Milton T. Stubbs,[∥,▽] Georg Lipps,[⊥,○] and Jochen Balbach*[,§,∥]

[§]*Institut für Physik, Biophysik, Martin-Luther-Universität Halle-Wittenberg, D-06120 Halle (Saale), Germany,* [∥]*Mitteldeutsches Zentrum für Struktur und Dynamik der Proteine (MZP), Martin-Luther-Universität Halle-Wittenberg, D-06120 Halle (Saale), Germany,* [⊥]*Laboratorium für Biochemie, Universität Bayreuth, D-95440 Bayreuth, Germany, and* [▽]*Institut für Biotechnologie, Martin-Luther-Universität Halle-Wittenberg, D-06120 Halle (Saale), Germany.* [#]*Present address: Department of Lead Discovery, Boehringer Ingelheim Pharma GmbH & Co. KG, D-88397 Biberach an der Riss, Germany.* [@]*Present address: Department of Medical Biochemistry and Biophysics, Karolinska Institutet, 17177 Stockholm, Sweden.* [○]*Present address: University of Applied Sciences Northwestern Switzerland, CH 4132 Muttenz, Switzerland.*

Received May 4, 2009; Revised Manuscript Received September 14, 2009

ABSTRACT: ORF56 is a small and thermodynamically extremely stable dimeric protein from the archaeon *Sulfolobus islandicus*. This DNA binding protein is encoded on plasmid pRN1 and possibly controls the copy number of the plasmid. We report the solution NMR structure as well as the crystal structure of ORF56 comprising a ribbon−helix−helix fold. The homodimer consists of an antiparallel intersubunit β-sheet and two α-helices per monomer, which is a common DNA binding fold of plasmid- and phage-encoded gene regulation proteins. NMR titration experiments with ORF56 and double-stranded DNA derived from its promoter binding site revealed that it is largely the β-sheets that interact with the DNA. The β-sheet experiences high local fluctuations, which are conserved among DNA binding ribbon−helix−helix dimers from mesophilic and hyperthermophilic organisms. In contrast, residues strongly protected against H−D exchange are localized in helix 2, forming the hydrophobic intermolecular core of the dimer. A structure-based comparison of the intermolecular binding surface and the change in accessible surface area upon unfolding of various ribbon−helix−helix dimers with the Gibbs free energy changes and *m* values show a correlation between hydrophobicity of these surface areas and stability. These findings provide possible explanations for the very high thermodynamic stability of ORF56 with retained DNA binding capacity.

Plasmid pRN1, a member of the highly conserved archaeal pRN plasmid family (*1*), is hosted as a high-copy number plasmid in the crenarchaeot *Sulfolobus islandicus*. This hyperthermophilic and acidophilic organism grows in acidic hot springs with optimal growth conditions of 80 °C and pH 3. *S. islandicus* is a model organism for studying basic principles of archaeal life (*2*) and thermodynamic protein stability.

pRN1 (5.3 kb) is fully sequenced and encodes three open reading frames (*orf56*, *orf80*, and *orf904*). The corresponding proteins have been described previously. ORF80 is also a highly sequence specific DNA binding protein; however, its role in replication is unclear (*3*). ORF904, the assumed initiator protein of plasmid replication, harbors ATPase, primase, helicase, and DNA polymerase activities (*4*). The crystal structure of the N-terminal region, which contains primase and polymerase activity, reveals a structural relationship to eukaryotic primases (*5*). *orf904* is located downstream of *orf56*, and both genes share a common promoter.

ORF56 is a small (6.5 kDa per monomer) dimeric protein. It shows a low degree of sequence homology (< 20%) to some bacterial ribbon−helix−helix (rhh)[1] proteins, such as the Arc repressor, CopG, and MetJ. Tetrameric ORF56 binds specifically to the DNA of the imperfect inverted repeat within its own promoter region (*6*), indicating that ORF56 controls the transcription of *orf904*, which in turn encodes the initiator protein and therefore acts as a plasmid copy number control protein.

Besides its function as a DNA binding protein, ORF56 shows a very high thermodynamic stability against unfolding by temperature, pH, and denaturants (*7*). The Gibbs free energy of unfolding (ΔG_U) is significantly higher compared to those of the wild type and stabilized variants of the structurally homologous Arc repressor (*8*, *9*), a well-characterized model system for the study of folding mechanisms of dimeric proteins.

In this study, we report the structure of dimeric ORF56 determined by NMR spectroscopy and X-ray crystallography together with a "structure-based" analysis of the thermodynamic stability of rhh proteins (*8*, *10*−*12*). The high stability can be explained by a high fraction of buried hydrophobic residues and of residues at the intermolecular interface. These regions are well-protected from amide proton exchange. Remaining local fluctuations were found for the double-stranded DNA (dsDNA)

[†]This research was supported by grants by the Deutsche Forschungsgemeinschaft (DFG BA1821/1-4) and the European Regional Development Fund (ERDF).

*To whom correspondence should be addressed: Institut für Physik, Biophysik, Martin-Luther-Universität Halle-Wittenberg, D-06120 Halle (Saale), Germany. Telephone: ++49 345 55 28550. Fax: ++49 345 55 27161. E-mail: jochen.balbach@physik.uni-halle.de.

[1]Abbreviations: rhh, ribbon−helix−helix; H−D, hydrogen−deuterium; NOE, nuclear Overhauser enhancement; hNOE, ^{15}N heteronuclear NOE; GdmCl, guanidinium chloride; MYL, R31M/E36Y/R40L variant of the Arc repressor; ASA, change in accessible surface area upon unfolding; rmsd, root-mean-square deviation.

 Published on Web 09/15/2009

binding site located on the intermolecular antiparallel β-sheet even for the hyperthermophilic rhh representative. This seems to be a general feature of dsDNA binding rhh dimers.

MATERIALS AND METHODS

Protein Expression and Purification. Cloning, expression, and purification were conducted as previously described (*6*). For ^{15}N, ^{15}N and ^{2}D, or ^{15}N and ^{13}C labeling, bacteria were grown on M9 minimal medium supplemented with 1 g/L $^{15}NH_4Cl$, 1 g/L $^{15}NH_4Cl$ and D_2O, or 1 g/L $^{15}NH_4Cl$ and 1 g/L [^{13}C]glucose, respectively.

DNA. Two oligonucleotide fragments (TTG CGG ATA CAA and TTG TAT CCG CAA), derived from half of the imperfect inverted repeat of the promotor, were obtained from biomers.net GmbH, dissolved and mixed in NMR buffer, heated to 95 and 65 °C for 2 min each, and cooled slowly to room temperature.

NMR Spectroscopy. All spectra were recorded on Bruker AMX 400, DRX 500, DRX 600, AvanceII 600, and Avance 700 spectrometers, using 50 mM sodium phosphate buffer (pH 5.0) containing 10% (v/v) D_2O at 25 °C. Backbone assignment was achieved with a ^{15}N and ^{13}C sample by a series of HNCO, HNCA, HNCACB, and CBCA(CO)NH spectra. Side chain assignment was achieved with an H(C)CH-COSY experiment and a ^{13}C-edited $^{1}H-^{13}C$ NOESY-HSQC experiment. Backbone and side chain assignment was assisted by ^{15}N-edited $^{1}H-^{15}N$ NOESY-HSQC and $^{1}H-^{15}N$ TOCSY-HSQC (MLEV, 60 ms mixing time) experiments. NOEs were derived with a $^{1}H-^{1}H$ NOESY spectrum (100 ms mixing time) and a ^{13}C-edited $^{1}H-^{13}C$ NOESY-HSQC spectrum (80 ms mixing time).

Intersubunit NOEs were derived by double-half-filter experiments (*13*) with a sample of mixed proteins (^{15}N and ^{13}C and ^{14}N and ^{12}C) in the NH−NH and CH−CH cases as well as by comparison of a two-dimensional (2D) NOESY spectrum and a ^{15}N-edited 2D NOESY spectrum of a ^{15}N- and ^{2}D-labeled sample and a ^{14}N- and ^{1}H-labeled sample, in the case of NH−CH. Samples were prepared by mixing equal amounts of unfolded protein.

H−D exchange experiments were performed on a Bruker AvanceII 600 spectrometer at pD 7.5. DNA titration was followed by a series of ^{15}N HSQC spectra at different titration points. The influence of binding was analyzed using the weighted average chemical shift difference (*14*). ^{15}N heteronuclear NOE experiments with the free and DNA bound forms of ORF56 were performed by recording two data sets without and with ^{1}H saturation for 3 s by successive 120° pulses. All data were processed and analyzed using Felix 2000.

NMR Structure Calculations. Structures were calculated with ARIA2.0 (*15*) using the ARIA standard parameters but by increasing the number of cooling steps to 200000 and reducing the windows for NOE assignment to 0.003 ppm (direct dimension) and 0.007 ppm (indirect dimension). All NOEs were treated ambiguously as inter- or intramolecular NOEs at the start of the structure calculation, caused by the homodimeric nature of ORF56. The structure calculation was guided by 20 experimentally identified intermolecular NOEs.

Besides the NOE data, H bond restrictions from H−D exchange experiments, ϕ angle restrictions derived from J couplings, and backbone conformation (ϕ and ψ angles) restrictions derived from the chemical shift index were used. The stereochemistry was analyzed with PROCHECK (*16*).

Crystallization, Data Collection, and Processing. Crystals of ORF56 were obtained using the hanging-drop vapor-diffusion method at 13 °C. Two microliters of a 20 mg/mL protein solution was diluted 1:1 from a reservoir containing 100 mM HEPES and 20% PEG 4000 (pH 7.5). Crystals ~0.2 mm in size grew in 30 days. The crystals of ORF were first soaked in suitable cryoprotectant solution (reservoir with 20% glycerol) and mounted on a loop in a cold nitrogen stream. Data were collected to a resolution of 2.0 Å at the BESSY-MX BL14.1 beamline (Berlin, Germany). Indexing and scaling were performed with MOSFLM (*17*) and SCALA (*18*). The data collection statistics are summarized in Table 2.

Structure Solution and Data Analysis. Initial indexing suggested space group $P4_32_12$, with two monomers in the asymmetric unit and the following unit cell dimensions: $a = b = 46.56$ Å, and $c = 85.84$ Å. Molecular replacement (MR) using PHASER (*19*) with a search model prepared from the ARC repressor structure (PDB entry 1PAR) allowed placement of a dimer in this crystal form and manual rebuilding of the ORF56 sequence into the electron density using Coot (*20*). The structure refined poorly however, with R_{cryst} and R_{free} values of 45 and 51%, respectively. Careful inspection and reindexing of the diffraction data suggested the lower symmetry space group $P4_3$ with the following unit cell dimensions: $a = b = 33.03$ Å ($= 46.56/\sqrt{2}$ Å), and $c = 85.69$ Å. Analysis of the data with phenix.xtriage of the PHENIX package (*19*) reported a strong merohedral twinning of the ORF crystal, with the twin operator $(h,-k,-l)$ and a twinning fraction of ~0.478. Despite the twinning problem, the structure could be determined by MR using the partially refined solution from the pseudo-space group; attempts to use the original ARC repressor structure failed. The asymmetric unit of the crystal again comprises one ORF56 dimer; the twinning axis is parallel to both the dimer axis and the crystallographic a axis. The structure was manually rebuilt and verified against omit difference Fourier maps, which despite the twinning were clearly interpretable (SI Figure 6). The final model consists of residues 7−51 in chain A and 6−51 in chain B; the missing amino acids resulted from the noninterpretable electron density map and are most likely disordered. The model was refined to R and R_{free} values of 17.06 and 23.70%, respectively, using phenix.refine of the PHENIX package with a refined twin fraction at the end of refinement of 0.481. The refined model has good geometry as judged by PROCHECK (*21*); there are no residues in disallowed regions of the Ramachandran plot (87.5% in most allowed regions and 11.2% in allowed regions).

GdmCl Transitions. GdmCl-induced equilibrium unfolding was followed by fluorescence spectroscopy using 2 μM ORF56 at 25 °C in 50 mM Tris-HCl (pH 7.5). Data were analyzed assuming a transition of a folded dimer to two unfolded monomers without intermediates as described previously (*7*).

Calculation of Dimer Interfaces and Changes in Accessible Surface Areas during Unfolding. Size and polar/hydrophobic contributions of the dimer interfaces and the change in accessible surface area during unfolding were calculated using STC (*22*), GETAREA (*23*), and NACCESS (Hubbard and Thornton, 1993) using PDB entries 1ARR, 1MYK, 1B28, 2CPG, 3FT7, and 2K9I. Standard parameters were used for calculations of the dimer interface with STC. Using GETAREA and NACCESS, the dimer interface was calculated by subtracting the surface of the dimer from the surface of the two theoretical monomers of the structures. The change in ASA was calculated using NACCESS by subtracting the surface of the dimer from the

surface of the two monomers in an all-β conformation as a model for the unfolded state according to ref *24*.

RESULTS

Determination of the Structure of ORF56. Standard three-dimensional (3D) NMR spectra [HNCO, HNCA, HNCACB, CBCA(CO)NH, H(C)CH-COSY, and ^{13}C and ^{15}N NOESY-HSQC] were analyzed for the complete assignment of [^{13}C,^{15}N]ORF56. All NMR spectra (e.g., ^{15}N HSQC spectrum in Figure 1) showed a single set of resonances for the homodimeric protein. Intermolecular NOEs were obtained by ^{15}N and ^{13}C double-half-filter experiments (*13*). All NOEs from ^{13}C and ^{15}N NOESY experiments were treated as ambiguous distance restraints (intra- and intermolecular) for structure calculations using ARIA 2.0 (*15*). We obtained a converging ensemble of structures only after manually restraining 20 interdomain distances unambiguously obtained from the double-half-filter experiments, however. Experimental restraints and structure statistics are summarized in Table 1.

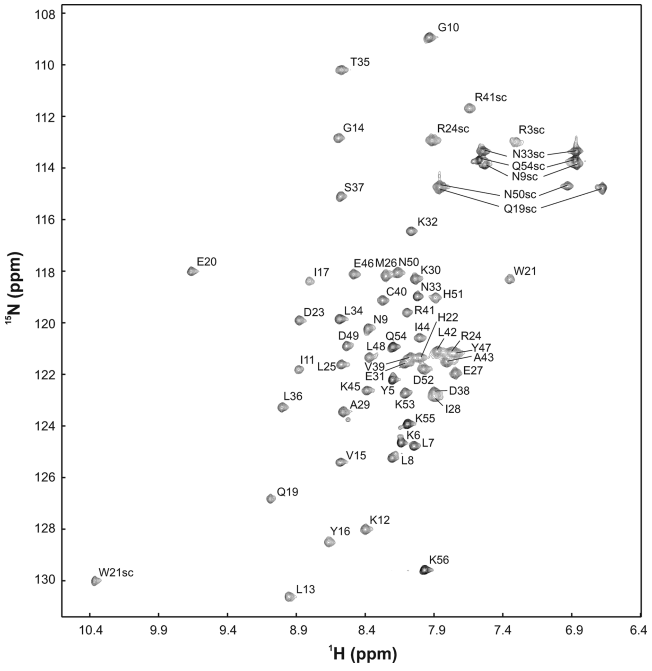

FIGURE 1: 2D ^{15}N HSQC spectrum of ORF56 at 25 °C and pH 5.0. Cross-peaks of the backbone amides are labeled by the one-letter amino acid code. Cross-peaks of amino acid side chains are marked with sc. A single set of resonances has been detected for the dimeric protein.

Crystals were obtained in 100 mM HEPES and 20% PEG 400 (pH 7.5) and measured at the BESSY-MX BL14.1 beamline to a resolution of 2.0 Å. Molecular replacement was performed using a monomer of the Arc repressor. The data collection statistics are summarized in Table 2.

The ensemble of the 20 lowest-energy water-refined structures is well-defined for residues 10−50 (Figure 2A), and the crystal structure shows electron density from residue 8 to 52 (Figure 2C). Both structures consist of an intermolecular β-sheet (residue 11−18), followed by two helices (helix 1: residue 19−32, helix 2: residue 36−50) exhibiting the common ribbon-helix-helix DNA binding motif (Figure 2B,C). Clearly, ORF56 adopts the same global fold as other rhh-proteins including the Arc repressor of salmonella bacteriophage P22, the plasmid-encoded transcription repressor CopG (RCSB protein data bank entries 1ARR.pdb (*25*), 1MYK.pdb (*8*), 1B28.pdb (*12*), 2CPG.pdb (*11*)), methionine repressor MetJ (*26*) and the antitoxin CcdA (*27*).

Functional Residues of ORF56 Localized by NMR Titration with DNA. ORF56 binds to its own promoter on plasmid pRN1 (*6*) as a tetramer. To locate the binding interface on the protein, varying amounts of double-stranded DNA fragments comprising half (to avoid tetramerization) of the imperfect reverted repeat (12 bp) of the promotor were titrated into ^{15}N-labeled ORF56. Residues with a change in the weighted averaged backbone chemical shift above 0.1 ppm, as monitored by ^{15}N HSQC spectra, are localized in the β-sheet and the N-terminal parts of the helices of ORF56 (colored red in Figure 3A) and allowed us to map the functional region of the protein. The identified binding interface contains all residues whose homologues show direct protein DNA contacts in the DNA complex structures of the Arc repressor (*28*) and CopG (*11*): K12, L13, G14, Y16, L36, and S37 (yellow in Figure 2C). Several other residues were identified in the direct structural neighborhood: I11, V15, I17, Q19, D23, C40, and R41 (orange in Figure 2C). Among the first set of residues, the position but not the type of amino acid is conserved, indicating a high variability for binding to specific DNA sequences. In addition to the local changes in the chemical environment of ORF56 upon dsDNA binding, the protein backbone dynamics change on a pico- to nanosecond time scale. The unstructured N-terminus becomes less flexible in the complex as indicated by

Table 1: Experimental Restraints and Structure Statistics for NMR Structure Calculation

no. of restraints		rmsd	
NOEs	1265 × 2	rmsd for bonds	0.00205 ± 0.00007 Å
dihedrals	82 × 2	rmsd for NOEs	0.0323 ± 0.0076 Å
J couplings	46 × 2	rmsd for angles	0.374 ± 0.006°
H bonds	27 × 2	rmsd for impropers	0.298 ± 0.014°
no. of NOE violations		rmsd for backbone atoms	2.31 ± 0.74 Å
NOE of >0.5	0.10 ± 0.30	rmsd for heavy atoms	2.81 ± 0.70 Å
NOE of >0.3	0.35 ± 0.48	rmsd for secondary structure backbone atoms	0.24 ± 0.06 Å
NOE of >0.1	34.80 ± 3.22	rmsd for secondary structure heavy atoms	0.69 ± 0.07 Å
energy		Ramachandran plot (residues 11−50)	
E_{tot}	384.2 ± 17.3	most allowed region	97.9%
E_{bond}	26.2 ± 1.6	allowed region	2.1%
E_{angle}	151.8 ± 10.5	generously allowed region	0%
$E_{improper}$	77.1 ± 8.7	disallowed region	0%
E_{NOE}	73.1 ± 8.4		
E_{cdih}	5.7 ± 1.3		
E_{coup}	50.3 ± 6.3		

an increased ^{15}N heteronuclear NOE (Figure 3C), indicating that the unstructured N-terminus is somehow involved in or at least affected by DNA binding.

Local Fluctuations of ORF56 Followed by H−D Exchange. The extremely high thermodynamic stability of ORF56 is reflected by a Gibbs free energy of stabilization of −20.3 kcal/mol and an extrapolated melting point of 107.5 °C (*7*). To find structure-based reasons for this high stability of ORF56, we analyzed local amide proton exchange of the protein on a residue-by-residue basis. During the experiment (2 days) with lyophilized protein dissolved in D_2O buffer, exchange was followed by a series of ^{15}N HSQC spectra. Plotting the cross-peak intensity against time reveals four subsets of exchange behavior. Amide protons that completely exchange during the dead time of the experiment (red in Figure 3B) are localized in the unstructured termini, the N-termini of each helix (DNA binding region), and residues in β-sheet that are not hydrogen bonded. Amide protons of the antiparallel β-sheet involved in H bonds show increased levels of protection but still exchanged within the first 2 h of the experiment (violet in Figure 3B). This exchange rate is much higher compared to the expected rate of exchange for residues reflecting only complete unfolding for achieving the exchange competent state, which corresponds to the denaturant-induced unfolding transitions monitored by fluorescence or CD spectroscopy (*7*). Amides that show a slow exchange in the range of hours (cyan in Figure 3B) or no exchange at all (blue in Figure 3B) cluster at the C-terminal part of the first helix and the core of the second helix. The most protected residues, localized in the second helices, mainly form the intermolecular hydrophobic core (depicted side chains of W21, L25, I28, L34, V39, and L42 of one monomer and A43, I44, Y47, L48, and H51 of the other monomer in Figure 3B).

These experiments indicate that the intermolecular hydrophobic core must be the origin of the thermostability. It is well separated from the DNA binding site, which exhibits limited local structural fluctuations. Such a large difference in the level of protection against exchange of H bonds in structured parts of the molecule (second helix vs β-sheet) is exceptional when compared to other single-domain model proteins for protein folding such as ubiquitine (*29*), RNase T1 (*30*), or α-spectrin SH3 (*31*, *32*). However, such differences have often been related to functional aspects, in particular, DNA binding proteins (*12*, *33*−*37*).

Comparison of Different rhh Structures Suggests Aspects That Could Influence Stability. A comprehensive understanding of the thermodynamic stability of a protein solely based on its structure is still difficult. While some aspects influencing protein stability (e.g., hydrogen bonds, salt bridges, unfavorable electrostatic contacts, and hydrophobic interactions) have been impacted in individual structures, no general rules have emerged despite the available complete genome sequences and structural protein databases. Apparently, individual strategies have evolved for each protein to increase thermostability without losing function (*38*, *39*).

The inter- and intramolecular contacts (hydrogen bonds, salt bridges, and hydrophobic contacts) of ORF56 were compared with those of much less stable but structurally homologous rhh proteins: wild type (wt), P8L, and R31M/E36Y/R40L Arc

Table 2: X-ray Data Collection and Refinement Statistics

Data Collection	
wavelength (Å)	λ = 0.91841
space group	$P4_3$
cell (Å)	a = 33.03
maximal resolution (Å)	2.00
R_{merge} [a]	8.00 (62.3) [b]
% complete	99.6 (99.7) [b]
no. of reflections	26364
no. of unique reflections	6218
$\langle I/\sigma(I)\rangle$	11.5 (2.8) [b]
Refinement	
resolution range (Å)	33.03−2.00 (2.12−2.00) [b]
completeness (working + test) (%)	96.2 (98.1) [b]
no. of reflections ($F > 0$)	6218 (163) [b]
Wilson B (Å^2)	26.3
R_{cryst} [c] (%)	17.06 (28.19) [b]
R_{free} [d] (%)	23.70 (34.37) [b]
no. of non-hydrogen atoms	791
protein	744
water	41
glycerol	6
Data Collection	
rmsd from ideality	
bond lengths (Å)	0.016
bond angles (deg)	1.933
dihedral angles (deg)	21.949
chirality (deg)	0.103
average B-factor (Å^2)	
protein atoms	49.52
main chain atoms	47.98
waters	40.46
glycerols	47.93
twinning information	
fraction	0.481
operator	$h,-k,-l$

[a] $R_{merge} = |I - \langle I\rangle|/\langle I\rangle$. [b] Values in parentheses correspond to the highest-resolution shell (2.12−2.00 Å). [c] $R_{cryst} = ||F_{obs}| - |F_{calc}||/|F_{obs}|$. [d] R_{free} is calculated as R_{cryst} for a test set comprising 9.1% of the reflections (the free reflections were chosen to obey the highest possible symmetry of the lattice in phenix.refine) not used in the refinement.

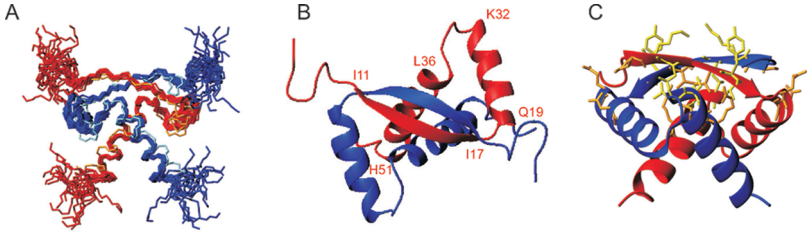

FIGURE 2: Solution and crystal structure of ORF56. (A) Backbone superposition of the 20 lowest-energy NMR structures (PDB entry 2k9i, dark blue and red) and the crystal structure (PDB entry 3ft7, light blue and orange) with an rmsd of 1.82 Å for backbone heavy atoms between the averaged NMR structure and the crystal structure. (B) Ribbon representation of the energy-minimized average NMR structure, rotated horizontally by 90° relative to panel A. (C) Crystal structure of ORF56, where side chains involved in DNA binding are colored yellow and orange. The two respective chains of the homodimer are colored red and blue.

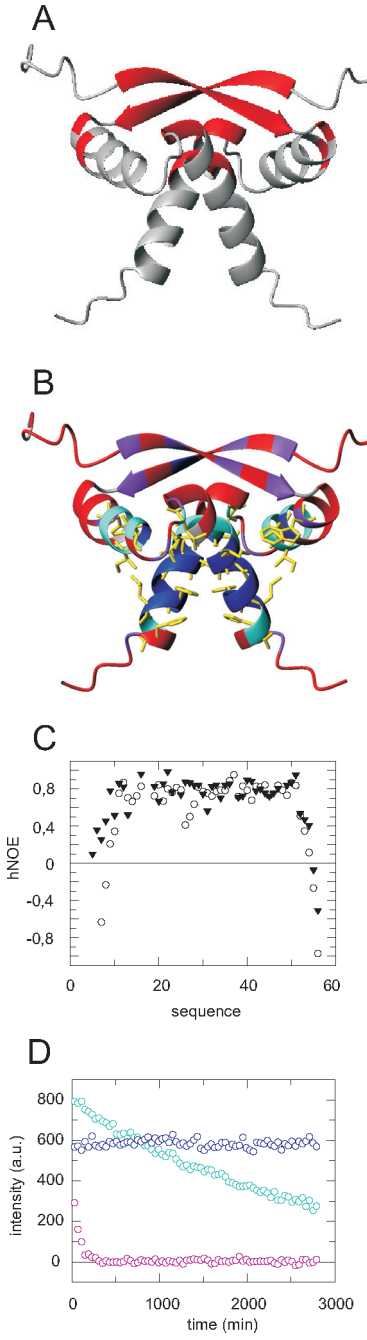

FIGURE 3: (A) Ribbon representation of the energy-minimized average structure of ORF56. Residues showing a shift of their amide cross-peak of >0.1 ppm (weighted average chemical shift) upon addition of a dsDNA fragment are colored red. (B) Residues are colored according to different amide exchange behaviors: red for exchange in dead time, magenta for exchange in the first hours, cyan for some exchange, and blue for no exchange over 2 days. Side chains involved in the hydrophobic cluster are colored yellow. (C) hNOEs of ORF56 free (○) and bound to DNA (▼). (D) Examples of residues exchanging with D_2O in different time regimes: dark blue for L48, light blue for A29, and pink for I17.

repressor as well as CopG (PDB entries 1ARR, 1MYK, 1B28, and 2CPG, respectively). It should be borne in mind that this approach neglects the unfolded state and surface electrostatics (*40*), which could also influence stability.

The H bond profile of backbone H bonds was characterized with DSSP (*41*), and LIGPLOT/HBPLUS (*42*, *43*) was employed for all H bonds across the dimer interface. Additionally, all H bonds as well as potentially missing H bond partners (donors or acceptors) were analyzed manually. A comparison of the Arc repressor variants and CopG revealed no significant differences between the examined structures, except one additional H bond donor in the β-sheet of ORF56 (I11) and CopG (K3) compared to the wt Arc repressor, which hosts P8 at the respective position. No additional stabilizing electrostatic interactions were obvious in ORF56, although the Arc repressor contains an unfavorable intramolecular interaction between R31 and R40 that is missing in CopG and ORF56.

Analysis of the dimerization surface of the rhh proteins using GETAREA (*23*), STC (*22*), and NACCESS (Hubbard and Thornton, 1993) reveals an almost invariant size of all the dimerization interfaces and therefore no influence on stability (Table 3 and SI Figure 7). The same holds for the change in accessible surface area upon unfolding (ASA) of all rhh dimers. However, the hydrophobic fraction of the change in ASA and the even higher hydrophobic fraction of the interface of the rhh proteins approximately increase with the stability of the dimer (ΔG_U values in Table 3 and SI Figure 8). Because of a comparable midpoint of the denaturant-induced unfolding transitions of all analyzed rrh dimers, the *m* values of the different homodimers also increase with the hydrophobic fraction of the ASA change and of the interface.

Analysis of Protein Variants. Two residues of ORF56 were substituted with the respective amino acid of the Arc repressor to investigate the role of different interactions in terms of stability (Figure 4). The thermodynamic stability was measured by fluorescence-detected equilibrium unfolding transitions of wild-type ORF56 and the derived variants using GdmCl as a denaturant (Figure 5 and Table 3). The I11P mutation was chosen because in the Arc repressor exchange of the corresponding Pro8 with any other residue increases stability by providing two additional backbone H bonds in the intermolecular β-sheet (*8*). Interestingly, the β-sheet of ORF56 itself shows an increased number of local fluctuations according to the H−D exchange results. The second mutation, W21V, was selected to test the role of the side chain of W21, which is buried in the core of the protein and is the amino acid that forms most of the inter- and intramolecular hydrophobic contacts. Both ORF56 mutants reveal a significant decrease in stability (Figure 5), showing that disturbance of the intra- and intermolecular interactions decreases stability. This suggests that hydrogen bonding of the first residue in the β-sheet and the core of ORF56 in terms of side chain packing of hydrophobic residues are optimized for high stability.

DISCUSSION

Structure and Global Fold. ORF56 shows the ribbon−helix−helix fold, consisting of an intermolecular antiparallel β-sheet and two intertwined helices per monomer. The rhh fold is a canonical DNA binding motif of small homodimeric proteins that function as repressors, e.g., the Arc repressor (*44*), CopG (*11*), and MetJ (*26*). The rmsd of heavy backbone atoms between ORF56 and CopG or the Arc repressor is ~2.0 Å. While the structure of ORF56 is well-defined from residue 10 to 50 indicated by a high degree of convergence in the NMR structure ensemble as well as electron density in the crystal structure, the positively charged termini of ORF56 combining R3 and K6 in the N-terminus as well as K53, K55, and K56 in the C-terminus are unstructured. While the length of the C-terminus is similar to that of the Arc repressor, the unstructured N-terminus is elongated. The N-terminus is not directly involved in DNA binding

Table 3: ΔG_U and *m* Values of Different rhh Proteins Taken from the Literature and Calculated Dimer Interfaces Using NACCESS, GETAREA, and STC

	ARC	ARC P8L	ARC MYL	CopG	ORF56-X-ray	ORF56-NMR
ΔG_U (kcal/mol_	9.52−9.82	12.37−12.5	14.28	13.42	20.31[a]	20.31[a]
m value (kcal mol^{-1} M^{-1})	1.31−1.39	1.20−1.27		1.92	2.27	2.27
ref	*45*	*8*	*52*	*46*	*7*	*7*
NACCESS						
ASA[b] (Å^2)	9230	8366	8136	6912	7422	7635
% hydrophobic ASA[b]	55	59	62	61	63	64
interface (Å^2)	3917	3569	3376	3167	3339	3059
% hydrophobic interface	67	67	75	75	73	80
GETAREA						
interface (Å^2)	3928	3554	3376	3167	3507	3136
% hydrophobic interface	71	71	78	78	76	81
STC						
interface (Å^2)	3926	3457	3511	3264	3512	3108
% hydrophobic interface	68	70	75	73	74	78
PDB entry	1ARR	1MYK	1B28	2CPG	3FT7	2K9I

[a]This ΔG_U taken from ref *7* differs from the value of 17.82 kcal/mol determined in this study because of a different buffer and a different protein concentration. [b]Change in accessible surface area upon unfolding.

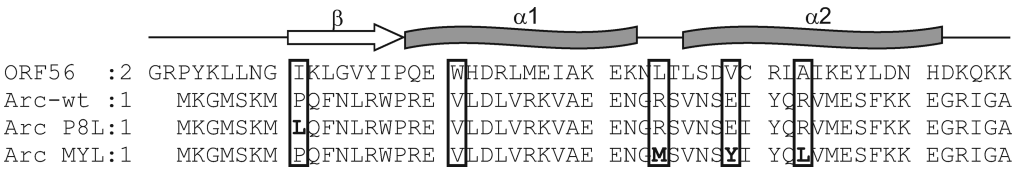

FIGURE 4: Sequence alignment of ORF with the wild-type Arc repressor and its stabilizing mutants P8L and R31M/E36Y/R40L (MYL). The positions of mutations are marked. At the top, the secondary structure elements are indicated.

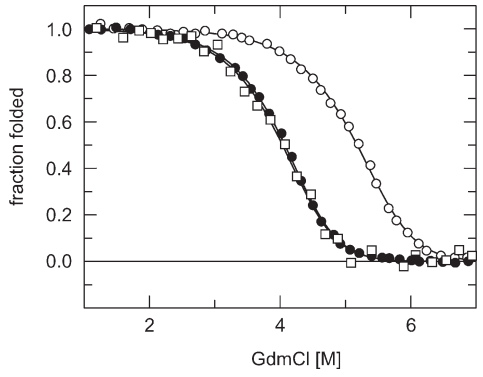

FIGURE 5: GdmCl-induced unfolding of wild-type ORF56 and two variants. All transitions were measured at a protein concentration of 2 μM at 25 °C and pH 7.5. Analysis of (○) wild-type ORF56 leads to a ΔG_U of 17.82 kcal/mol and an *m* value of 1.95 kcal mol^{-1} M^{-1}. The corresponding stabilities and *m* values for (●) I11P (16.26 kcal/mol and 2.09 kcal mol^{-1} M^{-1}, respectively) and (□) W21V (15.65 kcal/mol and 2.07 kcal mol^{-1} M^{-1}, respectively) are significantly lower.

according to the NMR titration experiment but becomes less flexible on a pico- to nanosecond time scale (hNOE) in the DNA complex.

Stability and Structure. For all rrh proteins discussed herein, the same unfolding mechanism at equilibrium has been reported: 2U ⇆ N_2 (*7*, *45*, *46*). For such a two-state mechanism, the experimentally determined *m* value from unfolding transitions shows a good correlation with the change in the accessible surface area upon unfolding (*24*). On the basis of the 3D structure of the Arc repressor, CopG, and ORF56, we calculated the change in ASA (Table 3). This parameter cannot explain the large difference in ΔG_U of ~10 kcal/mol and in the *m* value of ~1 kcal mol^{-1} M^{-1}, which is rather invariant with respect to these thermodynamic properties (SI Figure 7). The surface area buried by the dimer interface is also invariant within the analyzed rrh proteins. A closer look at these buried areas revealed that the hydrophobic fraction shows some dependence on the thermodynamic stability of the rrh proteins (SI Figure 8). For both the change in ASA and the dimer interface, the hydrophobic fraction of buried surface increases with protein stability in a manner independent of the program used to calculate these areas.

This result has to be discussed in the context of several assumptions and simplifications. First, we assumed that all investigated rrh proteins follow a two-state unfolding model. At least one report for the Arc repressor challenges this assumption from H−D exchange data in water (*47*). The presence of an intermediate generally decreases the experimentally determined *m* value relative to the ASA change between the native and fully unfolded state (*48*). For the comparison of the different stabilities of rrh dimers, this would mean that the experimentally determined *m* value of the Arc repressor is lower than expected for a two-state mechanism. Second, the unfolded state of the analyzed rrh dimers might be different, which also has a prominent influence on protein stability (*49*).

From the experiments presented here and despite simplifying assumptions, we can conclude only that the hydrophobic fractions of changes in ASA and of the buried dimer interface according to the known 3D structures increase with protein stability. This contribution can only partially explain the large difference in ΔG_U and the *m* value between the mesophilic and thermophilic representatives of rrh proteins. The reduced stability of the W21V variant is in line with this conclusion. Note that we can exclude higher oligomeric states for ORF56 from ultracentrifugation experiments (*7*).

Stability and Function at a Residue-by-Residue Level. The residues of ORF56 localized in the β-sheet experience a large difference in amide proton exchange rates with the solvent

compared to its α-helices and especially its core residues of the second helix. Such large differences are generally observed for proteins (e.g., between secondary structures and exposed loop regions), and different mechanisms have been suggested for this observation, including partial unfolding, limited structural fluctuations, and global unlocking (*50*, *51*). For a clear discrimination between these models and for a quantitative analysis, H−D exchange data at different denaturant concentrations are required. For ORF56, we determined amide proton exchange protection only in water but suggest that limited structural fluctuations are the source of the low level of protection in the β-sheet. The reason is an earlier measured equilibrium GdmSCN transition monitored by NMR, which shows two-state behavior for residues from all sites of ORF56 with very close ΔG_U and m values (*7*). The two-state mechanism was confirmed in the same study by kinetic un- and refolding experiments, which excludes kinetic intermediates that can be detected by fluorescence or CD spectroscopy. Therefore, we exclude partial unfolding or global unlocking as a mechanism for less protected amide protons as a first approximation.

Beside these local fluctuations, the steric freedom and the number of H bonds in the antiparallel β-sheet are required for high stability, because the I11P substitution strongly destabilizes ORF56. The fluctuations and dynamics in the β-sheet are expected to be important for the function of ORF56, which is binding to dsDNA: enhanced flexibility allows improved DNA interaction by adjusting the local geometry in the complex. The same observation has been reported for the Arc repressor (*12*). The three substitutions (R31M/E36Y/R40L) increased the stability of the protein, but the local dynamics of the β-sheet remained. For the *lac* repressor, a more detailed picture arose from H−D exchange and dynamics studies (*33*, *34*). For the fast location of the target site, mainly electrostatic but nonspecific interactions with the dsDNA are responsible. This nonspecific complex retains its dynamics and low level of protection of amide protons against exchange. Both properties are missing after formation of a specific complex.

These results show that ORF56 follows these same general rules. The very high global stability of the protein can in part be explained by an increased fraction of hydrophobic interaction and the maximum number of possible H bonds in the interdomain β-sheet. Thus, ORF56 has an optimized thermodynamic stability in hyperthermophilic *S. islandicus*. The flexibility in the β-sheet is indispensably linked to the DNA binding function in evolutionarily very distant Arc repressor and ORF56.

SUPPORTING INFORMATION AVAILABLE

The $2F_o - F_c$ omit electron density maps and a graphical representation of Table 3. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Peng, X., Holz, I., Zillig, W., Garrett, R. A., and She, Q. (2000) Evolution of the family of pRN plasmids and their integrase-mediated insertion into the chromosome of the crenarchaeon *Sulfolobus solfataricus*. *J. Mol. Biol. 303*, 449–454.
2. Lipps, G. (2006) Plasmids and viruses of the thermoacidophilic crenarchaeote Sulfolobus. *Extremophiles 10*, 17–28.
3. Lipps, G., Ibanez, P., Stroessenreuther, T., Hekimian, K., and Krauss, G. (2001) The protein ORF80 from the acidophilic and thermophilic archaeon *Sulfolobus islandicus* binds highly site-specifically to double-stranded DNA and represents a novel type of basic leucine zipper protein. *Nucleic Acids Res. 29*, 4973–4982.
4. Lipps, G., Rother, S., Hart, C., and Krauss, G. (2003) A novel type of replicative enzyme harbouring ATPase, primase and DNA polymerase activity. *EMBO J. 22*, 2516–2525.
5. Lipps, G., Weinzierl, A. O., von Scheven, G., Buchen, C., and Cramer, P. (2004) Structure of a bifunctional DNA primase-polymerase. *Nat. Struct. Mol. Biol. 11*, 157–162.
6. Lipps, G., Stegert, M., and Krauss, G. (2001) Thermostable and site-specific DNA binding of the gene product ORF56 from the *Sulfolobus islandicus* plasmid pRN1, a putative archael plasmid copy control protein. *Nucleic Acids Res. 29*, 904–913.
7. Zeeb, M., Lipps, G., Lilie, H., and Balbach, J. (2004) Folding and association of an extremely stable dimeric protein from *Sulfolobus islandicus*. *J. Mol. Biol. 336*, 227–240.
8. Schildbach, J. F., Milla, M. E., Jeffrey, P. D., Raumann, B. E., and Sauer, R. T. (1995) Crystal structure, folding, and operator binding of the hyperstable are repressor mutant PL8. *Biochemistry 34*, 1405–1412.
9. Srivastava, A. K., and Sauer, R. T. (2002) Mutational studies of protein stability and folding of the hyperstable MYL Arc repressor variant. *Biophys. Chem. 101−102*, 35–42.
10. Breg, J. N., van Opheusden, J. H., Burgering, M. J., Boelens, R., and Kaptein, R. (1990) Structure of Arc repressor in solution: Evidence for a family of β-sheet DNA-binding proteins. *Nature 346*, 586–589.
11. Gomis-Rüth, F. X., Sola, M., Acebo, P., Parraga, A., Guasch, A., Eritja, R., Gonzalez, A., Espinosa, M., del Solar, G., and Coll, M. (1998) The structure of plasmid-encoded transcriptional repressor CopG unliganded and bound to its operator. *EMBO J. 17*, 7404–7415.
12. Nooren, I. M., Rietveld, A. W., Melacini, G., Sauer, R. T., Kaptein, R., and Boelens, R. (1999) The solution structure and dynamics of an Arc repressor mutant reveal premelting conformational changes related to DNA binding. *Biochemistry 38*, 6035–6042.
13. Folkers, P. J. M., Folmer, R. H. A., Konings, R. N. H., and Hilbers, C. W. (1993) Overcoming the Ambiguity Problem Encountered in the Analysis of Nuclear Overhauser Magnetic-Resonance Spectra of Symmetrical Dimer Proteins. *J. Am. Chem. Soc. 115*, 3798–3799.
14. Grzesiek, S., Stahl, S. J., Wingfield, P. T., and Bax, A. (1996) The CD4 determinant for downregulation by HIV-1 Nef directly binds to Nef. Mapping of the Nef binding surface by NMR. *Biochemistry 35*, 10256–10261.
15. Linge, J. P., Habeck, M., Rieping, W., and Nilges, M. (2003) ARIA: Automated NOE assignment and NMR structure calculation. *Bioinformatics 19*, 315–316.
16. Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR 8*, 477–486.
17. Leslie, A. G. W. (1992) Recent changes to the MOSFLM package for processing film and image plate data. Joint CCP4 and ESF-EACMB Newsletter on Protein Crystallography Vol. 26, BBSRC, Warrington, U.K.
18. Evans, P. R. (1997) Scaling of MAD Data. Proceedings of the CCP4 Study Weekend on Recent Advances in Phasing, ESRF, Grenoble, France.
19. Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K., and Terwilliger, T. C. (2002) PHENIX: Building new software for automated crystallographic structure determination. *Acta Crystallogr. D58*, 1948–1954.
20. Emsley, P., and Cowtan, K. (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D60*, 2126–2132.
21. Laskowski, R. A., Macarthur, M. W., Moss, D. S., and Thornton, J. M. (1993) Procheck: A Program to Check the Stereochemical Quality of Protein Structures. *J. Appl. Crystallogr. 26*, 283–291.
22. Lavigne, P., Bagu, J. R., Boyko, R., Willard, L., Holmes, C. F., and Sykes, B. D. (2000) Structure-based thermodynamic analysis of the dissociation of protein phosphatase-1 catalytic subunit and microcystin-LR docked complexes. *Protein Sci. 9*, 252–264.
23. Fraczkiewicz, R., and Braun, W. (1998) Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. *J. Comput. Chem. 19*, 319–333.
24. Myers, J. K., Pace, C. N., and Scholtz, J. M. (1995) Denaturant m values and heat capacity changes: Relation to changes in accessible surface areas of protein unfolding. *Protein Sci. 4*, 2138–2148.
25. Bonvin, A. M., Vis, H., Breg, J. N., Burgering, M. J., Boelens, R., and Kaptein, R. (1994) Nuclear magnetic resonance solution structure of the Arc repressor using relaxation matrix calculations. *J. Mol. Biol. 236*, 328–341.

26. Old, I. G., Phillips, S. E., Stockley, P. G., and Saint Girons, I. (1991) Regulation of methionine biosynthesis in the Enterobacteriaceae. *Prog. Biophys. Mol. Biol. 56*, 145–185.
27. Madl, T., Van Melderen, L., Mine, N., Respondek, M., Oberer, M., Keller, W., Khatai, L., and Zangger, K. (2006) Structural basis for nucleic acid and toxin recognition of the bacterial antitoxin CcdA. *J. Mol. Biol. 364*, 170–185.
28. Raumann, B. E., Rould, M. A., Pabo, C. O., and Sauer, R. T. (1994) DNA recognition by β-sheets in the Arc repressor-operator crystal structure. *Nature 367*, 754–757.
29. Brand, T., Cabrita, E. J., Morris, G. A., Gunther, R., Hofmann, H. J., and Berger, S. (2007) Residue-specific NH exchange rates studied by NMR diffusion experiments. *J. Magn. Reson. 187*, 97–104.
30. Steegborn, C., Schneider-Hassloff, H., Zeeb, M., and Balbach, J. (2000) Cooperativity of a Protein Folding Reaction Probed at Multiple Chain Positions by Real-Time 2D NMR Spectroscopy. *Biochemistry 39*, 7910–7919.
31. Casares, S., Lopez-Mayorga, O., Vega, M. C., Camara-Artigas, A., and Conejero-Lara, F. (2007) Cooperative propagation of local stability changes from low-stability and high-stability regions in a SH3 domain. *Proteins 67*, 531–547.
32. Sadqi, M., Casares, S., Abril, M. A., Lopez-Mayorga, O., Conejero-Lara, F., and Freire, E. (1999) The native state conformational ensemble of the SH3 domain from α-spectrin. *Biochemistry 38*, 8899–8906.
33. Kalodimos, C. G., Boelens, R., and Kaptein, R. (2002) A residue-specific view of the association and dissociation pathway in protein−DNA recognition. *Nat. Struct. Biol. 9*, 193–197.
34. Kalodimos, C. G., Biris, N., Bonvin, A. M., Levandoski, M. M., Guennuegues, M., Boelens, R., and Kaptein, R. (2004) Structure and flexibility adaptation in nonspecific and specific protein-DNA complexes. *Science 305*, 386–389.
35. Zeeb, M., Max, K. E., Weininger, U., Löw, C., Sticht, H., and Balbach, J. (2006) Recognition of T-rich single-stranded DNA by the cold shock protein Bs-CspB in solution. *Nucleic Acids Res. 34*, 4561–4571.
36. Löw, C., Homeyer, N., Weininger, U., Sticht, H., and Balbach, J. (2009) Conformational switch upon phosphorylation: Human CDK inhibitor p19INK4d between the native and partially folded state. *ACS Chem. Biol. 4*, 53–63.
37. Löw, C., Weininger, U., Zeeb, M., Zhang, W., Laue, E. D., Schmid, F. X., and Balbach, J. (2007) Folding mechanism of an ankyrin repeat protein: Scaffold and active site formation of human CDK inhibitor p19(INK4d). *J. Mol. Biol. 373*, 219–231.
38. Jaenicke, R. (2000) Stability and stabilization of globular proteins in solution. *J. Biotechnol. 79*, 193–203.
39. Shoichet, B. K., Baase, W. A., Kuroki, R., and Matthews, B. W. (1995) A relationship between protein stability and protein function. *Proc. Natl. Acad. Sci. U.S.A. 92*, 452–456.
40. Perl, D., Mueller, U., Heinemann, U., and Schmid, F. X. (2000) Two exposed amino acid residues confer thermostability on a cold shock protein. *Nat. Struct. Biol. 7*, 380–383.
41. Kabsch, W., and Sander, C. (1983) Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers 22*, 2577–637.
42. Wallace, A. C., Laskowski, R. A., and Thornton, J. M. (1995) LIGPLOT: A program to generate schematic diagrams of protein-ligand interactions. *Protein Eng. 8*, 127–134.
43. McDonald, I. K., and Thornton, J. M. (1994) Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol. 238*, 777–793.
44. Susskind, M. M. (1980) A new gene of bacteriophage P22 which regulates synthesis of antirepressor. *J. Mol. Biol. 138*, 685–713.
45. Milla, M. E., and Sauer, R. T. (1994) P22 Arc repressor: Folding kinetics of a single-domain, dimeric protein. *Biochemistry 33*, 1125–1133.
46. Wales, T. E., Richardson, J. S., and Fitzgerald, M. C. (2004) Facile chemical synthesis and equilibrium unfolding properties of CopG. *Protein Sci. 13*, 1918–1926.
47. Burgering, M. J., Hald, M., Boelens, R., Breg, J. N., and Kaptein, R. (1995) Hydrogen exchange studies of the Arc repressor: Evidence for a monomeric folding intermediate. *Biopolymers 35*, 217–226.
48. Pace, C. N. (1975) The stability of globular proteins. *CRC Crit. Rev. Biochem. 3*, 1–43.
49. Shortle, D. (1995) Staphylococcal nuclease: A showcase of m-value effects. *Adv. Protein Chem. 46*, 217–247.
50. Mayo, S. L., and Baldwin, R. L. (1993) Guanidinium Chloride Induction of Partial Unfolding in Amide Proton Exchange in RNase-A. *Science 262*, 873–876.
51. Qian, H., Mayo, S. L., and Morton, A. (1994) Protein hydrogen exchange in denaturant: Quantitative analysis by a two-process model. *Biochemistry 33*, 8167–8171.
52. Hendsch, Z. S., Jonsson, T., Sauer, R. T., and Tidor, B. (1996) Protein stabilization by removal of unsatisfied polar groups: Computational approaches and experimental tests. *Biochemistry 35*, 7621–7625.