BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

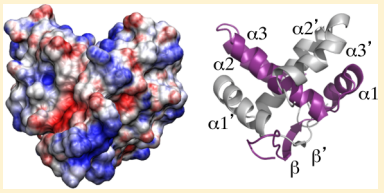
Structural Studies of E73 from a Hyperthermophilic Archaeal Virus Identify the "RH3" Domain, an Elaborated Ribbon−Helix−Helix Motif Involved in DNA Recognition

Casey Schlenker,[†] Anupam Goel,[†] Brian P. Tripet,[†] Smita Menon,[†,⊥] Taylor Willi,[†] Mensur Dlakić,[‡] Mark J. Young,[‡,§,∥] C. Martin Lawrence,*[,†,∥] and Valérie Copié*[,†,∥]

[†]Department of Chemistry and Biochemistry, [‡]Department of Microbiology, [§]Department of Plant Sciences and Plant Pathology, and [∥]Thermal Biology Institute, Montana State University, Bozeman, Montana 59717, United States

Supporting Information

ABSTRACT: Hyperthermophilic archaeal viruses, including *Sulfolobus* spindle-shaped viruses (SSVs) such as SSV-1 and SSV-Ragged Hills, exhibit remarkable morphology and genetic diversity. However, they remain poorly understood, in part because their genomes exhibit limited or unrecognizable sequence similarity to genes with known function. Here we report structural and functional studies of E73, a 73-residue homodimeric protein encoded within the SSV-Ragged Hills genome. Despite lacking significant sequence similarity, the nuclear magnetic resonance (NMR) structure reveals clear similarity to ribbon−helix−helix (RHH) domains present in numerous proteins involved in transcriptional regulation. In vitro double-stranded DNA (dsDNA) binding experiments confirm the ability of E73 to bind dsDNA in a nonspecific manner with micromolar affinity, and characterization of the K11E variant confirms the location of the predicted DNA binding surface. E73 is distinct, however, from known RHH domains. The RHH motif is elaborated upon by the insertion of a third helix that is tightly integrated into the structural domain, giving rise to the "RH3" fold. Within the homodimer, this helix results in the formation of a conserved, symmetric cleft distal to the DNA binding surface, where it may mediate protein−protein interactions or contribute to the high thermal stability of E73. Analysis of backbone amide dynamics by NMR provides evidence of a rigid core, fast picosecond to nanosecond time scale NH bond vector motions for residues located within the antiparallel β-sheet region of the proposed DNA-binding surface, and slower microsecond to millisecond time scale motions for residues in the $\alpha1-\alpha2$ loop. The roles of E73 and its SSV homologues in the viral life cycle are discussed.



The past decade has seen resurgent interest in viruses outside traditional agricultural and medical interests. Reasons include growing appreciation for their enormous abundance, their huge impact on global carbon and nitrogen cycles, the recognition that viruses represent the greatest reservoir of genetic diversity on the planet, and their central role in evolution, where viruses clearly play a major role in horizontal gene transfer.[1−3] In addition, roles for viruses in several major evolutionary transitions have also been proposed, including facilitating the development of DNA and DNA replication mechanisms[4] and the origin of the eukaryotic nucleus.[5] Finally, there is also considerable interest in viral genesis and evolution in and of itself.[3] To evaluate broad evolutionary hypotheses and relationships between viruses, knowledge of viruses infecting the Archaea, the third domain of life, is clearly essential.

Remarkably, the virosphere has now been shown to extend to almost every known environment on earth, including the extreme acidic, thermal, and saline environments where archaeal organisms may dominate.[6,7] It is from these environments that the best studied archaeal viruses have been isolated. However, relative to the bacterial and eukaryotic domains, where more than 5000 viruses have been studied in detail, archaeal viruses are vastly understudied, with fewer than 50 viral species described in any detail.[3]

The Crenarchaea and Euryarchaea are the most intensively studied archaeal phyla. Euryarchaeal viruses are most frequently head-and-tail phages,[8] although there are notable exceptions.[3,9,10] In contrast, viruses infecting the Crenarchaea exhibit remarkably unusual and diverse morphotypes.[3,6,7,11] On the basis of these unusual morphologies and their genetic diversity, crenarchaeal DNA viruses have been classified into 10 viral families, with additional viruses awaiting assignment.[3]

The Fuselloviridae,[12−14] or spindle-shaped viruses (SSVs), which are ubiquitous in acidic hot springs around the world, were among the first crenarchaeal viral families to be recognized and are important model systems for archaeal viruses. SSVs are generally characterized by ∼60 nm × 100 nm lemon-shaped virions, with tail fibers emanating from one end, although some members show more pleomorphic morphologies and package circular double-stranded DNA (dsDNA) genomes approximately 16 kb in size.[14] The genomic sequence for SSV1,[15] the type virus for the Fuselloviridae, contains 34 putative open reading frames (ORFs). The virion has been shown to contain five proteins: VP1−VP3 and small amounts of D244 and

Received: December 6, 2011
Revised: March 8, 2012
Published: March 12, 2012

C792.[16,17] VP1 and VP3 are putative capsid proteins, while VP2 is a small DNA binding protein thought to be involved in packaging viral DNA.

Importantly, only a few ORFs in these viral genomes are significantly identical to genes with known function. For SSV1, ORF D335 encodes an integrase[18] and has been experimentally characterized,[19−22] while B251 shows weak homology to DNaA.[23] More recently, improved bioinformatics approaches have also identified putative zinc finger motifs in B129, C102a, and A79 and suggest the presence of ribbon−helix−helix motifs in C80 and E51.[17,24,25] Despite such predictions, bioinformatics approaches are currently unable to provide functional predictions for a large majority of the SSV1 and other crenarchaeal viral proteomes.

However, because three-dimensional (3D) structural similarities persist longer than similarities in genomic or amino acid sequence,[11,26,27] structural studies can identify distant evolutionary relationships or other structural features that, in turn, suggest functional roles for these proteins.[11,17,27−35] Combined structural and biochemical work on fuselloviral proteins has thus identified D63 as a potential ROP-like adaptor protein,[11,28] F93 and F112 as winged-helix DNA binding proteins,[17,27] and SSV1 D244 and SSV-RH D212 as nucleases with similarity to Holliday junction cleavage enzymes.[34]

Here, we turn our structural focus to ORF E73 from SSV-RH.[14] Though not among the current set of 13 core genes common to all fuselloviruses,[13] E73 and its orthologs are found in six of the eight SSV genomes currently in the public databases,[13,36] where 86% of the residues in SSV-RH E73 are strictly conserved in a multiple-sequence alignment with SSV2 79a, SSV4 73, SSV5 GP23, and SSV6 GP17. The level of identity decreases to 31% between E73 and SSV1 E51, which lacks a C-terminal extension present in other SSV orthologs, but gene synteny is conserved. Similarity to E51 is also noteworthy because of the work of Frols et al.,[24] who found that UV irradiation of the infected host results in a chronologically regulated SSV1 transcription cycle. This begins with a small UV inducible immediate early transcript, and ensuing expression of the three early transcripts, including T5 that encodes E51. This is followed by late transcripts and subsequent viral production.

CD[37] and BLAST[38] searches with E73 do not identify conserved domains or significant similarities to proteins of known function. However, more recent profile−profile search methods like HHpred[39−41] suggest distant similarity to the ribbon−helix−helix DNA binding motif within the N-terminus of E73,[17,24,25] and a C-terminal extension of unknown function. Specifically, four of the top five hits from the HHpred search (pdb70_2Mar12 database) map the first 50 residues of E73 to proteins containing the RHH motif, with *e* values of 7.4×10^{-3} for ParG,[42] 5.5×10^{-2} for CcdA,[43] 0.92 for *Shigella flexneri* yiiF (Northeast Structural Genomics Consortium, unpublished results), and 17 for PutA.[44] Here we report the NMR-based structure and backbone dynamics of E73 that reveal an interesting elaboration on a well-known ribbon−helix−helix motif that we term the RH3 domain, complemented with biochemical work that demonstrates a role for the E73 RH3 domain in DNA recognition and confirms the predicted DNA binding surface. Potential roles for E73 as a transcriptional regulator early in the SSV life cycle are discussed.

EXPERIMENTAL PROCEDURES

Protein Expression and Purification. Cloning, expression, and purification were conducted as previously described.[45] For uniform ^{15}N or tandem ^{15}N and ^{13}C labeling, transformed BL21(DE3) (Invitrogen) bacteria were grown in minimal medium supplemented with 1 g of $^{15}NH_4Cl$ per liter and 2.5 g of [^{13}C]glycerol per liter, respectively, as described by Studier et al.[46] The vector containing the E73 K11E mutant was generated using the QuickChangeII kit (Agilent). Expression and purification of the variant protein were as described above.

Chemical and Thermal Stability. Guanidine hydrochloride (GdnHCl)-induced equilibrium unfolding was conducted by monitoring the molar ellipticity at 222 nm on a Jasco J-810 spectropolarimeter (Jasco Inc., Easten, MD) at a concentration of 10 μM (monomer concentration), as assessed by absorbance measurements at 280 nm (OD_{280} readings) in a buffer that consisted of 0.1 M KCl and 0.05 M K_2HPO_4 (pH 7.0) at 25 °C. Similarly, thermal denaturation was followed by the change in molar ellipticity at 222 nm from 4 to 98 °C in a 1 cm path length cell with the temperature increasing at a rate of 1 °C/min. For both GdnHCl and temperature denaturation experiments, ellipticity readings were normalized to the fraction of protein folded (f_f) or unfolded (f_u) using the standard equation $f_f = ([\theta] - [\theta]_u)/([\theta]_n - [\theta]_u)$ and $f_u = 1 - f_f$, where $[\theta]_n$ and $[\theta]_u$ represent the ellipticity for the fully folded and fully unfolded species, respectively; $[\theta]$ is the observed ellipticity at 222 nm at a given GdnHCl concentration or temperature. Data were analyzed assuming a two-state transition of a folded dimer to two unfolded monomers without intermediates, consistent with denaturation profiles reported for the extremely thermostable RHH gene product ORF56 from *Sulfolobus islandicus* plasmid pRN1.[47] $\Delta G_u(H_2O)$ and dissociation constant K_d were calculated as described by Zeeb et al.[48] and Chao et al.[49]

NMR Spectroscopy. All protein spectra were recorded on a Bruker DRX 600 spectrometer at 312 K (39 °C) in a 50 mM potassium phosphate, 1 mM EDTA, 0.1 mM PMSF, 0.01% sodium azide buffer (pH 5.0) containing 5% (v/v) D_2O or 100% D_2O [for acquisition of 3D ^{13}C-edited $^1H-^1H$ TOCSY, $^1H-^1H$ NOESY, and two-dimensional (2D) $^1H-^{13}C$ CT-HSQC spectra] on E73 protein samples with a dimer concentration of 1 mM (as determined by Bradford assays and OD_{280} readings). Chemical shifts of backbone and side chain atoms were extracted from a series of double- and triple-resonance NMR experiments, including HNCA,[50] HNCACB,[51] CBCA(CO)NH,[52] C(CO)NH,[53] HBHA(CO)NH,[54] HCC-(CO)NH,[53] $^1H-^{13}C$ CT-HSQC,[55] and HCCH-TOCSY[56] as reported in ref 45. ^1H, ^{13}C, and ^{15}N chemical shifts were indirectly referenced to DSS. $^1H-^2H$ solvent exchange experiments were performed by first lyophilizing 500 μL of a 1 mM (dimer) ^{15}N-labeled E73 protein in sample buffer followed by resuspension of the lyophilized protein in the same volume of 100% D_2O, and acquisition of a series of 2D $^1H-^{15}N$ HSQC correlation spectra at subsequent time points following resolubilization of protein in D_2O. A control lyophilization experiment was conducted prior to the H_2O-D_2O exchange whereby a 2D $^1H-^{15}N$ HSQC spectrum of the protein was recorded following lyophilization and resolubilization in H_2O and was found to be identical to the 2D $^1H-^{15}N$ HSQC spectrum recorded in H_2O or a buffer of unlyophilized E73, indicating that the protein structure is unaffected by the lyophilization process.

NMR Structure Calculations. Structure calculations were performed using CNS version 1.2[57] ARIA version 2.2[58] with noncrystallographic symmetry (ncs.def) restraints to account for the fact that E73 is a symmetric homodimer.[45] The initial structure models were guided by 23 experimentally identified intermolecular ^{1}H−^{1}H NOEs (based on a hypothetical model of the E73 RHH core domain) combined with 61 dihedral angle constraints generated by TALOS based on chemical shift data.[59] These models permitted the manual assignment of 527 interproton NOEs. The assignment of ambiguous ^{1}H−^{1}H NOEs was performed with ARIA. ARIA calculations proceeded in nine iterations of spectral calibration and assignment of ambiguous and unambiguous ^{1}H−^{1}H NOEs and estimates of resulting interproton distance restraints followed by calculation of an ensemble of lowest-conformational energy structures. Values for the molecular dynamics and simulated annealing protocol (anneal.inp) and the noncrystallographic symmetry (ncs.def) rotation matrix are reported in Table S0 of the Supporting Information. Following each computational cycle, resulting structures were manually inspected and any restraint violations analyzed and corrected. The ARIA calculations were repeated using NMR assignments generated from previous ARIA iterations until convergence was achieved. For the final set of calculations, 100 structures were generated, and the 20 lowest-energy structures were selected for analysis. In total, 2095 NOE-derived interproton distance restraints, 244 dihedral angle restraints, and 100 hydrogen bond distance restraints (i.e., two distance restraints were assigned per NH identified as being protected from H_2O−D_2O exchange) were used as input for ARIA and CNS (Table 1).

The stereochemistry and quality of these structures were analyzed with PROCHECK-NMR,[60] MolProbity,[61] Verify3D,[62] ProsaII,[63] and electrostatic surfaces calculated with APBS.[64] An ensemble of 20 lowest-energy conformers of E73 has been deposited in the Protein Data Bank as entry 4aai, and code r4aaimr for the NMR restraints. Figures were prepared with PYMOL.[65] Structural homology was identified with the DALI[66] and SSM[67] web servers; superposition utilized the Least Square Alignment protocol of Kabsch et al.[68] as implemented in LSQMAN.[69]

^{15}N NMR Relaxation Experiments. Backbone amide ^{15}N NMR relaxation experiments, including measurements of ^{15}N T_1, ^{15}N T_2, and heteronuclear ^{15}N−^{1}H NOE, were conducted in duplicate on a Bruker DRX 600 MHz solution NMR spectrometer using standard NMR pulse sequences[70−72] at 312 K and pH 5.0, in a manner analogous to that of our recent work with tryptophan repressor variants.[73,74] Briefly, ^{15}N T_1 relaxation profiles were sampled at eight different relaxation delay time points of 40, 96, 200, 400, 600, 1000, and 1200 ms. ^{15}N T_2 relaxation profiles were sampled at eight different relaxation delay periods of 8, 16, 32, 40, 64, 80, 104, and 152 ms, with the delay between the series of ^{15}N 180° pulses applied in the CPMG sequence set to 0.5 ms.[75,76] The ^{15}N T_1 and ^{15}N T_2 data were collected using 512 complex points in the ^{1}H acquisition time dimension (t_2) and 256 complex data points in the ^{15}N (t_1) indirect time evolution dimension, using a WALTZ-16 ^{15}N decoupling scheme during data acquisition. ^{15}N T_1 and ^{15}N T_2 were calculated from the series of NMR spectra using a two-parameter single-exponential decay function:

$$I(t) = I_o \exp(-t/T_{1,2}) \quad (1)$$

where $I(t)$ is the peak height after a delay time t and I_o is the signal height at time zero.

Heteronuclear ^{15}N−^{1}H NOEs were measured using a water-flip-back 2D heteronuclear ^{1}H/^{15}N NOE pulse sequence and the results corrected for the finite repetition delay according to the method of Grzesiek and Bax.[71] For each amide, the ^{15}N−^{1}H NOE was established as the ratio of peak intensities (I/I_o) from NMR data sets acquired with (I) or without (I_o) ^{1}H presaturation. ^{15}N−^{1}H NOE spectra were recorded using

Table 1. Experimental Restraints and Structural Statistics for an Ensemble of the 20 Lowest-Energy Accepted NMR E73 Structures

no. of NMR restraints	
total NOE distance restraints	2095
intraresidue	453 × 2
sequential	239 × 2
short-range	24 × 2
medium-range	191 × 2
long-range	46 × 2
interprotomer	189
H-bond distance restraints[d]	50 × 2
dihedral angle restraints	122 × 2
no. of NMR restraint violations	
NOEs >0.5 Å	0
dihedrals >0.3°	0
energies[e]	
E_{total} = 487.1 ± 8.4	
E_{bond} = 22.7 ± 0.8	
E_{angle} = 158.8 ± 2.5	
E_{NOE} = 109.2 ± 4.8	
E_{cdih} = 4.0 ± 0.7	
E_{vwd} = 172.6 ± 4.2	
E_{imp} = 19.8 ± 1.2	
rmsd (Å)	
bonds	0.00300 ± 0.00005
NOEs	0.0258 ± 0.0007
impropers	0.324 ± 0.010
dihedrals	0.367 ± 0.029
backbone atoms[a,b]	0.50 ± 0.09
heavy atoms[a,c]	1.02 ± 0.14
Ramachandran analysis from Procheck_NMR[a]	
most favored regions	86.6%
additionally favored regions	11.7%
generously allowed regions	1.7%
disallowed regions	0%
Ramachandran plot summary from Molprobity[f]	
most favored regions	94.1%
allowed regions	4.3%
disallowed regions	1.7%
	global quality raw score
Verify3D	0.23
ProsaII	0.55
Procheck[f] ($\phi - \psi$)	−0.18

[a]Root-mean-square deviation (rmsd) and Ramachandran plot analysis calculated for residues 10−70 and derived from Procheck_NMR. [b]Backbone atoms refer to N, C_α, C′, and O atoms of the protein polypeptide chain. [c]Heavy atoms refer to all atoms except hydrogen atoms. [d]Two H-bond distance restraints were assigned per amide hydrogen protected from H−D exchange. [e]Refers to average energies (kilocalories per mole) for the 20 structures of E73 in the ensemble calculated from CNS version 1.2. [f]Molprobity analysis for selected residue ranges: 10−70 (protomer A) and 10′−70′ (protomer B).

dx.doi.org/10.1021/bi201791s | Biochemistry 2012, 51, 2899−2910

1024 and 256 complex data points in t_2 and t_1, respectively, using 48 scans per t_1 increment. The $^{15}N-^{1}H$ NOE experiments were conducted using a delay of 4.5 s between scans to minimize the introduction of systematic errors in measured $^{15}N-^{1}H$ NOEs that could arise from incomplete signal recovery and H_2O saturation.[71] For the data sets acquired with ^{1}H saturation, the 4.5 s recovery delay incorporated a ^{1}H presaturation pulse sequence whereby a series of 90 μs pulses interspersed with a 50 ms delay were applied at an radiofrequency field strength of 5 kHz for ~4 s.

To minimize the impact of magnetic field drift, ^{15}N T_2 and $^{15}N-^{1}H$ NOE data were collected in an interleaved manner, while 2D ^{15}N T_1 data sets were acquired consecutively using a list of shuffled relaxation delay time points.

All relaxation data were processed using NMRPipe, NMRDraw, and Sparky.[77,78] The resulting ^{15}N T_1, ^{15}N T_2, and $^{15}N-^{1}H$ NOE data were then analyzed in terms of spectral density functions as described by Bracken et al.[79] and more recently by Tripet et al.[74] The data were also analyzed using FastModelFree,[80] which interfaces with ModelFree version 4.0.1 of Palmer and colleagues[81] and is based on the extended model-free formalism developed by Lipari and Szabo.[82,83] The ^{15}N NMR relaxation data have been deposited in the BioMagResBank (BMRB) as entry 17069 and are also reported in Tables S1–S3 of the Supporting Information.

Electromobility Shift Assays (EMSAs). A 19-member library of 1000 bp DNA fragments covering the entire SSV-RH genome was generated by PCR and used to screen for E73 DNA recognition. To eliminate the possibility of missing a binding site at the junctions, each 1000 bp fragment (except for one) overlapped the previous fragment by 100 bp. The electrophoretic mobility of each DNA fragment in the presence of E73 was analyzed on agarose gels. Each DNA fragment (100–200 ng) was incubated in 20 μL of binding buffer [10 mM Tris-HCl (pH 8.0), 10 mM HEPES, 1 mM EDTA, 50 mM KCl, 50 μg/mL bovine serum albumin, and 1000 ng of unrelated DNA (300 and 600 bp fragments generated by PCR from the *Staphylococcus aureus* IsdB gene)] for 30 min at 48 °C with increasing amounts of E73 (from 0 to 20 μM). Loading buffer (5 μL of 20 mM Tris-HCl, 10 mM EDTA, and 50% glycerol) was added, and 25 μL of the sample mixture was analyzed on a 1.2% agarose gel with 0.5× TBE running buffer. Samples were run at 135 V for 10 min, followed by gel migration for 1 h at 100 V, and visualized under UV light following sypro-gold (Invitrogen, Inc.) staining.

RESULTS

Thermostable Dimer. Previously reported biophysical characterization utilizing size exclusion chromatography, analytical ultracentrifugation, dynamic light scattering, and 2D $^{1}H-^{15}N$ NMR spectroscopy indicated that E73 is a homodimer in solution with a molecular mass of 20 kDa.[45] We have now also assessed the stability of the E73 protein to chemical and temperature denaturation. Chemical denaturation of E73 with guanidinium hydrochloride follows a two-state unfolding process with a relatively high unfolding midpoint at 4.8 M (Figure 1). Consistent with these data, thermal denaturation analysis also showed the dimeric E73 protein to be extremely thermostable with a T_m of ≥98 °C (Figure 1B, inset). Calculation of the free energy change for unfolding of the E73 dimer [i.e., $\Delta G_u(H_2O)$] yielded a value of 26.5 kcal/mol

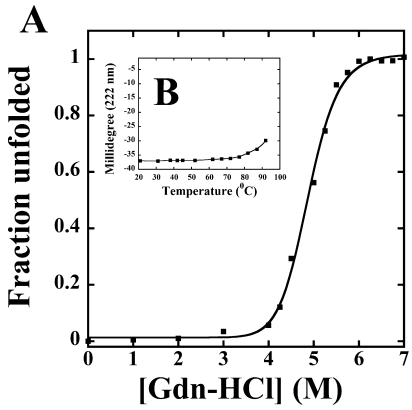


Figure 1. (A) GdnHCl-induced unfolding of E73 monitored by CD at 222 nm and 25 °C in 0.1 M KCl, 0.05 M PO_4 (pH 7) buffer with varying concentrations of GdnHCl. The fraction of unfolded protein (f_u) was calculated as described in Experimental Procedures. The E73 protein concentration was 10 μM. (B) Temperature-induced unfolding of the E73 protein monitored by CD at 222 nm in a 0.1 M KCl, 0.05 M PO_4 buffer (pH 7). The protein concentration was 10 μM.

with an m value of −4.3 kcal mol^{-1} M^{-1} and an estimate of K_d for dissociation of the protomer from the E73 dimer of <1 pM.

Structure of E73. To determine the solution NMR structure of E73, sequential assignments of ^{1}H, ^{13}C, and ^{15}N resonances of backbone and side chain atoms were performed using well-established heteronuclear (^{1}H, ^{15}N, and ^{13}C) multidimensional NMR experiments. If the N-terminal methionine was excluded, the previously reported backbone and side chain resonance assignments were nearly complete (>97%) and have been deposited in the BMRB as entry 16177.[45] $^{13}C_\alpha$, $^{13}C_\beta$, and $^{1}H_\alpha$ chemical shifts, short-range and sequential $^{1}H-^{1}H$ NOEs, and amides protected from hydrogen–deuterium (H/D) exchange were used to identify secondary structural elements of E73 (Figure S1 of the Supporting Information). Each E73 protomer is comprised of an extended β-strand spanning residues 8–16 followed by three α-helices designated $\alpha1-\alpha3$ that span residues 18–31, 35–53, and 61–69, respectively.

We identified an initial set of 15 interprotomer $^{1}H-^{1}H$ NOEs consistent with formation of an antiparallel β-sheet and eight interprotomer $^{1}H-^{1}H$ NOEs orienting helix $\alpha2$ close to helix $\alpha2'$. Using these and other intraprotomer restraints, torsion angle dynamics and simulated annealing calculations using CNS version 1.2 and ARIA version 2.2 were conducted to determine an initial 3D structure of dimeric E73 in solution. Subsequent rounds of manual and automated NOE assignments using ARIA yielded a final set of NOEs (see Table 1) that were used to determine a final ensemble of 20 lowest-energy structures with CNS (Figure 2). This ensemble indicated that the E73 dimer adopts a well-ordered conformation that spans residues 10–69, with disordered C- and N-termini. The rmsd for the final set of E73 structures relative to the mean structure was 0.50 Å for backbone atoms of residues 10–15, 18–31, 35–53, and 56–69, which correspond to the well-defined secondary structural elements of E73, and 1.02 Å for all heavy atoms (Table 1). The N- and C-termini of the protein are disordered, as evidenced by the lack of short- and long-range $^{1}H-^{1}H$ NOEs and amide hydrogen–deuterium (H/D) exchange (Figure S1 of the Supporting Information). ^{15}N NMR relaxation experiments (see below) also indicated that NH bond vectors of amides in these segments of the

dx.doi.org/10.1021/bi201791s | *Biochemistry* 2012, 51, 2899–2910

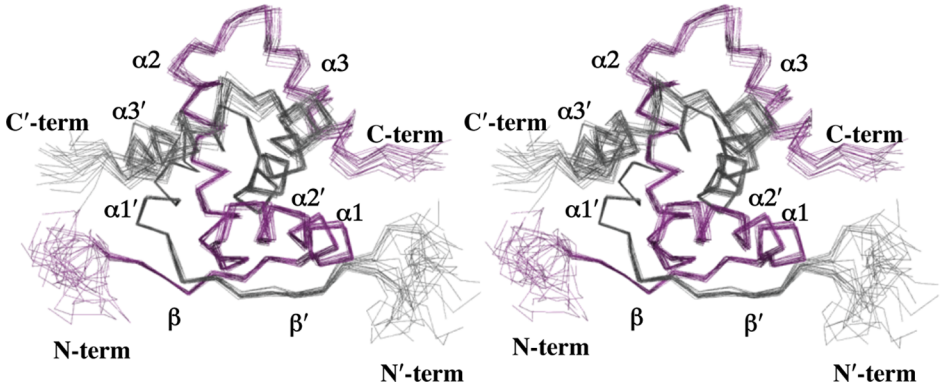


Figure 2. Family of accepted E73 structures. The wall-eye stereoview is shown for the overlay of the backbone heavy atoms (N, C_{α}, and C′) of residues 10−69 for the 20 lowest-energy structures and was generated with PYMOL. The overlay of C_{α} backbone traces is colored in gray for one of the E73 protomers and purple for the second protomer.

protein undergo large amplitude picosecond to nanosecond time scale motions, indicating that the N- and C-termini of E73 are also very flexible.

As suggested by the initial bioinformatics work and secondary structure assignments from the chemical shift index analysis, the structure of the E73 protomer is comprised of an N-terminal β-strand, followed by the three α-helices (α1−α3). Collectively, the β-strand and helices α1 and α2 form a recognizable ribbon−helix−helix (RHH) substructure, with the β-strand of each E73 protomer interacting together to form a short antiparallel β-sheet (Figure 3). The β-strand and first α-

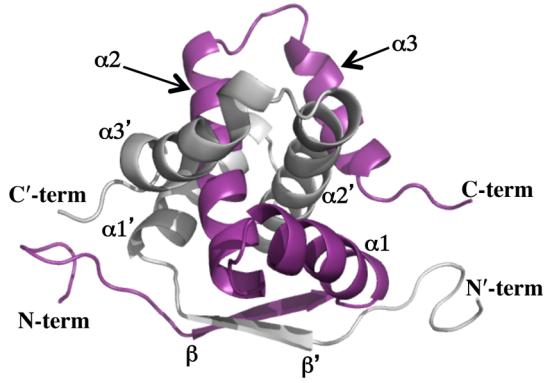


Figure 3. Ribbon representation of a representative conformer from the ensemble of 20 low-energy NMR structures of E73. The two respective chains of the E73 homodimer are colored gray and purple. The secondary structure elements are labeled β (β-sheet, residues 11−16), α1 (α-helix 1, residues 18−32), α2 (α-helix 2, residues 35−53), and α3 (α-helix 3, residues 57−71). The prime symbol denotes the second protomer chain.

helix are connected by a sharp turn centered at Asp17. The solvent-exposed α1−α2 loop is similarly comprised of charged and polar amino acid residues. Most noteworthy is the fact that the amino acid sequence in this loop deviates from the conserved G-X-S/T/N motif found in the loop connecting helices α1 and α2 in the canonical RHH motif.[84] Hydrophobic side chains within helices α1 and α2 (e.g., Leu31, Ile38, Val43, and Trp46) engage in numerous van der Waals contacts within the protein and are responsible for forming a large portion of the E73 hydrophobic core.

This RHH substructure is augmented by a notable and intriguing embellishment, the addition of a third helix (α3) that is tightly integrated within the RHH homodimer. Specifically, helix α2 is followed by a sharp turn (α2−α3) that positions helix α3 to run back toward the N-terminus of α1 (Figure 3). Within the context of the homodimer, this sharp α2−α3 turn serves to fold helix α3 over helix α2′ of the symmetry-related subunit. At the junction of helices α1 and α3, we find a constellation of hydrophobic residues that include Phe16, Phe69, Phe70, and Ile7′. These residues appear to position the C-terminal end of helix α3 toward the N-terminal ends of α1 and β′ in the symmetry-related subunit. The net result is that helices α1−α3 from one subunit completely encircle helix α2′ in the neighboring subunit (Figure 3). Similarly, helices α1′−α3′ of the symmetry-related subunit also encircle α2. This gives a tightly intertwined dimer, one that cannot dissociate without significant rearrangement of the α3 helices.

We also note that the addition of helix α3 to the standard RHH fold creates a structural cleft distal to the antiparallel β-sheet of E73's RHH domain (Figure 4). The tight integration of this third helix within the structural domain and the resulting cleft are features that, to the best of our knowledge, are unique to E73 and have not been observed thus far in structural studies of the RHH fold (see below). The α2−α3 turn creates the upper ridges of this shallow, V-shaped cleft, while the bottom of the cleft is formed by the intertwining of helices α3 and α2′ (Figures 3 and 4). The α2−α3 loop is largely solvent-exposed, with the exception of Leu54 whose side chain forms van der Waals contacts with Leu58 of helix α3′ and Leu31 of helix α1′.

Structural Homology to Ribbon−Helix−Helix Proteins. DALI[66] and SSM[67] were used to search for structural homologues of E73. Significant structural similarity was limited to ∼45 residues within the β-strand and helices α1 and α2, which constitute the well-known RHH fold described above. However, considering the limited sequence identity, the E73 RHH substructure is quite similar to those of other bacterial and viral RHH domains. Similar RHH folds identified by the SSM search included PutA (PDB entry 2RBF, 1.52 Å rmsd for 44 aligned residues with 16% sequence identity[85]), CopG (PDB entry 1EA4, 1.55 Å rmsd for 44 aligned residues with 18% sequence identity[86]), the Arc repressor (PDB entry 1MYK, 1.56 Å rmsd for 44 aligned residues with 14% sequence identity[87]), and SvtR from *S. islandicus* rod-shaped virus (SIRV1), another hyperthermophilic crenarchaeal virus (PDB entry 2KEL, 1.71 Å for 45 aligned residues sharing 27% sequence identity[88]). As an illustration of the structural similarities of the E73 RHH domain to that of other RHH

dx.doi.org/10.1021/bi201791s | *Biochemistry* 2012, 51, 2899−2910

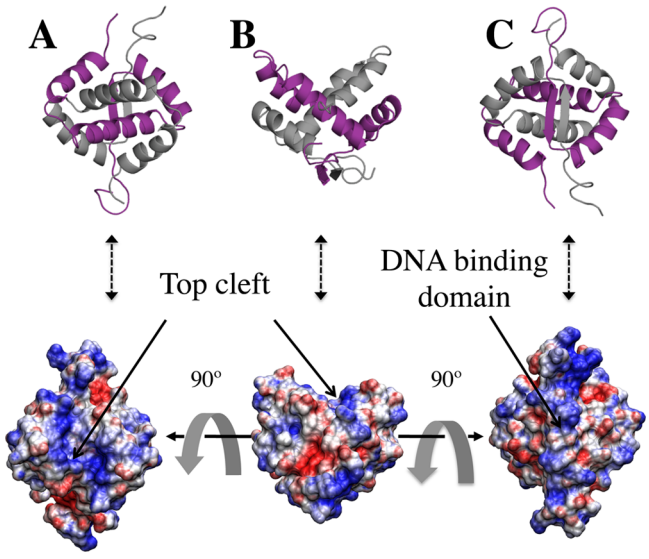


Figure 4. Three distinct orientations of the electrostatic potential surface of E73 (blue for positive charge and red for negative charge) calculated as described in the text. The top figures display the different orientations of the ribbon diagram structure of E73 to highlight the orientations of the electrostatic surfaces with respect to the secondary structural elements of the protein. The color scheme for the two protomers is identical to that of Figure 3. (A) Electrostatic potential surface of E73 highlighting the positively charged nature of the top cleft of E73 (arrow). (B) View of the electrostatic potential surface of E73 rotated by 90° relative to the orientation shown in panel A, with the DNA binding surface of E73 pointing downward and the protein cleft oriented at the top (arrow) and distal to the β-sheet. (C) Molecular surface of E73 rotated by 90° relative to panel B and highlighting the positively charged nature of the protein region engaged in DNA binding.

proteins, a structural overlay of the RHH motif of E73 with PutA and SvtR is shown in Figure 5. However, none of the

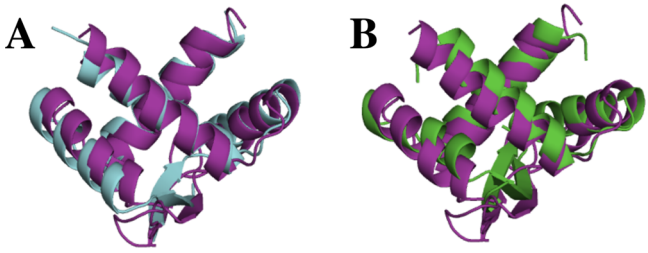


Figure 5. Structural similarity between E73 and other RHH-containing proteins. Overlay of the ribbon–helix–helix (RHH) domain of E73 (purple) (A) with that of DNA-bound PutA (cyan, PDB entry 2RBF) and (B) with the RHH structure of the hyperthermophilic protein SvtR (green, PDB entry 2KEL). All structures are shown in their ribbon representations. The third helix of E73 has been omitted for the sake of clarity and to emphasize the close structural similarities of the RHH motifs.

RHH proteins that are close homologues of E73 show structural similarity beyond the core RHH motif.

Further scrutiny of the potential structural homologues returned by the DALI and SSM servers did identify more distant proteins with an RHH substructure followed by an additional α-helix, but none of these proteins incorporates the third helix in a manner similar to that seen in the E73 RH3 fold (Figure 6). Specific examples of proteins with a third helix

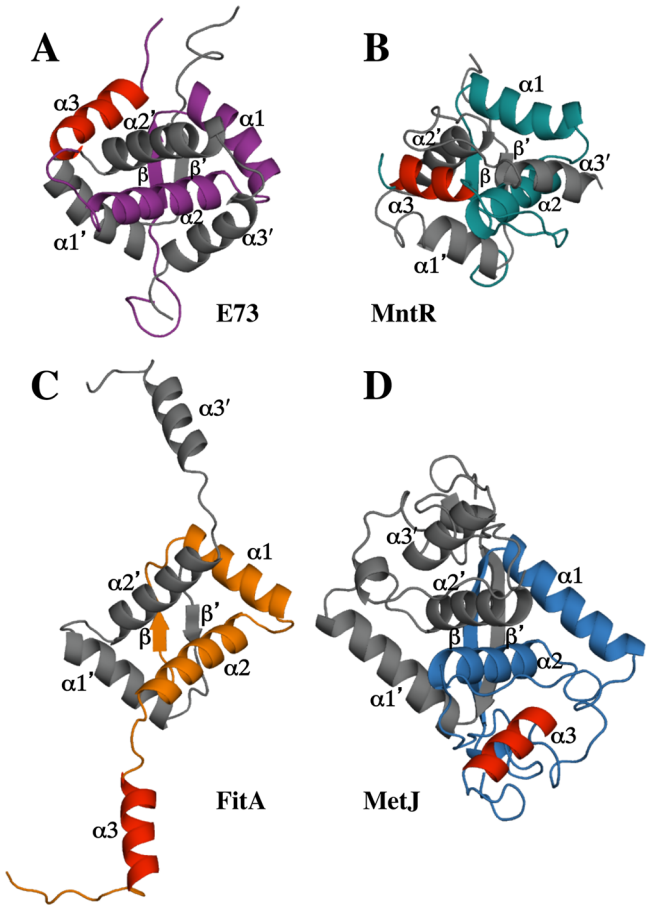


Figure 6. Comparison of the structure of E73 (A) with that of other RHH proteins containing a third helix, including (B) MntR (PDB entry 1MNT), (C) FitA (PDB entry 2BSQ), and (D) MetJ (PDB entry 1MJM). All structures are shown as ribbons with one protomer colored gray and the other colored purple (E73), green (MntR), orange (FitA), or blue (MetJ). The third helix of the colored protomers is colored red for all four proteins to highlight that the orientation of the third helix in E73 relative to its RHH motif is unique to E73 and that all three homologues have very distinct third helix topologies.

include MetJ (PDB entry 1MJM[89]), the antitoxin FitA (PDB entry 2BSQ[90]), and the Mnt repressor (PDB entry 1MNT[91]). MetJ consists of a larger domain that incorporates the RHH substructure, with a third helix that follows the RHH motif. However, the helix does not pack against the RHH substructure and clearly lacks structural equivalence with the third helix of the RH3 fold (see Figure 6A,D). For the smaller FitA in the FitA–FitB toxin–antitoxin complex, the RHH domain is followed by a third helix that extends away from the RHH domain to mediate interactions with FitB (Figure 6C). In the absence of FitB, the residues in this third helix are disordered. In MntR, an additional poorly ordered helix follows the RHH substructure but runs perpendicular to helix α2′, rather than crossing over it to complete encirclement of helix α2′ (Figure 6B).

DNA Binding Site and Superpositional Docking. Least-squares superposition was used to dock E73 on the structure of the PutA DNA operator complex to produce a model E73–DNA complex (Figure 7A). This docking model orients a positively charged surface at the E73–DNA interface and places the antiparallel β-sheet within the DNA major groove where it

dx.doi.org/10.1021/bi201791s | *Biochemistry* 2012, 51, 2899–2910

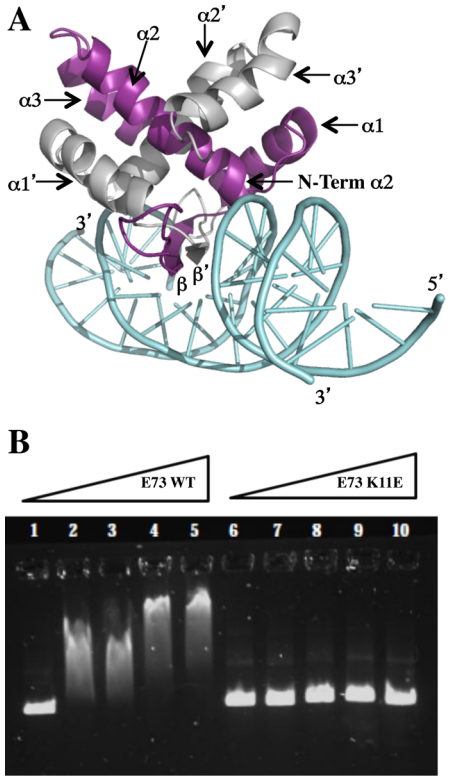


Figure 7. (A) Ribbon representation of the E73 homodimer docked by superposition onto PutA DNA (PDB entry 2RBF). The two protomers are colored purple and gray, and DNA is colored cyan. Similar to PutA, the β-sheet interacts with the DNA major groove, and the N-terminus of helix α2 appears to interact with the ribose. (B) EMSA demonstrating the loss of nonspecific viral DNA binding affinity by the E73 K11E variant. The E73–DNA intreaction was followed by an EMSA. Increasing concentrations of wild-type E73 (lanes 1–5) and the E11K variant (lanes 6–10) were incubated with 182 ng of SSV-RH DNA (SSV-RH oligonucleotides 2900–3900). Protein (monomer) concentrations were 0 μM (lanes 1 and 6), 5 μM (lanes 2 and 7), 10 μM (lanes 3 and 8), 15 μM (lanes 4 and 9), and 20 μM (lanes 5 and 10). In contrast to wild-type E73, the K11E variant failed to shift DNA at any of the concentrations examined (1–20 μM).

is poised to make a number of base specific interactions, as well as more generic interactions with the ribose–phosphate backbone (Figure 7A). This superposition suggests potential base specific interactions mediated by Lys11, Thr13, or Ala15, while interactions with the ribose–phosphate backbone are likely to be mediated by basic residues in the N-terminus of E73, including Lys5, Lys6, Lys9, Lys10, and perhaps Lys11. In addition, the main chain amides at the N-terminal end of helix α2 that do not participate in hydrogen bonding and the α2 helix dipole are also suggested to interact with the negatively charged phosphate in the DNA backbone. Two side chains at the N-terminal end of helix α2 may also be involved. Specifically, Arg33 and Thr36 might each interact with the ribose–phosphate backbone. Alternatively, Thr36 could also make a base specific interaction. In addition, the docking exercise also places the C-terminus of helix α3 near the ribose–phosphate backbone, and additional interactions between E73 and DNA could be mediated by basic residues in helix α3, such as Lys73. Overall, with the exception of helix α3, this structural arrangement is quite similar to that observed for the PutA and Arc repressor–DNA complexes.[85,92] This further suggests that basic residues within the disordered N-terminus of E73 will interact with the ribose–phosphate backbone.

E73 Binds dsDNA. A series of overlapping 1000 bp fragments covering the entire SSV-RH genome was generated by PCR and used in an initial screen for potential E73 binding sites in the SSV-RH genome. Each 1000 bp fragment was incubated with increasing amounts of E73 (0–20 μM monomer, concentrations at which E73 is a stable dimer as assessed by CD and the thermal denaturation experiments) and analyzed by agarose gel electrophoresis. In general, 1 μM E73 shifted each of the 19 PCR-amplified products, but as E73 concentrations fell below 1 μM (monomer), the DNA band shift was uniformly lost for each of the 19 fragments, suggesting the shifts at higher concentrations were due to nonspecific interactions. This was confirmed with control experiments using unrelated plasmid DNA that also gave a dissociation constant of ~1 μM. Thus, while we were unable to identify specific E73 binding sites within the SSV-RH genome, this work does confirm the ability of E73 to recognize dsDNA in a nonspecific manner with relatively high affinity.

To verify the predicted E73 DNA binding surface, we then characterized the E73 K11E variant in which Lys11 had been replaced with a glutamic acid residue (Figure 7B) (see Figure S6 of the Supporting Information). In comparison to data for wild-type E73 (lanes 2–5), the EMSA data show that the K11E variant (lanes 7–10) has lost the ability to interact with dsDNA (up to a monomer concentration of 20 μM). The weakened affinity is not due to a structural change, because the K11E variant retains similar biophysical properties and, with the exception of the NMR resonance of K11, exhibits a 2D ^{1}H–^{15}N HSQC spectrum identical to that of wild-type E73.

E73 Backbone Amide Dynamics. ^{15}N NMR relaxation parameters (^{15}N T_1, ^{15}N T_2, and ^{15}N–^{1}H NOE) were measured for 58 of the 70 assigned non-proline residue amides of E73. Data for the 12 remaining residues were omitted from the analysis because of overlapped, very weak, or unobservable ^{15}N–^{1}H NMR resonances. The ^{15}N NMR data (see Figures S2–S4 and Tables S1 and S2 of the Supporting Information) were analyzed in terms of spectral density functions and the extended Lipari–Szabo ModelFree formalism, which characterize the internal dynamics of the protein on the nanosecond to picosecond and microsecond to millisecond time scales. As discussed in the Supporting Information, backbone amides in the well-ordered segment of the protein (residues 16–69) display relatively uniform relaxation profiles that are typical of well-folded secondary structural elements of a protein with limited internal dynamics, with a few residues deviating slightly from these trends. More importantly, the ^{15}N NMR relaxation profiles of E73 are very similar to what has been observed for other crenarchaeal viral proteins from hyperthermophiles, including RHH proteins ORF56[47] and SvtR[88] from *S. islandicus*. This is in contrast to what has been observed for the antitoxin ParD protein whose N-terminal RHH domain, although structured in solution, is more flexible on a picosecond to nanosecond time scale than E73, ORF56, or StvtR.[93] ParD is also less thermostable than E73 (T_m of 64 °C),[94] supporting the notion that restricted internal dynamics is an important factor in the thermostability of hyperthermophilic crenarchaeal viral proteins like E73.

DISCUSSION

RHH domain proteins represent a large and diverse family of proteins found throughout all domains of life but are

dx.doi.org/10.1021/bi201791s | *Biochemistry* 2012, 51, 2899–2910

particularly common in bacteria, archaea, and their viruses where they typically serve as transcription factors.[84] For example, the RHH domain is found within larger multidomain proteins such as NikR, the nickel-dependent regulator of nickel uptake in *Escherichia coli*,[95] and PutA, the proline utilization flavoprotein.[85] In addition to their RHH domains, these proteins contain associated ligand-binding or catalytic domains. The RHH domain is also found as a stand-alone module. Examples here include CopG, a transcriptional repressor that in turn represses plasmid replication,[96] the Arc repressor, whose DNA binding function initiates the switch between the lytic and lysogenic states in *Salmonella* bacteriophage P22,[92] and SvtR, a putative transcriptional repressor from archeal rod-shaped virus SIRV1.[88] Structural homology analysis, however,[67,97,98] does not identify any RHH domain-containing protein with a third helix structurally equivalent to helix α3 in E73. The structure of E73 is thus the first observation of the RH3 fold, in which a third helix (α3) is utilized to complete the encirclement of helix $\alpha2'$ in the neighboring RHH subunit.

Interestingly, the tightly intertwined E73 homodimer would appear to be incapable of dissociating into monomers without significant rearrangement of helix α3. At the same time, the $^{1}H-^{2}H$ exchange and ^{15}N NMR relaxation data indicate that helix α3 is well-ordered on the surface of the E73 dimer and is not particularly flexible (Figures S2−S4 of the Supporting Information). Combined, these data and the intertwined nature of the E73 structure suggest that helix α3 plays an important role in stabilizing an intrinsic E73 homodimer. In addition, because helix α3 increases the volume of the hydrophobic core, and because oligomerization is also a well-recognized contributor to thermostability,[99] helix α3 is also likely to make a significant contribution to the overall thermostability of E73 (Figure 1). Thus, helix α3 clearly plays an important structural role.

However, as evidenced by SvtR,[88] RHH domains are clearly thermostable in the absence of α3-like helices. Thus, functional roles for helix α3 should also be considered. Specifically, within the context of the E73 homodimer, the $\alpha2-\alpha3$ turn and the N-terminal end of the third helix combine to form a 2-fold symmetric cleft that is distal to the canonical DNA binding surface of the RHH fold (Figure 4). This positively charged cleft is clearly the largest pocket or cavity on the protein surface and for this reason should be considered as a potential ligand binding site.[100,101] We cannot say, however, whether this potential ligand-binding surface is highly conserved, as the unusually high level of sequence conservation among the E73 orthologs does not truly allow the discrimination of conserved surface features. If this is a ligand-binding site, its positively charged surface suggests it would recognize a ligand with complementary negative charge. It remains to be determined whether this might be a viral or host protein, a nucleic acid, or a regulatory small molecule.

Using an EMSA, we have screened the entire SSV-RH genome for potential E73 binding sites using a series of overlapping 1000 bp fragments.[88] While E73 is capable of shifting each of these fragments, it is unable to discriminate among them, binding each of them with dissociation constants of ~1 μM. This suggests a nonspecific binding interaction with viral DNA, perhaps because of electrostatic interactions of the five lysine residues at the extreme N-terminus (i.e., K5, K6, K9, K10, and K11) with the negatively charged ribose−phosphate backbone. A structure-based alignment of the amino acid sequence of E73 compared to the amino acid sequence of other RHH proteins also highlights the unusual cluster of lysine residues at the extreme N-terminus of E73 that are missing in other RHH homologues (see Figure S5 of the Supporting Information). It is interesting to note that replacement of Lys with a Glu at position 11 abrogates nonspecific DNA binding, providing supporting evidence of a functional role of the N-terminal lysine residues of E73 in DNA binding.

The observation of nonspecific interactions with viral DNA leads us to postulate that instead of regulating viral DNA expression, the true DNA targets of E73 may instead reside within the host genome and that E73 may regulate expression of *Sulfolobus* host genes. This conclusion is consistent with the rich abundance of putative transcription factors found in fuselloviral genomes.[11,17,27] However, a simple gene candidate approach is impractical for screening the larger host genome, and strategies that evaluate the whole genome such as ChIP-on-Chip or ChIP-Seq will probably be required to address this issue. We also note that each of the stand-alone RHH domain-containing proteins showing the greatest structural similarity to E73, i.e., CopG, Arc repressor, and SvtR, serves as a transcriptional repressor. In light of the minimal structural features present in E73, it is likely the E73 will also serve as a repressor, targeting the host genome.

This scenario would also be consistent with a recent microarray study that observed tight temporal control of SSV1 gene expression and viral production following UV irradiation.[24] The process begins with expression of the immediate early gene B49, followed by expression of the T5 and T6 early transcripts. SSV1 E51, an E73 ortholog, is found in transcript T5 and is thus expressed early in the viral life cycle. Were it to serve as general repressor of viral transcription, this would be at odds with the continued temporal expression of the viral genome and viral production. It is also at odds with a CopG-like function for E73 and its homologues, suggesting that direct functional analogies to CopG, which represses expression of the RepB nickase needed to initiate rolling circle replication, should be avoided.[6]

In conclusion, our structural and functional analysis of SSV-RH E73 has revealed a unique adaptation on the RHH fold that we designate RH3. The addition of the third helix results in a tightly intertwined RHH homodimer, which in the case of E73 shows significant thermostability, and a notable cleft distal to the canonical DNA binding site. Structural similarity to CopG, the Arc repressor, and SvtR suggests that E73 is most likely to act as a transcriptional repressor. However, the inability of E73 to discriminate a high-affinity binding site in the SSV-RH genome suggests that rather than regulating transcription of the viral genome, it will instead target elements of the host genome. Further, in conjunction with the microarray study of Frols et al. on SSV1,[24] it appears this will occur early in the viral life cycle.

■ ASSOCIATED CONTENT

Supporting Information

A summary of amide H−D exchange data, patterns of sequential and short-range NOEs, and differences in $^{13}C_{\alpha/\beta}$ and $^{1}H_{\alpha}$ chemical shifts from random coil values for SSV-RH E73 (Figure S1); plots of ^{15}N NMR relaxation profiles and dynamics parameters resulting from reduced spectral density function and ModelFree analysis of the ^{15}N NMR relaxation data of E73 (Figures S2−S4); a structure-based amino acid sequence alignment of E73 highlighting structurally equivalent residues among RHH homologues (Figure S5); a structural representation of E73 residues that may be involved in DNA

binding (Figure S6); a list of experimental parameters used in the structure calculations of E73 (Table S0); ^{15}N NMR relaxation parameters ^{15}N T_1, ^{15}N T_2, and ^{15}N−^{1}H NOE (Table S1); reduced spectral density functions $J_{eff}(0)$, $J(\omega_N)$, and $J(0.87\omega_H)$ (Table S2); motional parameters (S^2, τ_e, R_{ex}, and model selection) computed for E73 (Table S3); and a short section of text describing the ^{15}N NMR relaxation analysis and internal dynamics results. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Department of Chemistry and Biochemistry, Montana State University, P.O. Box 173400, Bozeman, MT 59717. V.C.: e-mail, vcopie@chemistry.montana.edu; phone, (406) 994-7244; fax, (406) 994-5407. C.M.L.: e-mail, Lawrence@chemistry.montana.edu; phone, (406) 994-5382; fax, (406) 994-5407.

Present Address

⊥Department of Biochemistry and Molecular Biology, Uniformed Services University of the Health Sciences, Bethesda, MD 20814.

Author Contributions

C.S., A.G., and B.P.T. contributed equally to this work.

Funding

This work has been supported by National Science Foundation Grant MCB-0920312. The NMR experiments were conducted at Montana State University on a Bruker DRX600 NMR spectrometer purchased in part with funds from the National Institutes of Health Shared Instrumentation Grant program (Grant 1-S10RR13878-01). The Montana State University Research Experience for Undergraduate program (NSF-0852043) provided support for T.W.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We acknowledge the Montana State University Research Experience for Undergraduate (REU) program (NSF-0852043), which supported REU students Guthrie Lewis and Kelly Mildenberger. We are grateful to Guthrie and Kelly for help with expression and purification of ^{15}N-labeled and ^{15}N- and ^{13}C-labeled protein.

ABBREVIATIONS

APBS, Adaptive Poisson−Boltzmann Solver; ARIA, Ambiguous Restraints for Iterative Assignment; BMRB, BioMagResBank; CD, circular dichroism; CPMG, Carr−Purcell−Meiboom−Gill multipulse sequence; DIPSI, decoupling in the presence of scalar interactions; dsDNA, double-stranded DNA; EDTA, ethylenediaminetetraacetic acid; EMSA, electromobility shift assay; GdnHCl, guanidine hydrochloride; HSQC, heteronuclear single-quantum coherence spectroscopy; IPTG, isopropyl β-thiogalactoside; NMR, nuclear magnetic resonance spectroscopy; NOESY, nuclear Overhauser enhancement spectroscopy; NOE, nuclear Overhauser effect; ORF, open reading frame; PCR, polymerase chain reaction; PDB, Protein Data Bank; RHH, ribbon−helix−helix; rmsd, root-mean-square deviation; SSM, secondary structure matching; SSV-RH, *Sulfolobus* spindle-shaped virus-Ragged Hills; TBE buffer, Tris/Borate/EDTA buffer; WALTZ, wideband alternative-phase low-power technique for zero residual splitting.

REFERENCES

(1) Breitbart, M., and Rohwer, F. (2005) Here a virus, there a virus, everywhere the same virus? *Trends Microbiol. 13*, 278−284.

(2) Suttle, C. A. (2007) Marine viruses: Major players in the global ecosystem. *Nat. Rev. Microbiol. 5*, 801−812.

(3) Pina, M., Bize, A., Forterre, P., and Prangishvili, D. (2011) The archeoviruses. *FEMS Microbiol. Rev. 35*, 1035−1054.

(4) Forterre, P. (2006) Three RNA cells for ribosomal lineages and three DNA viruses to replicate their genomes: A hypothesis for the origin of cellular domain. *Proc. Natl. Acad. Sci. U.S.A. 103*, 3669−3674.

(5) Bell, P. J. (2001) Viral eukaryogenesis: Was the ancestor of the nucleus a complex DNA virus? *J. Mol. Evol. 53*, 251−256.

(6) Ortmann, A. C., Wiedenheft, B., Douglas, T., and Young, M. (2006) Hot crenarchaeal viruses reveal deep evolutionary connections. *Nat. Rev. Microbiol. 4*, 520−528.

(7) Prangishvili, D., Forterre, P., and Garrett., R. A. (2006) Viruses of the Archaea: A unifying view. *Nat. Rev. Microbiol. 4*, 837−848.

(8) Haring, M., Peng, X., Brugger, K., Rachel, R., Stetter, K. O., Garrett, R. A., and Prangishvili, D. (2004) Morphology and genome organization of the virus PSV of the hyperthermophilic archaeal genera *Pyrobaculum* and *Thermoproteus*: A novel virus family, the Globuloviridae. *Virology 323*, 233−242.

(9) Bath, C., Cukalac, T., Porter, K., and Dyall-Smith, M. L. (2006) His1 and His2 are distantly related, spindle-shaped haloviruses belonging to the novel virus group, Salterprovirus. *Virology 350*, 228−239.

(10) Bath, C., and Dyall-Smith, M. L. (1998) His1, an archaeal virus of the Fuselloviridae family that infects Haloarcula hispanica. *J. Virol. 72*, 9392−9395.

(11) Lawrence, C. M., Menon, S., Eilers, B. J., Bothner, B., Khayat, R., Douglas, T., and Young, M. J. (2009) Structural and functional studies of archaeal viruses. *J. Biol. Chem. 284*, 12599−12603.

(12) Peng, X., Blum, H., She, Q., Mallok, S., Brugger, K., Garrett, R. A., Zillig, W., and Prangishvili, D. (2001) Sequences and replication of genomes of the archaeal Rudiviruses SIRV1 and SIRV2: Relationships to the archaeal Lipothrixvirus SIFV and some eukaryal viruses. *Virology 291*, 226−234.

(13) Redder, P., Peng, X., Brugger, K., Shah, S. A., Roesch, F., Greve, B., She, Q., Schleper, C., Forterre, P., Garrett, R. A., and Prangishvili, D. (2009) Four newly isolated fuselloviruses from extreme geothermal environments reveal unusual morphologies and a possible interviral recombination mechanism. *Environ. Microbiol. 11*, 2849−2862.

(14) Wiedenheft, B., Stedman, K., Roberto, F., Willits, D., Gleske, A.-K., Zoeller, L., Snyder, J., Douglas, T., and Young, M. (2004) Comparative Genomic Analysis of Hyperthermophilic Archaeal Fuselloviridae Viruses. *J. Virol. 78*, 1954−1961.

(15) Palm, P., Schleper, C., Grampp, B., Yeats, S., McWilliam, P., Reiter, W. D., and Zillig, W. (1991) Complete nucleotide sequence of the virus SSV1 of the archaebacterium *Sulfolobus shibatae*. *Virology 185*, 242−250.

(16) Reiter, W. D., Palm, P., Henschen, A., Lottspeich, F., Zillig, W., and Grampp, B. (1987) Identification and characterization of the genes encoding three structural proteins of the *Sulfolobus* virus-like particle SSV1. *Mol. Genet. Genomics 206*, 144−153.

(17) Menon, S. K., Maaty, W. S., Corn, G. J., Kwok, S. C., Eilers, B. J., Kraft, P., Gillitzer, E., Young, M. J., Bothner, B., and Lawrence, C. M. (2008) Cysteine usage in *Sulfolobus* spindle-shaped virus 1 and extension to hyperthermophilic viruses in general. *Virology 376*, 270−278.

(18) Argos, P., Landy, A., Abremski, K., Egan, J. B., Haggard-Ljungquist, E., Hoess, R. H., Kahn, M. L., Kalionis, B., Narayana, S. V., Pierson, L. S. III, et al. (1986) The integrase family of site-specific recombinases: Regional similarities and global diversity. *EMBO J. 5*, 433−440.

(19) Muskhelishvili, G., Palm, P., and Zillig, W. (1993) SSV1-encoded site-specific recombination system in *Sulfolobus shibatae*. *Mol. Gen. Genet. 237*, 334−342.

(20) Clore, A. J., and Stedman, K. M. (2007) The SSV1 viral integrase is not essential. *Virology 361*, 103−111.

(21) Letzelter, C., Duguet, M., and Serre, M. C. (2004) Mutational analysis of the archaeal tyrosine recombinase SSV1 integrase suggests a mechanism of DNA cleavage in trans. *J. Biol. Chem. 279*, 28936−28944.
(22) Serre, M. C., Letzelter, C., Garel, J. R., and Duguet, M. (2002) Cleavage properties of an archaeal site-specific recombinase, the SSV1 integrase. *J. Biol. Chem. 277*, 16758−16767.
(23) Koonin, E. V. (1992) Archaebacterial virus SSV1 encodes a putative DnaA-like protein. *Nucleic Acids Res. 20*, 1143.
(24) Frols, S., Gordon, P. M., Panlilio, M. A., Schleper, C., and Sensen, C. W. (2007) Elucidating the transcription cycle of the UV-inducible hyperthermophilic archaeal virus SSV1 by DNA microarrays. *Virology 365*, 48−59.
(25) Prangishvili, D., Garrett, R. A., and Koonin., E. V. (2006) Evolutionary genomics of archaeal viruses: Unique viral genomes in the third domain of life. *Virus Res. 117*, 52−67.
(26) Moult, J., and Melamud, E. (2000) From fold to function. *Curr. Opin. Struct. Biol. 10*, 384−389.
(27) Kraft, P., Oeckinghaus, A., Kummel, D., Gauss, G. H., Gilmore, J., Wiedenheft, B., Young, M., and Lawrence, C. M. (2004) Crystal Structure of F-93 from *Sulfolobus* Spindle-Shaped Virus 1, a Winged-Helix DNA Binding Protein. *J. Virol. 78*, 11544−11550.
(28) Kraft, P., Kummel, D., Oeckinghaus, A., Gauss, G. H., Wiedenheft, B., Young, M., and Lawrence, C. M. (2004) Structure of D-63 from *Sulfolobus* Spindle-Shaped Virus 1: Surface Properties of the Dimeric Four-Helix Bundle Suggest an Adaptor Protein Function. *J. Virol. 78*, 7438−7442.
(29) Larson, E. T., Eilers, B., Menon, S., Reiter, D., Ortmann, A., Young, M. J., and Lawrence, C. M. (2007) A winged-helix protein from *Sulfolobus* turreted icosahedral virus points toward stabilizing disulfide bonds in the intracellular proteins of a hyperthermophilic virus. *Virology 38*, 249−261.
(30) Larson, E. T., Eilers, B. J., Reiter, D., Ortmann, A. C., Young, M. J., and Lawrence., C. M. (2007) A new DNA binding protein highly conserved in diverse crenarchaeal viruses. *Virology 363*, 387−396.
(31) Larson, E. T., Reiter, D., Young, M., and Lawrence., C. M. (2006) Structure of A197 from *Sulfolobus* turreted icosahedral virus: A crenarchaeal viral glycosyltransferase exhibiting the GTA fold. *J. Virol. 80*, 7636−7644.
(32) Khayat, R., Tang, L., Larson, E. T., Lawrence, C. M., Young, M., and Johnson, J. E. (2005) Structure of an archaeal virus capsid protein reveals a common ancestry to eukaryotic and bacterial viruses. *Proc. Natl. Acad. Sci. U.S.A. 102*, 18944−18949.
(33) Lintner, N. G., Frankel, K. A., Tsutakawa, S. E., Alsbury, D. L., Copie, V., Young, M. J., Tainer, J. A., and Lawrence, C. M. (2011) The structure of the CRISPR-associated protein Csa3 provides insight into the regulation of the CRISPR/Cas system. *J. Mol. Biol. 405*, 939−955.
(34) Menon, S. K., Eilers, B. J., Young, M. J., and Lawrence., C. M. (2010) The Crystal Structure of D212 from *Sulfolobus* Spindle-shaped Virus Ragged Hills reveals a new member of the PD-(D/E)XK nuclease superfamily. *J. Virol. 84*, 5890−5897.
(35) Lintner, N. G., Kerou, M., Brumfield, S. K., Graham, S., Liu, H., Naismith, J. H., Sdano, M., Peng, N., She, Q., Copie, V., Young, M. J., White, M. F., and Lawrence, C. M. (2011) Structural and functional characterization of an archaeal clustered regularly interspaced short palindromic repeat (CRISPR)-associated complex for antiviral defense (CASCADE). *J. Biol. Chem. 286*, 21643−21656.
(36) Stedman, K. M., She, Q., Phan, H., Arnold, H. P., Holz, I., Garrett, R. A., and Zillig., W. (2003) Relationships between fuselloviruses infecting the extremely thermophilic archaeon *Sulfolobus*: SSV1 and SSV2. *Res. Microbiol. 154*, 295−302.
(37) Marchler-Bauer, B. S. (2004) CD-Search: Protein domain annotations on the fly. *Nucleic Acids Res. 32*, W327−W331.
(38) Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res. 25*, 3389−3402.
(39) Soding, J. (2005) Protein homology detection by HMM-HMM comparison. *Bioinformatics 21*, 951−960.
(40) Soding, J., Biegert, A., and Lupas, A. N. (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res. 33*, W244−W248.
(41) Dlakic, M. (2009) HHsvm: Fast and accurate classification of profile-profile matches identified by HHsearch. *Bioinformatics 25*, 3071−3076.
(42) Golovanov, A. P., Barilla, D., Golovanova, M., Hayes, F., and Lian, L. Y. (2003) ParG, a protein required for active partition of bacterial plasmids, has a dimeric ribbon-helix-helix structure. *Mol. Microbiol. 50*, 1141−1153.
(43) 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.
(44) Larson, J. D., Jenkins, J. L., Schuermann, J. P., Zhou, Y., Becker, D. F., and Tanner, J. J. (2006) Crystal structures of the DNA binding domain of *Escherichia coli* proline utilization A flavoprotein and analysis of the role of Lys 9 in DNA recognition. *Protein Sci. 15*, 2630−2641.
(45) Schlenker, C., Menon, S., Lawrence, C. M., and Copié, V. (2009) ^{1}H, ^{13}C, ^{15}N backbone and side chain NMR resonance assignments for E73 from *Sulfolobus* spindle-shaped virus ragged hills, a hyperthermophilic crenarchaeal virus from Yellowstone National Park. *Biomol. NMR Assignments 3*, 219−222.
(46) Studier, F. W. (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expression Purif. 41*, 207−234.
(47) Weininger, U., Zeeb, M., Neumann, P., Lowm, C., Stubbs, M. T., Lipps, G., and Balbach, J. (2009) Structure-based stability analysis of an extremely stable dimeric DNA binding protein from *Sulfolobus islandicus*. *Biochemistry 48*, 10030−10037.
(48) 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.
(49) Chao, H., Houston, M. E. Jr., Grother, S., Kay, C. M., O'connor-McCourt, M., Irvin, R. T., and Hodges, R. S. (1996) Kinetic study on the formation of a de novo designed heterodimeric coiled-coil: Use of surface plasmon resonance to monitor the association and dissociation of polypeptide chains. *Biochemistry 35*, 12175−12185.
(50) Kay, L. E., Ikura, M., Tschudin, R., and Bax, A. (1990) Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. *J. Magn. Reson. 89*, 496−514.
(51) Wittekind, M., and Mueller, L. (1993) HNCACB, a high-sensitivity 3D NMR experiment to correlate amide-proton and nitrogen resonances with α and proton and nitrogen resonances with α and β carbon resonances in proteins. *J. Magn. Reson. 101*, 201−205.
(52) Grzesiek, S., and Bax, A. (1992) An Efficient Experiment for Sequential Backbone Assignment of Medium-Sized Isotopically Enriched Proteins. *J. Magn. Reson. 99*, 201−207.
(53) Grzesiek, S., Anglister, J., and Bax, A. (1993) Correlation of backbone amide and aliphatic side-chain resonances in ^{13}C/^{15}N-enriched proteins by isotropic mixing of ^{13}C magnetization. *J. Magon. Reson., Ser. B 101*, 114−119.
(54) Grzesiek, S., and Bax, A. (1992) Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *J. Am. Chem. Soc. 114*, 6291−6293.
(55) Vuister, G. W., and Bax, A. (1992) Resolution enhancement and spectral editing of uniformy ^{13}C-enriched proteins by homonuclear braodband ^{13}C decoupling. *J. Magn. Reson. 98*, 428−435.
(56) Bax, A., Clore, M. G., and Gronenborn, A. M. (1990) ^{1}H-^{1}H correlation via isotropic mixing of ^{13}C magnetization, a new three-dimensional approach for assigning ^{1}H and ^{13}C spectra of ^{13}C enriched proteins. *J. Magn. Reson. 88*, 425−431.
(57) Brunger, A. T., Adams, P. D., Clore, M. G., Delano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system.CNS: A new software suite for macromolecular structure determination. *Acta Crystallogr. D54*, 901−921.

(58) Habeck, M., Rieping, W., Linge, J. P., and Nilges, M. (2004) NOE assignment with ARIA 2.0: The nuts and bolts. *Methods Mol. Biol. 278*, 379−402.
(59) Cornilescu, G., Delaglio, F., and Bax, A. (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR 13*, 289−302.
(60) 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.
(61) Chen, V. B., Arendall, W. B. III, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. 66*, 12−21.
(62) Luthy, R., Bowie, J. U., and Eisenberg, D. (1992) Assessment of protein models with three-dimensional profiles. *Nature 356*, 83−85.
(63) Wiederstein, M., and Sippl, M. J. (2007) ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res. 35* (Suppl. 2), W407−W410.
(64) Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001) Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A. 98*, 10037−10041.
(65) DeLano, W. L. (2002) *The PyMOL Molecular Graphics System*, DeLano Scientific, San Carlos, CA.
(66) Holm, L., Kaariainen, S., Rosenstrom, P., and Schenkel, A. (2008) Searching protein structure databases with DaliLite v.3. *Bioinformatics 24*, 2780−2781.
(67) Krissinel, E., and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D60*, 2256−2268.
(68) Kabsch, W. (1978) A discussion of the solution for the best rotation to relate two sets of vectors. *Acta Crystallogr. A34*, 827−828.
(69) Kleywegt, G. J.; Zou, J. Y.; Kjeldgaard, M.; Jones, T. A. In *"International Tables for Crystallography, Vol. F. Crystallography of Biological Macromolecules"*. Rossmann, M. G. and Arnold, E., Eds. Kluwer Academic Publishers: Dordrecht, The Netherlands, 2001, Ch 17, pp 353−356 and 366−367.
(70) Barbato, G., Ikura, M., Kay, L. E., Pastor, R. W., and Bax, A. (1992) Backbone dynamics of calmodulin studied by ^{15}N relaxation using inverse detected two-dimensional NMR spectroscopy: The central helix is flexible. *Biochemistry 31*, 5269−5278.
(71) Grzesiek, S., and Bax, A. (1993) The importance of not saturating H_2O in protein NMR: Application to sensitivity enhancement and NOE measurements. *J. Am. Chem. Soc. 115*, 12593−12594.
(72) Kay, L. E., Torchia, D. A., and Bax, A. (1989) Backbone dynamics of proteins as studied by ^{15}N inverse detected heteronuclear NMR spectroscopy: Application to staphylococcal nuclease. *Biochemistry 28*, 8972−8979.
(73) Goel, A., Tripet, B. P., Tyler, R. C., Nebert, L. D., and Copie, V. (2010) Backbone amide dynamics of apo-L75F-TrpR, a temperature sensitive mutant of the tryptophan repressor protein (TrpR): Comparison with the ^{15}N NMR relaxation profiles of wild type and A77V mutant TrpR apo-repressors. *Biochemistry 49*, 8006−8019.
(74) Tripet, B. P., Goel, A., and Copié, V. (2011) Internal dynamics of the tryptophan repressor (TrpR) and two functionally distinct TrpR variants, L75F-TrpR and A77V-TrpR, in their L-Trp bound forms. *Biochemistry 50*, 5140−5153.
(75) Carr, H. Y., and Purcell, E. M. (1954) Effects of Diffusion on Free Precession in Nuclear Magnetic Resonance Experiments. *Phys. Rev. 94*, 630.
(76) Meiboom, S., and Gill, D. (1958) Modified Spin-Echo Method for Measuring Nuclear Relaxation Times. *Rev. Sci. Instrum. 29*, 688−691.
(77) Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR 6*, 277−293.
(78) Goddard, T. D., and Kneller, D. G. (2004) *SPARKY 3*, University of California, San Fransisco.
(79) Bracken, C., Carr, P. A., Cavanagh, J., and Palmer, A. G. (1999) Temperature dependence of intramolecular dynamics of the basic leucine zipper of GCN4: Implications for the entropy of association with DNA. *J. Mol. Biol. 285*, 2133−2146.
(80) Cole, R., and Loria, J. P. (2003) FAST-model free: A program for rapid automated analysis of solution NMR spin-relaxation data. *J. Biomol. NMR 26*, 203−213.
(81) Palmer, A. (1998) *ModelFree*, version 4.0, http://cpmcnet.columbia.edu/dept/gsas/biochem/labs/palmer.
(82) Lipari, G., and Szabo, A. (1982) Model-Free approach to the interpretation of nuclear magnetic-resonance relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc. 104*, 4546−4559.
(83) Lipari, G., and Szabo, A. (1982) Model-Free approach to the interpretation of nuclear magnetic-resonance relaxation in macromolecules. 2. Analysis of experimental results. *J. Am. Chem. Soc. 104*, 4559−4570.
(84) Schreiter, E. R., and Drennan, C. L. (2007) Ribbon-helix-helix transcription factors: Variations on a theme. *Nat. Rev. Microbiol. 5*, 710−720.
(85) Zhou, Y., Larson, J. D., Bottoms, C. A., Arturo, E. C., Henzl, M. T., Jenkins, J. L., Nix, J. C., Becker, D. F., and Tanner, J. J. (2008) Structural Basis of the Transcriptional Regulation of the Proline Utilization Regulon by Multifunctional PutA. *J. Mol. Biol. 381*, 174−188.
(86) Costa, M., Sola, M., del Solar, G., Eritja, R., Hernandez-Arriaga, A. M., Espinosa, M., Gomis-Ruth, F. X., and Coll, M. (2001) Plasmid Transcriptional Repressor CopG Oligomerises to Render Helical Superstructures Unbound and in Complexes with Oligonucleotides. *J. Mol. Biol. 310*, 403−417.
(87) 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 Arc repressor mutant PL8. *Biochemistry 34*, 1405−1412.
(88) Guilliere, F., Peixeiro, N., Kessler, A., Raynal, B., Desnoues, N., Keller, J., Delepierre, M., Prangishvili, D., Sezonov, G., and Guijarro, J. I. (2009) Structure, function, and targets of the transcriptional regulator SvtR from the hyperthermophilic archaeal virus SIRV1. *J. Biol. Chem. 284*, 22222−22237.
(89) Garvie, C. W., and Phillips, S. E. (2000) Direct and indirect readout in mutant Met repressor-operator complexes. *Structure 8*, 905−914.
(90) Mattison, K., Wilbur, J. S., So, M., and Brennan, R. G. (2006) Structure of FitAB from *Neisseria gonorrhoeae* bound to DNA reveals a tetramer of toxin-antitoxin heterodimers containing pin domains and ribbon-helix-helix motifs. *J. Biol. Chem. 281*, 37942−37951.
(91) Burgering, M. J. M., Boelens, R., Gilbert, D. E., Breg, J. N., Knight, K. L., Sauer, R. T., and Kaptein, R. (1994) Solution Structure of Dimeric Mnt Repressor (1−76). *Biochemistry 33*, 15036−15045.
(92) Breg, J. N., van Opheusden, J. H. J., Burgering, M. J. M., Boelens, R., and Kaptein, R. (1990) Structure of Arc represser in solution: Evidence for a family of β-sheet DMA-binding proteins. *Nature 346*, 586−589.
(93) Oberer, M., Zangger, K., Gruber, K., and Keller, W. (2007) The solution structure of ParD, the antidote of the ParDE toxin-antitoxin module, provides the structural basis for DNA and toxin binding. *Protein Sci. 16*, 1676−1688.
(94) Oberer, M., Lindner, H., Glatter, O., Krafky, C., and Keller, W. (1999) Thermodynamic properties and DNA binding of the ParD protein from the broad host-range plasmid RK2/RP4 killing system. *J. Biol. Chem. 380*, 1413−1420.
(95) Schreiter, E. R., Wang, S. C., Zamble, D. B., and Drennan, C. L. (2006) NikR-operator complex structure and the mechanism of repressor activation by metal ions. *Proc. Natl. Acad. Sci. U.S.A. 103*, 13676−13681.
(96) Rasooly, A., and Rasooly, R. S. (1997) How rolling circle plasmids control their copy number. *Trends Microbiol. 5*, 440−446.
(97) Holm, L., and Sander, C. (1993) Protein structure comparison by alignment of distance matrices. *J. Mol. Biol. 233*, 123−138.

(98) Gibrat, J.-F., Madej, T., and Bryant, S. H. (1996) Surprising similarities in structure comparison. *Curr. Opin. Struct. Biol. 6*, 377−385.
(99) Vieille, C., and Zeikus, G. J. (2001) Hyperthermophilic enzymes: Sources, uses, and molecular mechanisms for thermostability. *Microbiol. Mol. Biol. Rev. 65*, 1−43.
(100) Liang, J., Edelsbrunner, H., and Woodward, C. (1998) Anatomy of protein pockets and cavities: Measurement of binding site geometry and implications for ligand design. *Protein Sci. 7*, 1884−1897.
(101) Laskowski, R. A., Luscombe, N. M., Swindells, M. B., and Thornton, J. M. (1996) Protein clefts in molecular recognition and function. *Protein Sci. 5*, 2438−2452.