NMR SOLUTION STRUCTURE AND BIOLOGICAL ACTIVITY OF E73, A DNA BINDING PROTEIN FROM SULFOLOBUS SPINDLE VIRUS RAGGED HILLS

by

Casey James Schlenker

A thesis submitted in partial fulfillment of the requirements for the degree

of

Master of Science

in

Chemistry and Biochemistry

MONTANA STATE UNIVERSITY
Bozeman, Montana

August 2011
This thesis has been read by each member of the thesis committee and has been found to be satisfactory regarding content, English usage, format, citation, bibliographic style, and consistency and is ready for submission to The Graduate School.

Dr. Valérie Copié

Approved for the Department of Chemistry and Biochemistry

Dr. David J. Singel

Approved for The Graduate School

Dr. Carl A. Fox
STATEMENT OF PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a master’s degree at Montana State University, I agree that the Library shall make it available to borrowers under rules of the Library.

If I have indicated my intention to copyright this thesis by including a copyright notice page, copying is allowable only for scholarly purposes, consistent with “fair use” as prescribed in the U.S. Copyright Law. Requests for permission for extended quotation from or reproduction of this thesis in whole or in parts may be granted only by the copyright holder.

Casey James Schlenker

August, 2011
# TABLE OF CONTENTS

1. INTRODUCTION ......................................................................................................................... 1

2. MATERIALS AND METHODS ...................................................................................................... 5
   DNA Binding Experiments - Electromobility Shift Assay ................................................... 19

3. RESULTS ...................................................................................................................................... 20
   Analysis of an Ensemble of 20 Lowest Energy Conformers of E73 ......................... 24
   Structural Homologs of SSV-RH E73 ................................................................................. 27
      Modeling of E73 Structure onto DNA – Based on its Structural Homology to the DNA-Bound PutA Structure ......................................................... 29

4. DISCUSSION ................................................................................................................................. 39

REFERENCES CITED ...................................................................................................................... 45
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. NMR Experimental Parameters</td>
<td>10</td>
</tr>
<tr>
<td>2. E73 Structure Calculation Statistics</td>
<td>18</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SDS PAGE of E73 Ni Column Fractions</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>2D N(^{15})-H(^{1}) HSQC of E73</td>
<td>8</td>
</tr>
<tr>
<td>3</td>
<td>E73 Chemical Shift Index</td>
<td>11</td>
</tr>
<tr>
<td>4</td>
<td>CD Spectrum of Lyophilized E73</td>
<td>12</td>
</tr>
<tr>
<td>5</td>
<td>E73 Secondary Structure Parameters</td>
<td>14</td>
</tr>
<tr>
<td>6</td>
<td>Structure of E73</td>
<td>22</td>
</tr>
<tr>
<td>7</td>
<td>Hydrophobic Residues of E73</td>
<td>23</td>
</tr>
<tr>
<td>8</td>
<td>Ensemble of 20 E73 Structures</td>
<td>25</td>
</tr>
<tr>
<td>9</td>
<td>Ramachandran Plot of E73 Structure</td>
<td>26</td>
</tr>
<tr>
<td>10</td>
<td>E73 PutA Structural Alignment</td>
<td>28</td>
</tr>
<tr>
<td>11</td>
<td>Structure Based Sequence Alignment</td>
<td>30</td>
</tr>
<tr>
<td>12</td>
<td>E73 Superpositional Docking to PutA DNA</td>
<td>32</td>
</tr>
<tr>
<td>13</td>
<td>Electrostatic Potential of β Sheets</td>
<td>34</td>
</tr>
<tr>
<td>14</td>
<td>Electrostatic Potential of Top Cleft</td>
<td>35</td>
</tr>
<tr>
<td>15</td>
<td>EMSA Gel</td>
<td>38</td>
</tr>
</tbody>
</table>
Sulfolobus solfataricus, a model organism for Archaea, lives in extreme thermal and acidic environments such as the hot springs of Yellowstone National Park, and is host to diverse archaeal viruses including Sulfolobus spindle shaped virus-1 (SSV1) and Sulfolobus spindle shaped virus-Ragged Hills (SSV-RH). SSV viruses exhibit remarkable morphology and genetic diversity, but are poorly understood as many proteins encoded by their genomes have very little sequence homology to proteins of known functions. Detailed structure-function studies have been undertaken to better understand the role played by SSV proteins in regulating viral gene expression, viral life cycle, and in mediating virus-host interactions. Herein, we report the 3D solution structure of E73, a 73-residue, homodimeric protein encoded within the SSV-RH genome and demonstrate its dsDNA binding capabilities. We find that E73 is comprised of an extended ribbon-helix-helix (RHH) structural domain, which is structurally homologous to the RHH domains of numerous proteins involved in regulation of gene transcription. The N-terminal β strands of E73’s protomers assemble into an anti-parallel β-sheet, which, based on structural homology with other RHH proteins, form base-specific interactions with dsDNA. E73 is notably distinct from known RHH proteins however as it contains a third helix which forms a positively charged structural cleft that we postulate is involved in protein-protein interactions. These findings are discussed in the context of E73’s potential role in regulating SSV-RH genome transcription and replication in its Sulfolobus host.
CHAPTER 1

INTRODUCTION

The hyperthermophilic Crenarchaeum *Sulfolobus solfataricus* is host to a large group of archaeal viruses, plasmids, and virus-like particles. Genetic elements characterized to date reveal notable diversity in the genomic content and the morphology of these viruses [1]. These viruses are found to thrive in extreme thermal and acidic environments such as the hot thermal springs of Yellowstone National Park, and based on their molecular diversity, have been classified into seven new viral families [2-4]. One of the best studied, and representing some of the abundant crenarchaeal viruses is the *Fuselloviridae* viral family. *Fuselloviridae* have been isolated from acidic (pH < 4) hot (Temp > 70 °C) springs throughout the world. Their genetic and morphological diversity is reflected in their gene products, as only a small fraction of the open reading frames (ORFs) encoded by these viral genomes displays significant sequence similarity to functionally well characterized proteins [5]. Representatives of the *Fuselloviridae* include *Sulfolobus* spindle shaped virus -1 (SSV-1) and *Sulfolobus* spindle shaped virus-Ragged Hills (SSV-RH) [6, 7].

SSV-1 has been extensively characterized, and morphologically resembles a lemon-shaped virion, 60 x 90 nm in size, with tail fibers emanating from one end, and typically includes a 15.5 kb circular double stranded DNA (dsDNA) genome [8]. The genome sequence of SSV-1, the first to be completely annotated among *Fuselloviridae* [9], was found to encode 34 open reading frames (ORFs) [6]. SSV-RH, genetically
distinct, but a close relative of SSV-1, is comprised of a 16.5kbp genome, which codes for 37 ORFs. Roughly half of the SSVRH ORFs are homologous to those encoded in SSV-1 genome [7]. The names of the ORFs in the genomic sequence, and corresponding protein in these viruses are derived from one of 6 reading frames (A-F), and number of encoded amino acids. Thus, ORF E73 in SSV-RH encodes a 73 amino acid gene product from reading frame E [7].

In spite of significant progress in annotating potential functions to crenarchaeal viral proteins, 26 of 34 of SSV-1 ORFs have not yet been reliably identified, and have required a structural genomics approach to provides a greater understanding of crenarchaeal protein structures and functions encoded in crenarchaeal viral genomes [5]. Sequence conservation is often much weaker than conservation of structural elements, thus supporting the notion that structural characterizations of crenarchaeal viral proteins using X-ray crystallography and NMR can provide functional and evolutionary knowledge about proteins that is not accessible by a simple inspections of protein amino acid sequences [10]. To this end, the three dimensional (3D) structure of D63 determined by X-ray crystallography [11] has shown that D63 is structurally homologous to ROP (repressor of primer), an adaptor protein that suppresses ColE1 plasmid replication and thus regulates plasmid copy number in *E. coli* [12]. These data suggest that D63 may play a similar role in regulation of crenarchaeal viral genome replication. Structural and functional characterization of F93 and F112 has revealed that these proteins are comprised of a winged-helix-turn-helix motif, suggesting that they function as transcription factors [13, 14]. Similarly D212, a protein from *Sulfolobus* Spindle-shaped
Virus-Ragged Hills (SSV-RH) and a homolog of SSV-1’s D244, exhibits close structural similarity to members of the nuclease family of enzyme, including the nuclease component of the Holliday Junction Cleavage complex [15].

The ORF E73 from SSV-RH (i.e. SSV-RH E73) is homologous to one of the very early T5 transcript of SSV-1 encoding the protein E51 [16]. SSV-RH E73 displays remarkably high amino acid sequence conservation among several SSV strains, but little sequence homology with proteins from other organisms, suggesting that its function in viral life cycle or host gene regulation is important. A sequence homology search using the Basic Local Alignment Search Tool (BLAST) algorithm indicates that E73 is conserved across six (of total eight) known SSV variants, including SSV1 isolated from Sulfolobus shibate, Beppu, Japan [9, 17]; SSV2 found to infect Sulfolobus islandicus, Reykjaness, Iceland and SSV4 from Arnautn, Iceland (Redder); SSV5 isolated from Sulfolobus islandicus strains found on the Kamchatka peninsula, Russia [8]; SSV6 from Sulfolobus strains found in Hveregedi, Iceland [18] and SSV-RH isolated from Sulfolobus sulfataricus in the Norris Geyser Basin in Yellowstone National Park, USA [7]. The amino acid sequence of E73 from SSV-RH is 82% identical to that of E73 homologs found in five of the SSV strains mentioned above but has very little sequence homology to other non Sulfolobus viral proteins. E73’s sequence homology to E51, a syntenic homologue in the sixth strain, SSV1, is lower with only 31% strict amino acid conservation between the two proteins [8]. However, this degree of sequence conservation is still significant, as other proteins encoded by the 34 to 37 ORFs of eight known SSV members of the Fuselloviridae crenarchaeal viral family display sequence
conservation as low as 11% [7]. E73 also bears sequence homology to A59, a recently annotated protein encoded in the genome of *Acidianus* spindle-shaped virus 1 (ASV1) [8]. The exceptionally high conservation and the early expression of the T5 transcript encoding the E73 gene suggest that SSV-RH E73 and its homologues in other *Sulfolobus* strains are key regulators of the temporally controlled SSV transcription / replication mechanism.

To gain a better understanding of the functional role of E73 in SSV genome expression and/or virus/host gene transcription regulation, structural and *in vitro* DNA binding studies have been undertaken on this protein. The three-dimensional (3D) solution structure of SSV-RH E73 has been solved using multidimensional heteronuclear (*^1^H, ^1^5^N, ^1^3^C*) nuclear magnetic resonance (NMR) spectroscopy. The results discussed herein indicate that E73 adopts a homodimeric ribbon-helix-helix (RHH) fold in solution as each E73 protomer is comprised of an N-terminal extended β-strand, followed in sequence by three α-helices. The structural results have suggested that like numerous RHH proteins, E73 binds dsDNA. The ability of SSV-E73 to bind DNA *in-vitro* was confirmed by electromobility shift assays (EMSA). Surprisingly, in addition to a conventional RHH motif, the 3D structure of SSVRH E73 presents a 3rd α-helix, a structural feature which is not commonly found in RHH domains. Based on structural homology with other RHH protein containing a third helix, it is possible this addition to the conventional RHH domain mediates protein-protein interactions, and suggests that E73 may be a bifunctional DNA-binding, and protein-binding protein.
CHAPTER 2

MATERIALS AND METHODS

The SSVRH E73 pDEST-14 clone used to recombinantly produce SSVRH E73 in *E. coli* was provided by Smita Menon of the C. Martin Lawrence lab, at Montana State University. As a result of the cloning, the recombinant E73 protein investigated contains a non-native Met-(His)x6 N-terminal extension linked to the native Met1-Val2-Glu3- sequence of E73-SSVRH. To express recombinant SSVRH E73, pDEST14-E73 (Invitrogen) was transformed into BL21(DE3) *Escherichia coli* cells (Stratagene) by electroporation. Electroporated cells were incubated at 37°C for one hour in 1 mL of Luria Bertani medium, and 100 μL of this inoculated broth was plated onto a Luria Bertani agar plate containing 100 μg/mL ampicillin. Plated cells were incubated at 37°C for 16 hours. A single well-isolated colony was used to inoculate 100 mL of MDG non-inducing media (Studier ref), with 100 μg/mL ampicillin, and incubated at 37°C while shaking at 250 rpm for 12 hours. In order to produce ≥99% isotopically enriched protein samples for NMR structural analysis, a 275 mL auto-induction culture of N-5052 media (for N15 isotope labeling) or C-750501 media (for C13/N15 isotope labeling), containing 100 μg/mL ampicillin, was inoculated with 275 μL of the MDG seed-culture. Auto-induction cultures were incubated, while shaking at 250 rpm, in 2.8 L baffled Fernbach flasks (Fisher Scientific), for 22 hours.
After incubation, cells were pelleted by centrifugation at 3,000 rpm (Sorvall RC2-B, and immediately resuspended in lysis buffer (10 mM Tris, pH 8.0, 5 mM imidazole, 300 mM NaCl, and 0.1 mM phenylmethylsulfonyl fluoride) by manual shaking. Suspended cells were lysed using an M-110L Microfluidizer (Microfluidics, Newton, MA). The resulting lysate was incubated at 65°C to denature non-thermophilic E. coli proteins, and centrifuged to pellet insoluble material. The supernatant, containing soluble E73, was applied by gravity flow to a column containing a 5 mL bed volume of HIS-Select™ Nickel Affinity Gel (Sigma-Aldrich). The column was washed with 50 mL of wash buffer (10 mM Tris, pH 8.0, 10 mM imidazole, 300 mM NaCl). E73 was then eluted in 10 mM Tris, pH 8.0, 200 mM imidazole, 50 mM NaCl (Figure 1).

Figure 1. SDS-PAGE of elution fractions of E73 from a 5 mL Nickel-NTA column. Ni column elution fractions are labeled numerically. The molecular masses of Each fraction consists of 5 mL of elution volume. Fractions 2-5 were pooled, and protein yield was determined by Bradford Assay (Bio-Rad). These fractions were then further purified by Size Exclusion Chromatography, dialyzed into NMR buffer (50 mM KPO₄). The molecular weight of the E73 construct (including an N-terminal Methionine + 6x Histidine tag) is 9578.2 daltons per protomer.
Following elution, fractions were assayed by SDS-PAGE to determine presence of E73 protein, and concentration was determined by Bradford Assay. Fractions with E73 present were pooled, concentrated, and loaded on a calibrated Superdex 75 Size Exclusion Chromatography column (GE Healthcare).

The size exclusion profile of E73 was consistent with a molecular mass of ~20,000 daltons. As the calculated, theoretical molecular mass for the recombinant SSVRH-E73 produced is 9758 Da, further biophysical characterization was performed to determine the oligomeric stated of SSV-RH in solution under NMR conditions. Analytical ultracentrifugation indicated a molecule with a mass of ~19,000 Da, with no evidence of a monomer-dimer equilibrium. Dynamic Light Scattering was used to determine that E73 had a particle size of ~3.8 nm, which is also consistent with a homodimer.

In addition to biophysical characterization, the resonance linewidths of amide $^1$H in the 2D $^1$H-$^{15}$N HSQC correlation spectrum were estimated using the spectral analysis program Sparky. Amide $^1$H linewidths were determined to be ~16Hz, consistent with a molecular correlation time for a protein with a mass of ~20,000 Da. If E73 were a monomer under solution NMR conditions, it would have a total correlation time leading to amide $^1$H linewidths of ~8Hz. In addition to having spectral linewidths corresponding to the molecular mass of a dimer, only one distinct set of peaks was observed in the 2D $^1$H-$^{15}$N HSQC correlation spectrum,
implying that E73 is a symmetric homodimer (Figure 2).

Figure 2. Two dimensional N\textsuperscript{15}-H\textsuperscript{1} correlation HSQC of C\textsuperscript{13}/N\textsuperscript{15} isotopically enriched E73. This spectrum was acquired at 312K, and 600 MHz (H\textsuperscript{1} Larmour frequency) magnetic field strength. All backbone NH resonances of E73 were successfully assigned, excluding that of M1. Unlabeled doublet peaks connected by a thin, black line are glutamine and asparagine side-chain NH\textsubscript{2} groups. A resonance signal for the indole NH of W46 is observed at 11.2 ppm, but is not shown.

To obtain chemical shift resonance assignment of each \textsuperscript{1}H, \textsuperscript{13}C, and \textsuperscript{15}N atom of E73, a standard set of 3D heteronuclear NMR experiments was performed. All NMR data were collected at a temperature of 312K on a four-channel Bruker DRX-600 spectrometer, with an inverse triple (\textsuperscript{15}N, \textsuperscript{13}C, \textsuperscript{1}H) resonance probe equipped with triple axis gradients. The States-TPPI method of Quadrature detection was
used for all multidimensional NMR experiments. [19]. Data were processed and
analyzed using the NMRPipe [20], Sparky [21], and TopSpin™ (Bruker Inc., Billerica,
MA) software. Two-dimensional $^{1}H$-$^{15}N$ HSQC spectra [22] were acquired with
spectral windows of 13.0 ppm in $t_2$ and 30.0 ppm in $t_1$, with the proton carrier
frequency set at 4.64 ppm and the nitrogen carrier set at 120.0 ppm. Data were
collected with 512 complex points in $t_2$ and 128 complex points in $t_1$, using WALTZ-
16 [23] for $^{15}N$ decoupling during data acquisition. Apodization was performed
using a sine bell squared function offset by 0.35 π radians and ending at 0.98 π
radians in $t_2$, and a sine bell function offset by 0.40 π radians and ending at 0.99 π
radians in $t_1$.

The chemical shifts of $^{1}H$, $^{15}N$, and $^{13}C$ backbone, and sidechain atoms were
assigned by determining sequential and intra residue resonances from a series of
double and triple resonance NMR experiments including HNCA [24], HNCACB [25],
CBCA(CO)NH [26], C(CO)NH [27], HBHA(CO)NH [28], HCC(CO)NH [27], $^{1}H$-$^{13}C$-CT
HSQC [29] and HCCH-TOCSY [30] as reported in [31]. $^{1}H$-$^{13}C$-CT HSQC and HCCH-
TOCSY experiments were performed in D$_2$O. A DIPSI pulse sequence scheme [32]
was utilized for $^{1}H$ decoupling during carbon evolution and the WALTZ-16 scheme
for $^{15}N$ decoupling during data acquisition. Apodization in all spectral dimensions
was performed using similarly shifted sine bell functions. $^{1}H$, $^{13}C$, and $^{15}N$, chemical
shifts were indirectly referenced to DSS, using absolute frequency ratios for $^{13}C$ and
$^{15}N$ chemical shift referencing. A list of all acquisition parameters for all NMR
experiments recorded on E73 is included in Table T1.
Table 1. A table of parameters for NMR experiments carried out on SSVRH E73 for chemical shift resonance assignment, and $^1$H-$^1$H nOe determination. All NMR data were acquired on a 1mM (dimer concentration) sample of SSVRH E73 at 312K.

Upon complete assignment of backbone and sidechain chemical shift resonances, secondary structure elements of SSVRH E73 were determined using two methods. The first method utilized the program Chemical Shift Index (CSI)[33], which determines secondary structure elements by analyzing $\mathrm{H}\alpha$, $\mathrm{C}\alpha$, and $\mathrm{C}\beta$ chemical shift deviations from random coil values (Figure 3). The second method was to predict the dihedral angles for each residue using the program TALOS+[34]. TALOS+ uses the deviation of $\mathrm{HN}$, $\mathrm{H}\alpha$, $\mathrm{C}\alpha$, $\mathrm{C}\beta$, $\mathrm{CO}$, and $\mathrm{N}$ chemical shift resonances from random coil values to empirically predict phi and psi torsion angles for each residue. Reliable dihedral angles were calculated for 61 residues of SSVRH E73, these dihedral angle values were used as restraints for structural calculation by the software program Crystallography and NMR System (CNS) [35].
In order to determine which backbone $^1$H-$^{15}$N groups engage in hydrogen bonding, $^1$H/$^2$H exchange experiments were carried out to monitor the disappearance of amide proton resonances in 2D $^1$H-$^{15}$N HSQC spectra upon exchange in D$_2$O. To perform the $^1$H/$^2$H exchange experiments, 500 μL of a 1mM sample of $^{15}$N-labeled SSVRH E73 was lyophilized and reconstituted with 500 μL of D$_2$O. To determine whether lyophilization would cause denaturation of the sample, a fresh, unlabeled sample of SSVRH E73 was analyzed by circular dichroism.
spectroscopy (CD). This sample was then lyophilized, reconstituted, and evaluated by CD once more to show that secondary structural characteristics of the protein were not affected (Figure 4).

Figure 4. The CD spectra of SSVRH E73, before and after lyophilization. Open circles represent the CD spectrum of the protein prior to lyophilization, and filled circles are the spectrum immediately after reconstitution. Evident from these CD spectra are the helicity of E73, as predicted by the CSI calculation, and the ability of the E73 to withstand lyophilization without notable loss of structure.
Following CD spectroscopic analysis of the effect of lyophilization on SSVRH E73, a 2D $^1$H-$^{15}$N HSQC spectrum was taken of a fresh $^{15}$N-labeled sample of SSVRH E73. This $^{15}$N-labeled sample was then lyophilized and reconstituted in H$_2$O, and allowed to equilibrate overnight. A 2D $^1$H-$^{15}$N HSQC spectrum was then recorded, and compared to the spectrum of the sample prior to lyophilization as a control to demonstrate that lyophilization had no adverse effect on the sample integrity or NMR spectral quality.

Once chemical shift resonance assignment of the backbone and sidechain atoms of SSVRH E73 was complete, 3D $^{15}$N-edited $^1$H-$^1$H-NOESY [36, 37] and 3D $^{13}$C-edited $^1$H-$^1$H-NOESY [38] experiments were carried out with nOe mixing times of 80ms, 100ms, and 120ms (for $^{15}$N-edited NOESY) and 100, 120ms, and 150ms (for $^{13}$C-edited NOESY) (See Table 1). $^{15}$N- and $^{13}$C-edited NOESY spectra recorded with 100ms nOe mixing periods were used for nOe assignments and to derive $^1$H-$^1$H distance estimates that were used as input NMR data in the structure calculations. Both $^{13}$C- and $^{15}$N- edited NOESY spectra were processed using NMRPipe, and analyzed and assigned using Sparky. The intensities of assigned $^1$H-$^1$H-nOe cross peaks were measured in Sparky, and designated strong, medium, and weak in order to determine a restraint value range for the interproton distance associated with that nOe crosspeak. Strong nOe intensities were assigned a distance restraint of 1.8 – 3.0 Å, medium nOe intensities were assigned a distance restraint of 1.8 -3.5 Å, and weak nOe intensities were assigned a distance restraint of 1.8 -5.0 Å. Based on the secondary structural elements calculated using CSI, an initial set of sequential (i-
i+1) and short-range (≤ i-i+3) nOe cross peaks was identified (Figure 5). This set of sequential and short-range nOe peaks was used to find a corresponding set of nOe-based interproton distance restraint values that, along with hydrogen bond distance restraint values determined for backbone amides protected from $^2$H/$^1$H exchange and dihedral angle restraints calculated from TALOS+, experimentally defined the secondary structural elements of SSVRH E73 (Figure 5).

Figure 5. SSVRH E73 sequence is shown. Below the sequence is a graphical representation of the H/D exchange protection pattern for the backbone amide protons, short-range NOE patterns, and $^{13}$C$_\alpha$, $^{13}$C$_\beta$ and $^1$H$_\alpha$ chemical shift deviations from random coil according to the method of the Chemical Shift Index (CSI) [39], corresponding to the secondary structure elements of E73 represented in the last line. The heights of the boxes for the (i,i+1) and the thickness of the lines for the (i,i+2), (i,i+3), and (i,i+4) NOE patterns indicates the intensity of the corresponding peak, which was used to derive that interproton distance.
To gain insight into possible inter-protomer nOes with which to build initial structure models, a predicted model of E73 RHH 3D fold was used to generate a potential set of inter-protomer nOes. From this initial analysis, 15 inter-protomer $^1$H-$^1$H nOes were assigned in the 3D NOESY data orienting the two β-strands into an antiparallel β-sheet, and 8 inter-protomer $^1$H-$^1$H nOes were identified orienting the αB helix in close proximity of the αB’ helix. Initial structural calculations were accomplished using the Crystallography and NMR System 1.2 (CNS 1.2) software [35] using manually assigned inter-protomer, sequential, and short range nOes; hydrogen bond distance restraints obtained from identification of 25 hydrogen-bonded amides (i.e. 25 per protomer) from $^1$H/$^2$H solvent exchange experiments; and 61 dihedral angle restraints (i.e. 61 per protomer) generated by TALOS+ analysis of chemical shift data.

A molecular topology file of the SSVRH E73 dimer was built from primary structure information and used as template for torsion angle molecular dynamics in CNS. Ten lowest energy structures with no nOe violations (< 0.5 Å) were selected from a total of fifty starting structure calculated with CNS and were used as models for the next round of manual nOe assignments. Using this approach, a total of 527 $^1$H-$^1$H nOe distance restraints were assigned manually.

To obtain the final set of nOe distance restraints, the software program Ambiguous Restraints for Iterative Assignment v2.2 (ARIA) was utilized [40]. ARIA structure calculations were carried out using a representative structure from the 20 lowest energy structures calculated using the 527 manually assigned nOes as a
template, and the 527 manually assigned nOes were treated as unambiguous. ARIA calculations proceeded in nine iterations of NOESY spectra calibration and assignment of ambiguous and unambiguous distance restraints followed by calculation of an ensemble of structures. For the initial run of ARIA, the remaining set of unassigned nOe crosspeaks was used as a set of ambiguous distance restraints. From this initial calculation, an additional 331 ambiguous nOe peaks were assigned automatically, for a total of 858. Within the nOe assignment protocol, unassigned nOe cross-peaks were matched by the ARIA software to each possible chemical shift within a frequency tolerance range of ± 0.04 ppm for protons and ± 0.40 ppm for nitrogen or carbon heteronuclei, generating several potential intra or inter-protomer nOes. The ARIA program treated these nOes as ambiguous distance restraints [40]. In the structure calculation step of each iteration, ARIA invoked the CNS 1.2 software [35], using as input data automatically and manually assigned interproton ¹H-¹H-nOe data, hydrogen bond distance restraints, dihedral angle restraints obtained with TALOS+ [34, 41], and the initial structural model calculated using CNS as described above.

After each run of ARIA, the automatically assigned nOe peaks were evaluated manually to ensure they did not arise from spectral artifacts, and for proper calibration. Following the automated assignment of the majority of the remaining nOe peaks, small sets (10 or fewer) of likely ambiguous peaks were identified by manual spectral analysis, and used in iterative runs of the ARIA program. After each ARIA calculation, automatically assigned nOe distance restraints were analyzed
manually and used in subsequent manual structure calculations using CNS. During these steps, large violations were removed (those yielding energies of roughly > 100 kcal/mol), and the ARIA calculations repeated using NMR assignments generated from previous ARIA iterations until convergence was achieved. Using this method, a total of 1070 unambiguous nOe distance restraints were identified. This final set of nOe distance restraints, along with the 25 hydrogen bond distance restraints identified by $^2$H/$^1$H exchange, and 61 dihedral angle restraints determined by TALOS+ was used to calculate final set of 100 structures using CNS (Table 2).

From the final 100, 20 lowest energy structures were selected for analysis using PYMOL [42], PROCHECK-NMR [43], MolProbity [44], and Adaptive Poisson-Boltzmann Solver (APBS) for evaluation of the electrostatic properties of SSVRH E73 [45]. Size and polar/hydrophobic contributions to the dimer interface and changes to in accessible surface area upon dimer formation were calculated using the PISA server [46]
Table 2: Experimental Restraints Used in the Structure Calculations of E73 and Statistical Analysis of 20 Lowest Energy Structures

<table>
<thead>
<tr>
<th>Experimental NMR Restraints Used for Structure Calculations</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOE distance restraints</td>
</tr>
<tr>
<td>Intraresidue</td>
</tr>
<tr>
<td>Sequential (i – i+1)</td>
</tr>
<tr>
<td>Short Range (i-i+3)</td>
</tr>
<tr>
<td>Medium Range (i-i+4)</td>
</tr>
<tr>
<td>Long Range (&gt;i+5)</td>
</tr>
<tr>
<td>Interprotomer</td>
</tr>
<tr>
<td>Total</td>
</tr>
<tr>
<td>H-bond distance restraints</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Structural Statistics for 20 Lowest Energy E73 Structures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Root-Mean-Square Deviations Relative to Average Structure (Å)</td>
</tr>
<tr>
<td>Backbone (Cα, C’, N) atoms in 2nd structure</td>
</tr>
<tr>
<td>Backbone all residues</td>
</tr>
<tr>
<td>Heavy atoms in 2nd structure</td>
</tr>
<tr>
<td>Heavy atoms in all residues</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Restraint Violations</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOE distances with violations &gt; 0.5Å</td>
</tr>
<tr>
<td>Dihedral with violations &gt;5°</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Final Energies from Simulated Annealing (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fvdw</td>
</tr>
<tr>
<td>Fele</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Deviation from Idealized Geometry</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bonds (Å)</td>
</tr>
<tr>
<td>Angles (deg)</td>
</tr>
<tr>
<td>Impropers (deg)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ramachandran Analysis (% of All Residues)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Residues in most favored region</td>
</tr>
<tr>
<td>Residues in additionally allowed region</td>
</tr>
<tr>
<td>Residues in generously allowed regions</td>
</tr>
<tr>
<td>Residues in disallowed regions</td>
</tr>
</tbody>
</table>

*Root-mean-squared deviations (rmsds) were calculated for regions of defined secondary structure for both protomers.*
DNA Binding Experiments - Electromobility Shift Assay

The structure similarity of SSVRH E73 to several members of the Ribbon Helix Helix (RHH) domain of DNA binding proteins indicated that it is capable of binding dsDNA, and may regulate transcription of genes in the SSVRH genome. The ability of SSVRH E73 to bind SSVRH dsDNA was investigated in vitro using 17 ~1,000 bp fragments of SSV-RH that were amplified and generated using Polymerase Chain Reaction PCR.

Primers were designed to produce fragments that overlapped by at least 100bp to ensure complete coverage of SSV-RH genome. The EMSA assays were conducted according to published protocols [47]. Each 1000 bp fragment of DNA was incubated at 65 °C in 20 mM Bis-Tris pH 6.5, and 300 mM KCl, 5% glycerol and 1 mM EDTA, with serial dilutions of E73 from 1 uM to 1pM. The total reaction volume was 20 uL, and the concentration of each 1000bp fragment was 8 nM. After incubation at 65 °C for 20 minutes, 20 μL fractions were separated by electrophoresis on an agarose gel composed of 1.6% agarose, 25 mM histidine, 30 mM MOPS at pH 6.3 with a tank buffer of 50 mM MES, pH 6.5. Samples were run at 50 mV for 2 hours, stained with SYBR Gold (Invitrogen), and imaged using UV light [14].
CHAPTER 3

RESULTS

Of the 70 possible backbone amide resonances, excluding the three prolines in the sequence of SSV-RH E73, 69 residues were assigned utilizing the strategy outlined in Methods. The only possible amide backbone resonance unassigned was that of Met1. Cα/Cβ chemical shifts were assigned for all 73 residues of SSVRH-E73, Hα/Hβ chemical shifts were assigned for all residues except Lys73, and ~95% of sidechain 13C, and 1H resonances were assigned. All resonance assignments made for 1H, 15N, and 13C atoms of SSVRH E73 have been deposited in the Biological Magnetic Resonance Data Bank (BMRB), under the accession number 16,177 [31].

Upon complete assignment of backbone atom chemical shift resonances of SSVRH E73, the Cα, Cβ, and Hα chemical shift deviations from random coil values were used to predict regions of secondary structure. Short range and sequential 1H-1H nOes, and amides protected from hydrogen/deuterium exchange, were then used as distance restraint values to calculate secondary structure elements using the software package Crystallography and NMR System (CNS) [35]. Each E73 protomer is comprised of a short extended β-strand spanning residues Lys10-Phe16 followed by three α-helices αA, αB and αC, spanning residues Glu18-Leu31, Met35-Lys53, and Lys61-Phe69, respectively. Helix αC is capped at its N-terminal end by Pro 60, while the N- (Met1-Ile 7) and C-termini (Phe70-Lys73) of the protein are disordered as judged by observed chemical shifts very close to random coil values, lack of
significant $^1$H-$^1$H nOes, and absence of amides protected from H/D exchange for these segments of the protein (see Figure 5).

The overall three-dimensional fold of E73 consists of a tightly intertwined dimer with the N-terminal $\beta$-strands from each E73 protomer interacting to form a short antiparallel $\beta$-sheet. This antiparallel $\beta$-sheet packs together with helices $\alpha_A$ and $\alpha_B$ (and $\alpha'_A$ and $\alpha'_B$ of E73’s second protomer), to form a canonical ribbon-helix-helix (RHH) domain (Figure 6). The C2 symmetry axis of the molecule is perpendicular to the plane formed by the antiparallel $\beta$-sheet. A sharp turn at Asp17 connects E73’s $\beta$-strands to helix $\alpha_A$, which extends between the C-terminus of helix $\alpha'_B$ and the N-terminus of helix $\alpha_B$ and forms hydrophobic contacts with helix $\alpha'_B$. Between helices $\alpha_A$ and $\alpha_B$ is a short, solvent exposed loop (L1) consisting primarily of charged and polar amino acid residues. Leu31 and Ile38 serve to form a network of hydrophobic contacts between loop L1, and helices $\alpha_A$, $\alpha_B$, and $\alpha'_B$. (Figure 6)
Figure 6. The three-dimensional fold of SSVRH E73 contains a core homodimeric Ribbon Helix Helix (RHH) domain comprised of the N-terminal β sheet followed by the A and B helices. Each separate protomer in the RHH region is colored blue and grey respectively. In addition to the canonical RHH domain, E73 presents the interesting inclusion of a C helix, highlighted in yellow for both protomers.

The dimer interface of SSVRH E73 is predominated by the two αB helices forming a core, with the anti-parallel β sheet, αA and αC helices wrapping around this core. Hydrophobic residues Val43, and Trp46 in the middle of each of the αB helices form hydrophobic contacts to pack these two helices tightly against one and other. A large majority of the middle residues of αB helices is hydrophobic (Ala42, Val43, Val44, Met45, Trp46, Leu47, Ile 48) and forms the central hydrophobic core of the molecule, leading to the highly entwined orientation of this protein. Phe16 on the C-terminal end of the β strands, Leu24 in the middle and Leu31 on the C-
terminal end of the \( \alpha_A \) helices, and Ile38 on the N-terminus of the \( \alpha_B \) helices also form key hydrophobic interactions in the core RHH fold of SSVRH E73. A turn at loop L2 orients the \( \alpha_C \) helices to clamp the core of the molecule formed by the crossing of the \( \alpha_B \) helices. Hydrophobic interactions between Phe69 and Phe 70 on the C-terminal end of the \( \alpha_C \) helices, and Val44 in the middle of the \( \alpha_B' \) helices orients the \( \alpha_C \) helices with respect to the core RHH domain, and complete the dimer interface (Figure 7). The dimerization of the two protomers of SSVRH E73 buries approximately 1050 Å\(^2\) per protomer.

Figure 7. A structural representation of SSVRH E73, rotated 90° toward the reader as compared to Figure 6, to look down the C2 symmetry axis and show hydrophobic contacts that maintain tertiary and quaternary structure. Both protomers are colored grey, and positions with a hydrophobic amino acid residue are colored blue.
The αC helices are a notable addition to the RHH domain of E73, as standalone members of the RHH superfamily rarely have additions to the core RHH fold [48]. In the 3D structure of SSVRH E73, these additional 3rd αC helices, and the loops L2 creates a cleft distal to the antiparallel β-sheet of E73’s RHH domain. The loop L2 bridging helices αB and αC creates the upper ridges of this cleft that is completed by the αC helix of each protomer entwining the αB’ helix of E73’s second subunit (Figure 7). The loop between helices αB and αC is largely solvent exposed with the exception of Leu54 whose sidechain forms hydrophobic contacts with Leu58 of helix αC’ and with Leu31 of helix αA’ (Figure 7). This structural element has not yet been observed in the 3D structure of any other RHH proteins. There are examples of RHH domain proteins with αC helices, however none in structurally analogous orientations to those of SSVRH E73.

Analysis of an Ensemble of 20 Lowest Energy Conformers of E73

To obtain the final ensemble of structures for SSVRH E73, 100 total structures were calculated using the software program Crystallography and NMR System (CNS) [35]. These structures were calculated based on experimentally determined 1H-1H-nOe signal intensities, to obtain interproton distance restraints, as well as hydrogen bond patterns inferred from H/D solvent exchange experiments, and φ and ψ torsion angle information obtained from chemical shift data analyzed with the programs Chemical Shift Index [33] and TALOS + [34]. From the total 100 structures calculated in CNS, 20 structures with the lowest energy, no
nOe distance violations greater than 0.5 Å, and favorable Ramachandran angles were selected. These structures were aligned in Pymol [42] based on the coordinates of the backbone N, Cα, and C' atoms from residues in regions of well defined secondary structure, the N-terminal β-strand (residues 10-16), helix αC (residues 18-31), helix αB (residues 35-53), and helix αC (residues 61-69) (Figure 8)

Figure 8. A stereo image of an ensemble of the 20 lowest energy SSVRH E73 dimer structures represented as a trace of the N, Cα, and C' atoms of the polypeptide backbone. The orientation of this figure is the same as that of Figure 6. 100 total structures were calculated, based on distance restraints determined from the NMR data, using the program Crystallography and NMR System (CNS), version 1.2 [35] This alignment was carried out using the molecular imaging software PyMOL [42] by aligning the N, Cα, and C’ atoms of residues 10-16, 18-32, 35-51, and 61-68. The positions of residues in these regions are well defined by short, medium, long-range, and interprotomer nOe distance restraints.

The RMSD, relative to the mean structure, for the lowest energy ensemble of the backbone atoms (N, Cα, and C’) of 20 SSVRH E73 structures is 0.45 Å for residues located in well-defined structural regions. Evident in Figure 8 is the close alignment of the backbone traces corresponding to the β strands and α helices of the protein as well as the well-defined position of the short loop spanning helices αA and αB. The N- and C-terminal segments of the proteins are not as well defined as
seen by the weaker agreement in the overlay of backbone atoms for these regions. This relative disorder is to be expected as no short, medium, long-range, or interprotomer nOe based distance restraints were observed for residues 1-6, as well as residues 70-73.

To examine the distribution of \( \Phi/\Psi \) dihedral angles in the final set of structures calculated for SSVRH E73, the web server PDBsum, which is based on the PROCHECK-NMR program was used [49]. No dihedral angle for any of the lowest energy structures of SSV RH E73 was found to be in a disallowed region, while 88% and 12% of \( \phi/\psi \) angles were present in the most favorable and generously allowed regions of the Ramachandran plot, respectively (Figure 9).

Figure 9. A plot of the Ramachandran angles of one representative structure of SSVRH E73. 88% of the Ramachandran angles of SSVRH E73 are found in most favorable orientation, and 12% are found in additionally allowed positions. There are no residues with dihedral angles disallowed positions.
Structural Homologs of SSV-RH E73

To search the Protein Data Bank (PDB) for structural homologues of SSVRH E73, the DALI [50] and SSM [51] web servers were used. All structural homologues to SSVRH E73 identified in the PDB consist of homodimeric proteins with one or more RHH motifs and aligned well for ~ 40 residues of the total 73 of each protomer of SSVRH E73, from residue 10 to residue 53. The structurally aligning residue positions corresponded to the β strand and the αA and αB segments of each SSV RH E73 protomer. The RHH domain of E73 is most structurally homologous (in descending order) to the structures of the RHH domains of PutA (proline utilization A) of *E. Coli* in its DNA-bound (PDB ID 2RBF [52]) and DNA-free forms (PDB ID 2GPE [53]), SvtR (PDB ID 2KEL, [54]), DNA-bound and DNA-free CopG (PDB code 1EA4 [55] and 1B01 [56], respectively), and the RHH domain of NikR (PDB code 2WVB) [57].

The superposed structures of SSVRH E73 (black) and the RHH domain of DNA-bound PutA (magenta) are shown in figure 10. As can be seen in figure 10, the structural alignment of E73 with DNA-bound PutA is quite satisfactory for secondary structural elements forming the core of protein RHH domains, yielding a DALI Z-score of 5.1, Cα root-mean square deviation (rmsd) of 1.38 Å, and a sequence identity of 16% over 44 structurally identical residues. E73 alignment with the structure of free PutA is not quite as good, with a Z-score 5.3, Cα rmsd of 1.56 Å and a sequence identity of 15% over 48 structurally identical residues. Notably, PutA has no structural elements beyond residue 52 that align well with the 3rd helix αC of E73.
Figure 10. The structure of SSVRH E73 aligned with that of its closest structural homologue, the RHH domain of PutA (PDB ID 2RBF). SSVRH is shown in black, and the RHH domain of PutA is shown in magenta. The alignment was performed using the Least Squares method developed by Kabsch [58], according to structurally conserved residue positions obtained by a DALI [50] search of the PDB. As can be seen, the alignment is quite good, with a RMSD of 1.38 Å.

The structural alignments of E73 with both the DNA and non-DNA bound structures of CopG yield DALI Z-scores of 4.9 and 4.8 respectively, and a Cα rmsd of 1.40 Å for both forms of CopG, and 19% sequence identity over 43 structurally conserved residue. These numbers do not change in the alignment of SSVRH E73 with the DNA bound and DNA free forms of CopG because CopG is known to undergo very little structural change upon binding of its DNA operator sequence [59]. E73 RHH domain also aligns well (Cα rmsd of 1.68 Å) with the RHH domain of
non-DNA bound NikR (PDB code 2WVB) from *H. pylori*. The majority of RHH domain proteins in the literature function as transcriptional repressors, as is the case for PutA and CopG, but a few, such as NikR from *H. pylori*, act as transcriptional activators.

Several RHH domain proteins in the Protein Data Bank present a third helix in addition to the core RHH fold. The functions for these molecules fall into one of three general categories. RHH proteins like FitA and ParD function as the antidote in a toxin/antitoxin pair, with the 3rd αC helices serving to bind to the toxin. The 3rd helices of the Mnt repressor form a tetramerization domain. While molecules like MetJ have a C-terminal domain with set of 3rd helices the bind a cofactor. Of the structures in the Protein Data Bank that contain a RHH domain with a C-terminal set of helices, none have residues in structurally homologous positions to the αC helices of SSVRH E73.

Modeling of E73 Structure onto DNA – Based on its Structural Homology to the DNA-Bound PutA Structure

Using structural homolog information, an amino acid sequence alignment was performed for SSVRH E73 and several RHH proteins whose structures superposed quite closely with that SSVRH E73, as judged by DALI Z-scores above 4.0. The amino acid sequence alignment of these proteins based on structurally equivalent residues as reported by DALI is shown in Figure 11. The three sequence positions highlighted in red correspond to structurally conserved positions known to make DNA base-pair specific interactions in canonical RHH domain proteins.
whose DNA-bound structures have been solved, and are referred to as the two, four, and six positions since they correspond to the 2nd, 4th, and 6th amino acids in the β strand of each RHH protomer [48]. In the SSVRH E73 sequence, these residues correspond to Lys11, T13, and A15, which, in the 3D structure of SSVRH E73 have their side chains extending out from the plane of the anti-parallel β sheet and are poised to form favorable interactions with DNA bases upon positioning of the protein’s β sheet into the major groove of DNA. Other significant conserved amino acid patterns observed in the alignment of structurally equivalent residues are the positions of conserved hydrophobic residues which are highlighted in yellow in Figure 11, and which in E73 correspond to Leu14, Phe16, Leu24, Ile38, Val43, and Leu47.

Figure 11. A structure based sequence alignment of RHH domain proteins from a DALI search of the Protein Data Bank. Amino acid positions represented with a capital letter are in a structurally conserved position with SSVRH E73. The sequences from Mnt repressor, CcdA, 1y9b, MetJ, ParG, and FitA all contain a set of C-terminal third helices. With the exception of the Mnt repressor, none of the RHH domain proteins with a set of third helices have any residues in a structurally conserved position homologous to the αC helices of SSVRH E73. The positions highlighted in red correspond to the residues whose side-chains extend into the major groove of DNA, and make base-pair specific interactions. The positions highlighted in yellow correspond to hydrophobic positions that are important to the RHH domain fold [48].
The conserved positions of these hydrophobic residues are hallmarks of the RHH fold, and provide the driving force for folding and structural integrity for this small structural motif as the stability of RHH is highly dependent on extensive hydrophobic contacts between protein protomers [48]. For E73, the compactness of the protein hydrophobic core and the presence of a canonical RHH 3D fold are supported by the observation of a relatively large number of interproton distance restraints derived from inter-protomer $^1$H-$^1$H nOes compared to the relatively small number on long range intra-protomer nOes (see Table 2), leading to two very intertwined polypeptide chains in the SSVRH E73 dimer. This also supports the notion that the structural stability of E73 is highly dependent on its homodimeric state.

Based on the low RMSD between E73 and the DNA-bound structure of PutA, a structural alignment of SSVRH E73 onto the DNA-bound RHH domain of PutA was used to dock E73 onto the 3D structure of the PutA operator DNA (PDB ID 2RBF) [52] (Figure 12).
Figure 12. A superpositional-docking model of SSVRH E73 on the PutA operator DNA. SSVRH E73 was aligned to the structure of the DNA bound form of PutA based on structurally equivalent residues, as determined from a DALI search, the PutA molecule was then removed to show how SSVRH E73 might orient its antiparallel β sheet into the major groove of DNA. Residues N-terminal to the β sheet of each protomer (Met1-Lys9) have been omitted for clarity.

This model implies that E73’s anti-parallel β sheet can orient into the major groove of DNA and is positioned to interact specifically with DNA bases in the major groove.
This strongly suggests that E73 can interact with DNA in a similar manner to PutA and other RHH DNA-binding proteins. The superpositionally docked model shows that residues Lys11, Thr13, and Ala 15 on the antiparallel β sheet, extend their sidechains into the major groove of DNA and may make base-pair specific interactions with DNA. It is also evident from this model that the backbone amide hydrogens of residues Met35 and Thr36 on the N-terminus of helix αB, as well as the sidechain of Tyr21 on helix αA can potentially make hydrogen bonds with the DNA PO₄ backbone.

The model of SSVRH E73 superpositionally docked to the Put operator DNA shown in Figure 12 is further supported by analysis of the electrostatic potential surface of the SSVRH E73, using the Adaptive Poisson-Boltzmann Solver (APBS) [45]. As seen in Figure 13, the electrostatic potential surface of the protein lining the anti-parallel β-sheet region of the protein is positively charged (Fig 13). This charged surface is comprised of 10 Lys residues (5 from each E73 protomer), 6 of which are part of the two β-strands and 4 are located on the unstructured N-termini of the E73 dimer. The positioning of 4 Lys residues on the N-terminal ends of E73’s β-strands places them just outside of the major groove of DNA in the E73-DNA model presented in Fig 12, positioning the sidechains of Lys 9 (Lys 9’), Lys 6 (Lys 6’) and Lys 5 (Lys 5’) in such a way as to form favorable electrostatic interactions with the negatively charged phosphate backing of the DNA (See Figure 13). This structural arrangement is similar to what has been observed in the 3D structure of the Arc repressor-DNA complex [60]. Of the remaining two Lys residues (Lys 10 and
Lys 11), Lys11 is most likely to engage in base-pair specific interactions with the major of groove DNA as judged from the model of SSVRH E73 docked onto the DNA bound structure of PutA (Figure 12).

Figure 13. The electrostatic potential of the surface of SSVRH E73, oriented with the antiparallel β sheet toward the reader. The surface of the antiparallel β sheet is quite electronegative due to a number of Lys residues (see text). Lys11 is hypothesized to form base pair specific interactions with DNA bases in the major groove upon DNA binding.

As mentioned earlier, one of the most interesting and unusual features of E73 3D structure is the presence of a 3rd helix c, which packs against the RHH core domain forming a positively charged structural cleft distal to the antiparallel β
sheet. The position and surface features of this cleft suggest that this face of the protein is another possible surface for SSV E73 to interact with negatively charged proteins or nucleic acid phosphate backbone chains. The electrostatic surface of the cleft created by the αC helices entwining the RHH domain of SSVRH E73 is generally electronegative due to the presence of residues Lys53, Lys61, and Arg68 on the upper surface of the two αC helices (Figure 14).

Figure 14. The electrostatic potential of the surface of the cleft formed by the αC helices entwining the RHH core of SSVRH E73. This cleft is located distal to the antiparallel β sheet, which is postulated to bind the major groove of DNA. The surface of the cleft is electropositive due to the presence of two Lys residues, Lys53 and Lys 61, and Arg68, on the top surface of each of the αC helices.
To gain further insight into a possible biological function of the ‘top cleft’ formed by the $\alpha_C$ helices, particular attention was given to RHH proteins with a set of C-terminal 3$^{\text{rd}}$ helices. Notably, structural alignments to proteins incorporating the 3$^{\text{rd}}$ helix of E73 were not as close (as judged by RMSDs and DALI Z-scores) as those that considered only the SSVRH E73 RHH domain, indicating that the presence and 3D positioning of these $\alpha_C$ helices is an addition to the RHH domain that has not been described previously (see figure 11).

One of the structural homologs found in this search was the Mnt repressor (PDB ID 1MNT) [61] from Salmonella Bacteriophage P22 yielding a DALI Z-score of 3.6, $\alpha$ rmsd of 3.9 Å, 8% sequence identity over 66 structurally equivalent residues. This form of the Mnt repressor corresponds to a truncated version of the native protein (residues 1-76) lacking 5 C-terminal amino acids, and contains a 3$^{\text{rd}}$ helix (in addition to an RHH domain) that aligns in a structural position somewhat similar to the $\alpha_C$ helices of SSVRH E73, with 10 C-terminal residues in structurally conserved positions (Figure 11). However, unlike SSVRH E73, wild-type Mnt repressor is a tetramer in solution [61] facilitated by inter-dimer interactions of the Mnt repressor 3$^{\text{rd}}$ helices [62]. As stated previously, SSVRH E73 shows no evidence of oligomerization in solution beyond dimer formation, providing evidence that the SSVRH E73 $\alpha_C$ helices serve a different function than that of the 3$^{\text{rd}}$ helices of the Mnt repressor.

Other RHH proteins that present a set of C-terminal 3$^{\text{rd}}$ helices are MetJ from E. coli (PDB ID 1MJM) [63], FitA from N. gonorrhoeae (PDB ID 2BSQ) [64], and RelB
from *E. coli* (PDBID 2K29) [65]. The function of the C-terminal 3rd helices in MetJ involves binding that protein’s cofactor, S-Adenosyl methionine (SAM). Once the cofactor is bound, MetJ acts as a transcriptional repressor of the proteins involved in methionine biosynthesis. The DALI Z-score for the alignment of MetJ and SSVRH E73 is 2.9 with an RMSD of 2.3 Å, and as is evident in Figure 11, there are no residues in the C-terminal 3rd helices that are in structurally conserved positions compared to the αC helices of SSVRH E73. FitA, and RelB both serve as antitoxins in a toxin/antitoxin pair. The DALI Z-score for the alignment of SSVRH E73 and FitA is 3.9, and the RMSD is 6.8 Å. The DALI Z-Score for the alignment of SSVRH and RelB is 2.7, and the RMSD is 2.3 Å. The residues of the C-terminal 3rd helices of FitA also have no structurally analogous positions to the residues in the αC helices of SSVRH E73 (see Figure 11), and in the 3D structure of RelB, NMR analysis indicates that the C-terminal 3rd helices are unstructured in solution when not bound to the toxin RelE [65].

Based on the structural homology of SSVRH E73 to the Ribbon-Helix-Helix domain of DNA binding proteins, *in vitro* Electromobility Shift Assays (EMSA) were carried out to determine if SSVRH E73 could be shown to bind DNA. The recombinant SSVRH E73 gene product used for these structural investigations was incubated at 65°C for 30 minutes with each of 17, ~1000 basepair pieces of the SSVRH viral genome. These ~1000 base-pair pieces were produced by Polymerase Chain Reaction (PCR), using the SSVRH genome isolated by the Mark Young Lab at Montana State University [7]. In the case of each of the 17 pieces a 10:1 molar ratio
of SSVRH E73 to SSVRH genomic DNA was shown to cause a significant gel band shift (Figure 15). SSVRH E73 was also shown to bind unrelated negative-control DNA under these same conditions, leading to the conclusion that SSVRH E73 indeed binds DNA in vitro, but does so nonspecifically.

Figure 15. A representative Electromobility Shift Assay conducted using SSVRH E73, and 1 of 17 ~1000 base-pair pieces of SSVRH genomic DNA. Lane 1 is the SSVRH E73 protein with no SSVRH DNA present. Lane 2 is a 10:1 molar ratio of SSVRH E73 to DNA. Lane 3 is a 1:1 molar ratio of SSVRH E73 to DNA. Lane for is a 1:10 molar ratio of SSVRH E73 to DNA. SSVRH was found to bind all 17 ~1000 base pair pieces of the SSVRH genome in a similar manner.
CHAPTER 4

DISCUSSION

The structure of E73 is that of a homodimeric Ribbon-Helix-Helix (RHH) domain. Each protomer consists of an N-terminal β strand followed by three α-helices, αₐ, α₉ and, αₐ. The β strands and the first two α helices (αₐ and α₉) of each protomer form the canonical RHH domain, shown in Figure 6 with the RHH domain portion of one protomer colored blue, and the other grey. The C-terminal αₐ helices (shown in Figure 6 colored yellow) of SSVRH E73 are not representative of the RHH family, however RHH domain proteins that present C-terminal 3rd helices exist.

Ribbon-Helix-Helix (RHH) proteins represent a diverse and large family of proteins that regulate gene transcription of genes throughout all kingdoms of life including Prokarya, Archaea, and viruses. Numerous examples of this structural fold exist [48] including many structural variations. This domain is found as part of larger protein scaffolds and enzymes, as is the case for NikR, the nickel-dependent regulator of nickel uptake in *E. coli* [57] and the proline utilization flavoprotein PutA [52], both of which are composed of a catalytic or ligand-binding domain as well as a regulatory RHH domain that negatively regulates transcription of their own gene.

The RHH domain can also exist as stand alone molecule, as in the case of CopG, which regulates plasmid copy number in many prokaryotes such as *Streptococcus* by regulating transcription of the RepB protein, which is responsible for initiating plasmid replication by nicking the double strand in order for a helicase
to unwind the DNA, and for DNA Polymerase to perform replication [66]. Another example of a small RHH protein is the Arc repressor whose DNA binding function initiates the switch between the lytic and lysogenic states in *Salmonella* bacteriophage P22 [60].

There are also examples of the RHH domain with C-terminal 3\textsuperscript{rd} helices that add greater levels of regulation to their DNA binding characteristics. The C-terminal third helices categorized for RHH proteins play several functional roles such as; the toxin binding domain of a toxin/antitoxin pair as in ParD (PDB ID 2AN7) [67] and FitA (PDB ID 2BSQ) [68], a tetramerization domain as presented by the Mnt repressor (PDB ID 1MNT) [62] and TrwA (No PDB entry) [69], and ligand binding as in MetJ (PDB ID 1MJB) where repressor activity is initiated when its ligand S-adenosylmethionine binds in the protein as a cofactor [70]. However, there is no known RHH domain with a set of C-terminal third helices that is structurally analogous to those of SSVRH-E73.

Characterization of 3D structure by NMR together with in vitro DNA binding experiments have provided valuable insights into the possible biological function of this protein in SSV-RH. Sequence analysis has shown that this protein is highly conserved among various SSV viral strains, indicating that this protein play a crucial role in the life cycle of *Sulfolobus* crenarchaeal viruses. Herein, the 3D structure of SSVRH E73 has revealed the presence of a highly conserved RHH domain, which unexpectedly is extended to contain a 3\textsuperscript{rd} $\alpha$-helix with profound consequences on the structural arrangement of SSVRH E73.
Based on the recently experimentally determined structures of many SSV proteins, a large number of them appear to bind, or interact with, double stranded DNA. These virally encoded proteins are likely to be crucial elements in regulating viral transcription, and replication, as well as regulating host factors that are necessary for viral transcription and replication. Insights into the life cycles of crenarchaeal viruses have been obtained from transcriptome analysis following *Sulfolobus* host infection and virus production [16]. In the case of SSV1, the ability of the virus to infect its hosts was at first challenging to demonstrate as the virus does not produce significant numbers of particles upon infection [17]. The genome of SSV1 rapidly integrates into a specific tRNA gene site of the host [71] accompanied by a short period of lesser growth for the host [17]. However even after infection of *Sulfolobus* host with excess of virus particles, the host recovers well and seems to grow even better than prior to virus infection, suggesting that the presence of the viral genome confers some advantage to the *Sulfolobus* archea [17]. In addition to its integrated genome, the circular dsDNA genome of SSV1 resides in *Sulfolobus* cells as a plasmid with 3 to 4 copies per host chromosome. Upon UV light irradiation of host cells, viral DNA replication is vigorously induced and results in the release of large amounts of viral particles (up to ~ 100 per cell) into the culture medium without cell lysis of *Sulfolobus* [17, 72].

Nine polycistronic transcripts of various lengths (T1-T9), starting from different promoters, have been mapped onto the SSV1 genome, along with several regulatory elements and a promoter for a UV-inducible transcript (T-ind) [73, 74].
Several transcripts, including the T1/2 and T4/7/8 transcripts, share a common start site but result in transcripts of different lengths due to terminator read-through [73]. With the exception of T-ind, all T1-T9 transcripts are produced at low levels in absence of UV light irradiation, indicating that these are expressed constitutively (albeit in low amounts) in the “latent” state of the virus [74].

A recent microarray study has established a chronological expression of transcripts following UV induction, a process which appears to be tightly temporally regulated and reminiscent of transcriptional regulation processes observed for bacteriophages and better known eukaryotic viruses [16]. The chronological order of transcription of the SSV1 genome in the *Sulfolobus* host has provided new insights into SSV1 gene function and regulation. Upon UV light irradiation, SSV1 genome transcription begins with expression of T-ind that encodes the SSV1 protein B49 [16]. This is followed by early expression of T5 and T6 transcripts followed by a shortly delayed expression of T9, while the T1/2, T3, and T4/7/8 transcripts are up-regulated at a later phase and encode proteins (VP1, VP2, VP3, C792) which have been isolated in purified virions [16]. This genome-wide transcriptional analysis suggests that the UV-induced transcription response observed in SSV1 is not linked to an SOS-like stress response as often observed in bacteria [5, 16]. The ORF E73 from the *Sulfolobus* spindle virus-ragged hills strain (SSVRH E73) is homologous to one of the early T5 transcripts of SSV1 encoding the protein E51, and as mentioned earlier SSVRH E73 displays remarkably high amino acid sequence conservation
among different SSV strains, suggesting that its function in viral life cycle regulation is important.

The expression of SSV1 E51 ORF, along with being one of the very earliest to be transcribed is also constitutively expressed, and transcript levels encoding E51 increase, even as expression of later transcripts is initiated. Four hours following induction, all of the T6 ORFs have are being transcribed, whereas for T5 all open reading frames are not transcribed until six hours. Of the so-called “late transcripts”, T9 transcription is initiated at five hours, T2/8 at six hours. The latest ORFs transcribed are those from T4/7/8 and T7/8, these aren’t all present until eight and a half hours post induction [16]. From the work of Frols et al. and by analogy with SSV-1 E51, SSV-RH E73 is one of the early transcripts to be expressed following UV induction and expression increases over time [16].

This suggests that SSVRH E73 functions differently than simple transcriptional repressors such as CopG, as the latter not only represses transcription of the downstream RepB gene, which in turn facilitates replication of the plasmid, but also represses its own transcription in a negative feedback fashion [59] [66]. The presence of a unique third helix corroborates the hypothesis that E73 has a much more complex function than that of CopG, as was suggested by comparison of the sequences. The cleft distal to the antiparallel β sheet created by the third helix of E73 is likely to interact with some other molecule, and therefore implies that there is the potential for a higher degree of regulation than that seen for the stand-alone RHH domains. It is likely that as E73 is transcribed early on, and
constitutively transcribed as the viral genome is replicated, it plays a key role in viral replication, perhaps by interacting with the *Sulfolobus* host. The cleft created by the loops L2 and the $\alpha_c$ helices is likely to be protein-protein interaction surface, or ligand binding domain.


