Structure, Vol. 9, 1135–1141, December, 2001, ©2001 Elsevier Science Ltd. All rights reserved. PII S0969-2126(01)00680-3

Structure of the Universal Stress Protein of *Haemophilus influenzae*

Marcelo C. Sousa and David B. McKay[1]
Department of Structural Biology
Stanford University School of Medicine
Stanford, California 94305

Summary

Background: The universal stress protein UspA is a small cytoplasmic bacterial protein whose expression is enhanced several-fold when cellular viability is challenged with heat shock, nutrient starvation, stress agents which arrest cell growth, or DNA-damaging agents. UspA enhances the rate of cell survival during prolonged exposure to such conditions, suggesting that it asserts a general "stress endurance" activity. However, neither the structure of UspA nor the biochemical mechanism by which it protects cells from the broad spectrum of stress agents is known.

Results: The crystal structure of *Haemophilus influenzae* UspA reveals an asymmetric dimer with a tertiary α/β fold similar to that of the *Methanococcus jannaschi* MJ0577 protein, a protein whose crystal structure revealed a novel ATP binding motif. UspA differs significantly from the MJ0577 structure in several details, including the triphosphate binding loop of the ATP binding motif; UspA shows no ATP binding activity.

Conclusions: Within the universal stress protein family that is delineated by sequence similarity, UspA is the only member which has been correlated with a cellular activity, and MJ0577 is the only member which has been assigned a biochemical activity, i.e., ATP binding. UspA has a similar fold to the MJ0577 protein but does not bind ATP. This suggests that members of this protein family will segregate into two groups, based on whether or not they bind ATP. By implication, one subset of the universal stress proteins presumably has an ATP-dependent function, while another subset functions in ATP-independent activities.

Introduction

The universal stress protein UspA is thought to function as a general responder to the arrest of cell growth in bacteria [1, 2]. In *Escherichia coli*, it has been shown that expression of UspA is enhanced in response to a broad spectrum of stress agents that arrest cell growth, including H_2O_2, $CdCl_2$, carbonylcyanide-3-chlorophenylhydrazone (CCCP), dinitrophenol (DNP), and osmotic stress with 1 M NaCl; additionally, UspA expression is enhanced in response to heat shock [2]. *E. coli* cells lacking a wild-type *uspA* gene show a reduced survival rate when exposed to the stress agents listed above [2] and also when exposed to DNA-damaging agents such as mitomycin C and irradiation with ultraviolet light [3]. When expressed at elevated levels during cell growth arrest, the UspA protein becomes phosphorylated; the phosphorylation is dependent on the *typA* gene [4].

Analysis of amino acid sequence similarities classify the bacterial UspA proteins as members of the "universal stress protein (USP) domain" family (e.g., Pfam accession number PF00582 [5]), a conserved domain of approximately 140–160 residues. Five *E. coli* proteins (one of which has two tandem USP domains) are members of this family. However, in no case has a biochemical function been identified for any of the *E. coli* or other bacterial USP proteins, and the UspA protein is the only one for which expression has been correlated with a specific biological state in the cell. A protein from a thermophilic archaeon, the product of the MJ0577 gene of *Methanococcus jannaschi*, also belongs to the USP family. The structure of MJ0577 has been solved [6]; it revealed an α/β protomer fold, and surprisingly, a tightly bound ATP molecule. Incubation of MJ0577 with *M. jannaschi* cell extracts induced ATP hydrolysis, suggesting that the protein functions as a cell factor-dependent ATPase.

Here, we report the structure of the UspA protein from *Haemophilus influenzae*, which shares 68% sequence identity with the *E. coli* UspA and 12% identity with MJ0577 in a structure-based alignment (this work). The *H. influenzae* UspA protomer has a tertiary fold similar to that of the *M. jannaschi* MJ0577 protein; at the same time, it has significant structural differences. In particular, the UspA protein shows no evidence of ATP binding, suggesting that members of the USP protein family segregate into at least two subfamilies, one of which binds ATP and has an ATP-dependent function, and one of which does not.

Results and Discussion

UspA Is an Asymmetric Dimer in the Crystal

Four UspA protomers are present in the crystallographic asymmetric unit; they can be assigned to two dimers (Figure 1), which correlates with gel filtration results showing UspA is a dimer of apparent molecular weight 34.5 kDa in solution (Figure 2), to be compared to the computed dimeric molecular weight of 32.0 kDa. The individual subunits have average B factors of 30.0, 28.4, 33.0, and 28.0 $Å^2$ for the A, B, C, and D protomers, respectively. The crossvalidated estimated coordinate error is 0.17 Å. A sulfate ion is sequestered between the subunits of each dimer by the side chains of residues Arg28 and His29 from both protomers. Data collection, phasing, and refinement statistics are summarized in Tables 1 and 2.

The two dimers can be superimposed with a root-mean-square deviation (rmsd) in atomic positions of 1.0 Å for main chain atoms. Within each dimer, the two

[1] Correspondence: dave.mckay@stanford.edu

Key words: UspA (*H. influenzae*); universal stress protein; MJ0577; chaperone; arrest of cell growth; ATP binding

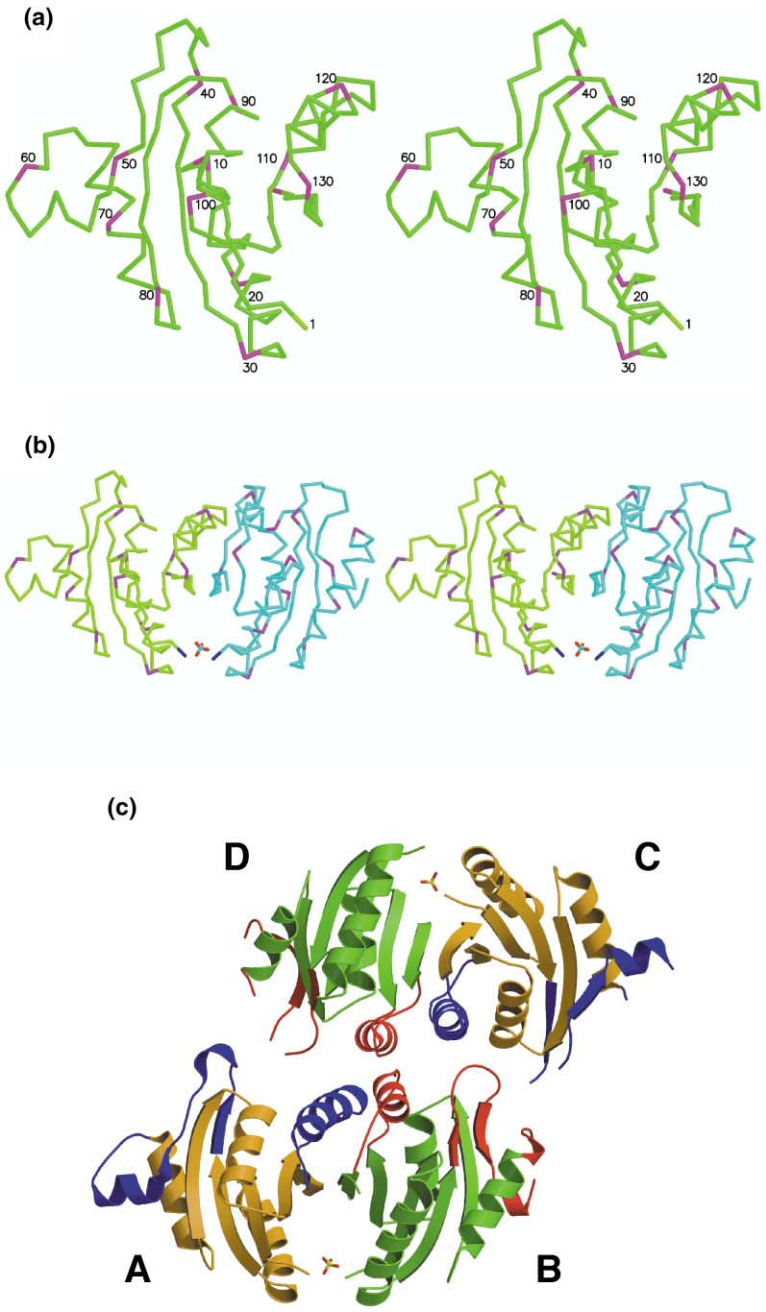

Figure 1. UspA Structure and Crystal Packing

(a) Stereo diagram of α carbon backbone trace of the "A" UspA protomer. Every tenth C_α is highlighted magenta and labeled.

(b) Stereo diagram of α carbon backbone trace of one UspA dimer. The color-coding is as follows: green, A subunit; cyan, B subunit; blue, amino-terminal residue Met1; magenta, every tenth C_α. The sulfate ion that is sequestered between subunits is also shown.

(c) Cartoon diagram of the two UspA dimers in the crystallographic asymmetric unit, showing the intermolecular packing. The color-coding is as follows: subunits A and C (gold and blue) have essentially the same conformation, and subunits B and D (green and red) have the same conformation. Regions in which the A and B subunits (or C and D subunits) of the dimer have the same conformation are gold and green, respectively; regions where the conformation differs substantially is shown in blue and red, respectively. Figures 1 and 3a–3b were made with MOLSCRIPT [17] and rendered with RASTER3D [18].

subunits are related by a 180° (two-fold) rotation when the common structural elements (shown in yellow in Figure 3a) are superimposed; however, the subunits differ significantly in conformation in two segments of polypeptide (shown in green versus magenta in Figure 3a), with the consequence that the dimers are notably asymmetric. One difference arises in the polypeptide segment between β strand β2 and helix α2, residues 41–63. This segment is flexible; in three of the four subunits, some residues could not be unambiguously placed in the model, and the residues that were placed have an average B factor of 47 Å^2. The differences we see between subunits in this region is likely to be due to inherent conformational heterogeneity of this segment of polypeptide.

The other major difference between subunits is at the dimer interface, where helix α4 is partially "unraveled" in one subunit, as compared to the other. The two peptide traces diverge at His110, which appears to be a swivel point. On subunit A, residues 115–129 form a helix, while on subunit B, residues 111–120 are in a helical conformation; the polypeptide paths subsequently converge at residue 128. In contrast to the loops at the periphery of the dimer, this region of the molecule is well ordered; the average B factor is 26 Å^2. The computed buried surface area between protomers in the A-B dimer is $\sim$2300 Å^2; between protomers of the C-D dimer, $\sim$2200 Å^2. Both dimers have a similarly asymmetric interface between subunits, even though the crystal packing environments of the two dimers are different.

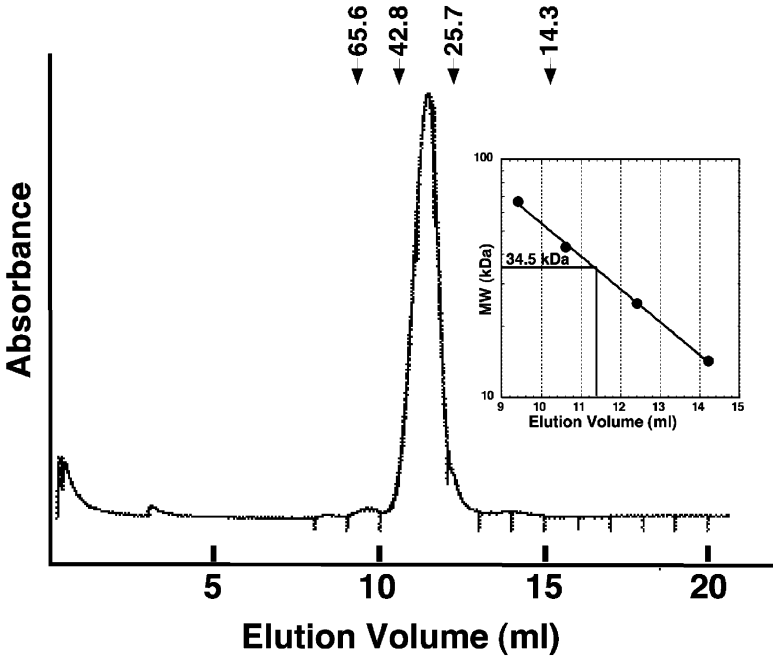

Figure 2. Gel Filtration Profile of UspA

Chromatogram of UspA on Superdex-75 (Pharmacia HR 10/30 column). Peak positions of molecular weight standards from calibration runs are indicated with arrows; the inset shows a semilogarithmic calibration graph, with elution volume and corresponding molecular weight of UspA indicated. Chromatography conditions are as follows: 200 mM $(NH_4)_2SO_4$, 50 mM Mes-K^+ (pH 6.5), flow rate 1 ml/min. Molecular weight standards are as follows: bovine serum albumin (65.6 kDa), ovalbumin (42.8 kDa), bovine chymotrypsinogen (25.7 kDa), and hen egg white lysozyme (14.3 kDa).

The major differences in subunit conformation occur in regions of the molecule that are involved in intermolecular crystal packing contacts. It is possible that the asymmetry between subunits is induced by crystallization and that the dimer is symmetric in solution. Model building reveals that it is possible to build a symmetric dimer with subunits in the B conformation, (although not in the A conformation, due to the steric clash that would result between helices α4 of the two subunits); a symmetric dimer is sterically permissible. Whether either of two alternative models for the asymmetry [i.e., either (1) UspA forms an intrinsically asymmetric dimer in solution, or (2) it is symmetric in solution but the dimer interface can readily be distorted through interactions with other proteins] is related to the molecular function of UspA is a question which deserves further examination.

UspA Is Similar to MJ0577

UspA is similar in protomer fold and dimer assembly to the gene product MJ0577 of *M. jannaschi* [6]. The MJ0577 protomer is a parallel α/β structure with five β strands; the UspA protomer has a similar structural core but differs somewhat by having a sixth, short, antiparallel β strand on the edge of the sheet distal to the dimer

Table 1. Data Collection and Phasing Statistics for SeMet-UspA

Data Statistics

Wavelength (Å)	Resolution (Å) (Highest Shell)	Completeness (%)	R_{sym}[a]	Fraction I > 3σ (I)	Average Redundancy
λ1 = 0.9794 (edge)	30–1.95 (2.02–1.95)	96.8 (97.5)	0.058 (0.240)	0.758 (0.483)	2.5 (2.5)
λ2 = 0.9792 (peak)	30–1.95 (2.02–1.95)	96.8 (97.3)	0.058 (0.319)	0.753 (0.468)	2.5 (2.5)
λ3 = 0.9567 (remote)	30–1.95 (2.02–1.95)	96.9 (97.6)	0.055 (0.261)	0.742 (0.458)	2.5 (2.5)

Diffraction Ratios and Phasing Statistics

	Anomalous Diffraction Ratios[b]			Phasing Power[c]	
Wavelength (Å)	λ1	λ2	λ3	(+) Friedel Mate	(−) Friedel Mate
λ1	0.098	0.048	0.062	0.91	2.49
λ2		0.085	0.059	0.99	2.46
λ3			0.071	reference	1.84

Figure of Merit <m>

Resolution (Å)	28.4–4.38	4.38–3.47	3.47–3.04	3.04–2.76	2.76–2.56	2.56–2.41	2.41–2.29	2.29–2.19	Overall
<m>	0.72	0.67	0.69	0.68	0.66	0.61	0.55	0.52	0.64

[a] $R_{sym} = \Sigma|I_{hkl} - \langle I_{hkl}\rangle|/\Sigma\langle I_{hkl}\rangle$ where I_{hkl} is a single value of measured intensity of hkl reflection, and $\langle I_{hkl}\rangle$ is the mean of all measured values for intensity of hkl reflection. Bijvoet measurements were treated as independent reflections for the MAD phasing data sets.

[b] Anomalous diffraction ratio values are $\langle(\Delta|F|)^2\rangle^{1/2}/\langle|F|^2\rangle^{1/2}$, where $\Delta|F|$ is the dispersive (off-diagonal element) or Bijvoet (diagonal element) difference, computed between 30.0 and 1.95 Å resolution.

[c] Phasing power = $\langle|F_H|\rangle/E$, where $\langle|F_H|\rangle$ is the rms structure factor amplitude for anomalous scatterers, and E is the estimated lack of closure error. Phasing power is listed for each lack-of-closure expression between the reference data set [i.e., (+)Friedel mate at λ3] and the (+) or (−) Friedel set at each wavelength. Phasing powers were calculated using all data between 30.0 and 2.2 Å.

Table 2. Data Collection and Refinement Statistics for Native UspA

Data Collection	
Wavelength (Å)	1.08
Resolution range (last shell) (Å)	30–1.85 (1.91–1.85)
Observations (total/unique)	339,160/49,786
Completeness (%)	99.4 (96.7)
Average I/σ	29.4 (14.4)
R_{sym} [a]	0.041 (0.103)
Refinement	
Resolution range (last shell) (Å)	28.78–1.85 (1.97–1.85)
R_{cryst} [b]	0.220 (0.240)
R_{free}	0.242 (0.282)
Number of reflections (working set)	46,819
Number of reflections (test set)	2,895
Number of protein atoms	4,090
Number of solvent/hetero atoms	232
Average B value, main chain atoms	26.6
Average B value, all protein atoms	27.6
Rmsd bond length (Å)	0.010
Rmsd angles (°)	1.5

[a] $R_{sym} = \Sigma|I_{hkl} - <I_{hkl}>|/\Sigma<I_{hkl}>$ where I_{hkl} is a single value of measured intensity of hkl reflection, and $<I_{hkl}>$ is the mean of all measured values for intensity of hkl reflection.

[b] $R_{cryst} = \Sigma|F_{obs} - F_{calc}|/\Sigma F_{obs}$ where F_{obs} is the observed structure factor amplitude and F_{calc} is the structure factor calculated from model. R_{free} is computed in the same manner as R_{cryst}, using the test set of reflections.

interface, as well as a shorter α2 helix (Figure 3). The UspA and MJ0577 protomers can be superimposed with an rmsd of backbone atomic coordinates of 1.8 Å for 101 residues of the A protomers and 1.9 Å for 93 residues of the B protomers (MJ0577 Protein Data Bank code 1MJH). The dimer interface is similar in the two proteins, being formed largely by an antiparallel interaction between neighboring β5 strands. MJ0577 forms a symmetric dimer, in contrast to UspA, and the relative orientation of subunits in the dimer differs by 27° between the two proteins.

Two of the regions where the two proteins differ substantially in structure are (1) the ATP binding pocket of MJ0577 near the dimer interface, specifically the triphosphate binding loop, and (2) helix α2 and the connecting polypeptide distal to the dimer interface. MJ0577 has an extended helix, while the α2 helix of UspA is much shorter and follows a segment of polypeptide that is more intimately associated with the core structure of the subunit. These segments correspond to regions where the sequence alignment shows insertions or deletions (Figure 4) and also to the regions where the individual protomers within the UspA dimer differ from each other.

UspA shares no structural identity with human serum response factor [7] or Mef2A [8], transcription factors of the MADS-box protein family, despite a prediction based on sequence similarity that the first ~40 residues of UspA would be similar to the extended DNA binding helix of these proteins [9]. Additionally, we find no obvious similarity in structure to any other known DNA binding motifs.

UspA Does Not Bind ATP in the Crystal

One notable feature that was first revealed by the crystal structure of MJ0577 is that it binds ATP tightly. Although ATP was not included during the purification or crystallization of MJ0577, it was present at full occupancy in the structure. In contrast, we see no evidence that UspA binds ATP. The native crystal whose structure is reported here was grown and maintained in the presence of 1 mM MgATP through adaptation to cryoprotectant for freezing and data collection. However, we see no bound nucleotide (ATP or hydrolysis product thereof) in the structure, which excludes tight binding of nucleotide to UspA.

UspA differs substantially from MJ0577 in the structure of the polypeptide corresponding to the triphosphate binding loop of the latter. The segment of polypeptide that aligns in the sequence with the triphosphate binding loop of MJ0577 is of sufficient length to accommodate the three phosphates of ATP, despite being three residues shorter than the MJ0577 loop and despite the fact that much of this segment is in an α-helical conformation in both UspA protomers. In the MJ0577 structure, many of the interactions with the triphosphate are from main chain atoms, not side chains. Notably, two glycines of MJ0577 which sandwich the α- and β-phosphates of ATP (Gly130 and Gly140) align with two bulky residues of UspA, Gln112 and Met119. Manual model building reveals that when the UspA polypeptide is superimposed on the MJ0577 backbone, the bulky side chains of UspA Gln112 and Met119 clash with the triphosphate of ATP. It is likely that the structural differences between MJ0577 and UspA in the polypeptide segment that forms the triphosphate binding loop of the former—and particularly the replacement of glycine residues of MJ0577 with bulky residues in UspA—impairs ATP binding by UspA.

Two Subfamilies for the MJ0577 Fold?

Comparison of the structures of UspA and MJ0577 suggests that the family of proteins which share the MJ0577 fold fall into two groups: those which bind ATP and those which do not. Although the consensus sequence for ATP binding has not been defined, consideration of the observed interactions of the triphosphate binding loop of MJ0577 in the context of structure-based sequence alignments suggests that one essential feature is a G-2X-G-9X-G(S/T) motif, i.e., three glycines separated by two and nine residues in sequence and a hydroxyl of a serine or threonine following the last glycine (in some cases, asparagine follows the last glycine, suggesting some flexibility in this latter constraint). A large fraction of the proteins that have been classified as members of the universal stress protein family (e.g., Pfam accession number PF00582 [5]) have this consensus set of glycines, while another subset, including UspA of *H. influenzae* and *E. coli*, does not. This suggests that members of this sequence family will be segregated on whether or not they bind ATP; by implication, they are likely to have diverged to carry out two or more (as yet unknown) biological functions, one of which is ATP-dependent and one of which is not.

Biological Implications

The universal stress protein UspA has been identified in *E. coli* as a protein whose expression level is elevated

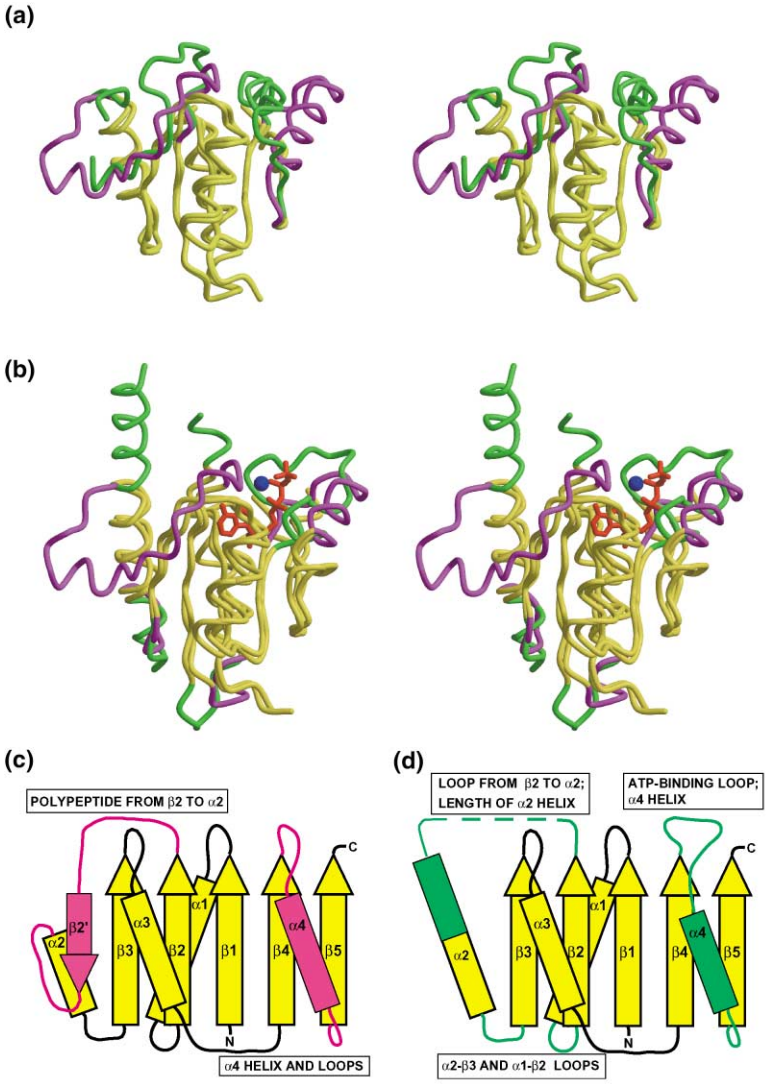

Figure 3. Superposition of Protomers of UspA and MJ0577

Subunits were superimposed with the program LSQMAN of the Uppsala Software Factory [19] and then examined visually for segments where the polypeptides diverged.

(a) Superposition of subunits A and B of UspA, shown as a stereo diagram of a worm representation of the α carbon backbone. The color-coding is as follows: segments of polypeptide where the two subunits superimpose with minimal difference are yellow, and other segments of polypeptide are magenta for subunit A and green for subunit B.

(b) Superposition of subunit A of MJ0577 on subunit A of UspA. Segments of polypeptide where the two subunits superimpose with minimal difference are yellow; other segments of polypeptide are magenta for UspA and green for subunit MJ0577. The ATP bound to MJ0577 is red; Mn^{2+} is blue.

(c) Schematic topology drawing showing where the conformation of the subunits of the dimer differ for UspA, shown in magenta, as compared to similar regions, shown in yellow. For simplicity, topology of a single protomer is shown. The short antiparallel β strand at the edge of the β sheet is labeled β2′ in order to maintain consistency with the numbering of the strands of the β sheet of MJ0577 [6].

(d) Schematic topology drawing showing where UspA and MJ0577 differ, shown in green.

and which is phosphorylated as cells enter a stationary growth state. Absence of UspA substantially reduces cell survival rates during exposure to heat shock, nutrient starvation, agents which impede cell growth, and DNA-damaging agents; UspA functions as a general "stress endurance" factor. However, no specific biochemical activity has been identified for UspA, nor have molecules with which it interacts in the cell been identified (other than the kinase that phosphorylates it). How UspA protects bacterial cells against a broad spectrum of deleterious conditions is unknown.

The crystal structure of the *H. influenzae* UspA, which shares 68% sequence identity with the *E. coli* homolog, reveals a tertiary fold and dimer assembly similar to that

```
                    β1            α1            β2          (β2')
USPA_HAEIN     MYKHILVAVD LSEESPILLK KAVGIAKRHD AKLSIIHVDV-[NFSDLYTGLI DVNMSSMQDR IS]-
USPA_ECOLI     AYKHILIAVD LSPESKVLVE KAVSMARPYN AKVSLIHVDV-[NYSDLYTGLI DVNLGDMQKR IS]-
Y577_METJA MSVMYKKILYPTD FSETAEIALK HVKAFKTLKA EEVILLHVID-[EREIKKRDIF SLLLGVAGLN KSVEEFENEL KNKLT]-
ATP CONTACTS:          AAA                          A

                α2                  β3          α3          β4          α4                    β5
USPA_HAEIN  TETQKALL DLAESV-[DYPI]-  SEKLSGSGDL GQVLSDAIEQ YDVDLLVTGH-[HQDFW...SKLM]-S STRQVMNTIK IDMLVVPLRD E
USPA_ECOLI  EETHHALT ELSTNA-[GYPI]-  TETLSGSGDL GQVLVDAIKK YDMDLVVCGH-[HQDFW...SKLM]-S SARQLINTVH VDMLIVPLAD
Y522_METJA  EEAKNKME NIKKEL-[EDVGFK]-VKDIIVVGIP HEEIVKIAED EGVDIIIMGS-[HGKTNLKEILLG]-S VTENVIKKSN KPVLVVKRKN S
ATP CONTACTS:                                                RR  RPPPP       P  P PR
P-BINDING MOTIF                                             -G.  .G.........G (S/T)-
```

Figure 4. Structure-Based Sequence Alignment of UspA and MJ0577

The amino acid sequences of *H. influenzae* UspA (USPA_HAEIN; SwissProt P44880), *E. coli* UspA (USPA_ECOLI; SwissProt P28242), and *M. jannaschi* MJ0577 (Y577_METJA; SwissProt Q57997) aligned based on the structures of *H. influenzae* UspA and MJ0577. Secondary structure of the A subunit of UspA is shown above the sequence; the short antiparallel β strand is labeled β2′ to retain consistency with the MJ0577 convention. Regions that contact the adenosine (A), ribose (R), or triphosphate (P) of MJ0577 (as described in [6]) are shown, along with residues that appear to be part of a consensus sequence in the ribose plus triphosphate binding loop.

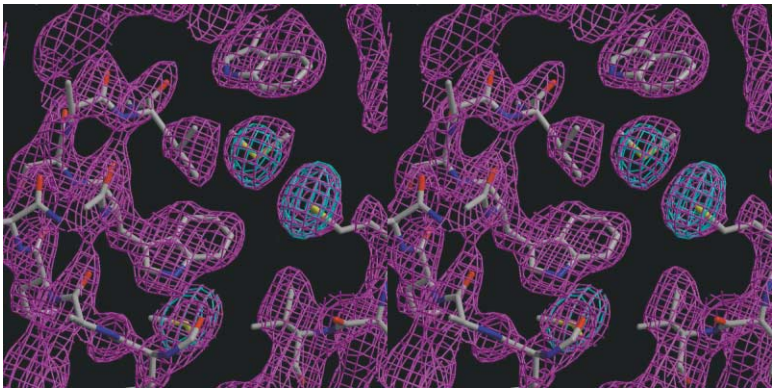

Figure 5. Experimental Electron Density Maps

Stereo diagram showing the experimental solvent-flipped 2.2 Å electron density map for SeMet-UspA, contoured at 1.2σ (magenta), and an anomalous difference Fourier map computed with the same experimental phases and contoured at 6.0σ (cyan). The refined UspA model is included for reference. This figure was made with BOBSCRIPT [20, 21] and rendered with RASTER3D [18].

of the *M. jannaschi* MJ0577 protein. However, unlike MJ0577, UspA shows no evidence of ATP binding activity. The demonstration that MJ0577 is a cell factor-dependent ATPase significantly restricts the subset of functions in which it is likely to be involved. Conversely, the demonstration that UspA does not bind ATP tightly suggests that its activities will be ATP independent, broadening the set of candidate cellular functions in which it might be involved. The UspA structure provides a foundation for structure-function studies to elucidate its biochemical activities.

Experimental Procedures

Cloning, Expression, and Purification

Genomic DNA was purified from *H. influenzae* strain KW20 (ATCC 51907) by standard methods [10]. The coding sequence for the *uspA* gene (TIGR locus HI0815 [11]) was PCR amplified by using a 5′ primer which incorporated an NdeI restriction site overlapping the start codon (5′-CCAGTCAGTGAAGGAGTCGCATATGTACAAACAC; restriction site underlined) and a 3′ primer which incorporated an XmaI restriction site (5′-CAATTTCTTGTCCCGGGTTCATCTCTAAG TGGAAC). The PCR product was cloned into the pTYB2 vector of the "IMPACT" expression system (New England Biolabs) by using the NdeI and XmaI restriction sites, yielding an in-frame fusion with the coding sequence of the intein and chitin binding domains of the vector.

The expression plasmid was transformed into *E. coli* strain BL21(DE3) (Novagen); cells were grown at 37°C to late log phase (A_{600} = 0.6); expression of the recombinant fusion protein was induced with 0.4 mM isopropyl-β-D-glucopyranoside (IPTG) at 28°C; cells were harvested 4 hr after induction. Initial purification and cleavage of the fusion were carried out on a chitin column using the manufacturer's recommended protocol: cells were sonicated in buffer A (0.5 M NaCl, 1 mM EDTA, 0.1% Triton X-100, 25 mM Tris-HCl [pH 7.5]) and centrifuged; the supernatant was loaded onto a chitin column preequilibrated with buffer A. The column was washed with 10 column volumes of buffer A and then quickly flushed with 3 column volumes of cleavage buffer (50 mM NaCl, 30 mM DTT, 25 mM Tris-HCl [pH 7.5]). The flow was stopped, and the column was incubated overnight at 4°C to allow self-cleavage of the UspA-intein fusion. The UspA was eluted with 2 column volumes of cleavage buffer, concentrated to 5 ml, and loaded onto a Superdex 75 column (Pharmacia) equilibrated with 200 mM NaCl, 0.1 mM EDTA, and 20 mM HEPES-K^+ (pH 7.0). UspA protein eluted with an apparent molecular weight of 34.5 kDa, corresponding to a dimer (Figure 2). The protein was concentrated and stored in 10 mM HEPES-K^+ (pH 7.0). The final product consists of the full-length *H. influenzae* UspA plus an additional Pro-Gly dipeptide at the carboxy terminus, which is a consequence of using the XmaI restriction site in the cloning procedure.

For the preparation of selenomethionine-labeled UspA (SeMet-UspA), a protocol for metabolic inhibition of methionine synthesis was employed [12]. BL21(DE3) cells with the expression plasmid were grown at 37°C in M9 minimal media. When the cell density reached A_{600} = 0.8, the temperature was lowered to 28°C, and methionine synthesis was inhibited by adding 100 mg/l each of D-lysine, D-phenylalanine, and D-threonine, 50 mg/l of both D-isoleucine and D-valine, and 60 mg/l D/L-selenomethionine (Sigma). After 15 min, expression of the UspA-intein fusion was induced with IPTG; 4 hr later, the cells were harvested and processed as described above. Amino acid analysis of the purified SeMet-UspA indicated a level of selenomethionine incorporation greater than 95%.

Crystallization and Data Collection

Crystals of both native and SeMet-UspA proteins grew in hanging drops containing equal volumes of protein stock (5–10 mg/ml) and a precipitant/mother liquor consisting of 21% PEG_{4000} (polyethylene glycol of average molecular weight 4000), 200 mM $(NH_4)_2SO_4$, and 100 mM Cacodylate-Na^+ (pH 6.5). Prism-shaped crystals grew to full size ($\sim 50 \times 100 \times 250\ \mu m^3$) in 7–10 days. Crystals were adapted to a cryoprotectant solution consisting of mother liquor plus 20% ethylene glycol with incremental steps of 5% ethylene glycol, then flash frozen in a nitrogen gas stream at 100°K. 1 mM MgATP was included in the crystallization mother liquor and cryoprotectant solutions for crystals of native UspA. Multiwavelength anomalous scattering data were measured at three wavelengths with an inverse beam strategy from a frozen crystal of SeMet-UspA on Stanford Synchrotron Radiation Laboratory (SSRL) beamline 9-2 using a Quantum-4 CCD detector (Area Detector Systems Corporation). The crystal space group is $P2_12_12_1$ with unit cell dimensions a = 64.14 Å, b = 65.19 Å, and c = 137.26 Å and has four UspA protomers per asymmetric unit. Diffraction data from a frozen native UspA crystal were collected on SSRL beamline 7-1 (λ = 1.08 Å) using a Mar345 image plate detector; unit cell dimensions are a = 64.13 Å, b = 63.21 Å, and c = 136.96 Å. All data were processed and scaled using the DENZO/SCALEPACK program package [13]. Data collection statistics for the SeMet-UspA and native UspA crystals are summarized in Tables 1 and 2, respectively.

Structure Determination and Refinement

Crystallographic calculations were effected with the program CNS [14]. An anomalous difference Patterson map was computed in the resolution range 15–3.0 Å using data measured at the absorption peak wavelength of the SeMet-UspA crystal. An automated Patterson search identified ten candidate selenium sites (out of a potential 24), which were used to calculate initial phases using the Philips-Hodgson method and maximum likelihood refinement of CNS. Two additional selenium sites were identified in an anomalous difference Fourier map using these initial phases. A final round of maximum likelihood refinement of the 12 selenium sites and phase calculation were carried out to 2.2 Å resolution (Table 1). The phases were then modified by solvent flipping and histogram matching as implemented in CNS. The resulting electron density map was readily interpretable without the use of noncrystallographic symmetry averaging (Figure 5), and an initial model including four copies of UspA was built using the program O [15, 16]. In all four chains, the polypep-

tide segment from residues 40–60 was poorly defined and was not included in the initial model.

Refinement of the SeMet-UspA model was carried out with the program CNS. Before refinement, a subset of the data was selected as a test set for crossvalidation. An overall anisotropic temperature factor correction was used throughout refinement. The initial SeMet-UspA model was subjected to rigid body refinement followed by a round of slow cool Cartesian annealing, positional refinement, and individual B factor refinement using the remote wavelength data between 20.0 and 2.4 Å and the MLHL target function. Manual rebuilding interspersed with several rounds of simulated annealing, positional refinement, and atomic B factor refinement were then carried out with data to 2.0 Å. At this stage (R_{cryst} = 0.268, R_{free} = 0.295) several solvent molecules and two sulfate ions were clearly visible and added to the model. Refinement using the MHLH target function and manual rebuilding continued until no further drop in the free R factor was observed (R_{cryst} = 0.252, R_{free} = 0.276).

At this point, stronger data to higher resolution from a native crystal became available (Table 2). The SeMet-UspA model was used as a starting point for further model building and refinement against native data (after SeMet was replaced with Met). The model was subjected to rigid body refinement followed by simulated annealing and several rounds of positional and atomic B factor refinement using a maximum likelihood amplitude target function (R_{cryst} = 0.237, R_{free} = 0.269). Manual rebuilding and positional and B factor refinement continued until no further improvement of the free R factor was observed. The final model contains residues 1–140 of chain A (of a full-length chain of 141 UspA residues plus the carboxy-terminal Pro-Gly), residues 1–55 and 58–139 of chain B, residues 1–43, 46–59, and 63–138 of chain C, residues 1–43, 45–55, and 65–137 of chain D, 248 water molecules, and 4 sulfate ions. Final refinement statistics are summarized in Table 2.

Acknowledgments

This work was supported by National Institutes of Health grant GM39928 to D.B.M. Portions of this research were carried out at the Stanford Synchrotron Radiation Laboratory (SSRL), a national user facility operated by Stanford University on behalf of the U.S. Department of Energy, Office of Basic Energy Sciences. The SSRL Structural Molecular Biology Program is supported by the Department of Energy, Office of Biological and Environmental Research, and by the National Institutes of Health, National Center for Research Resources, Biomedical Technology Program, and the National Institute of General Medical Sciences.

Received: July 26, 2001
Revised: September 25, 2001
Accepted: October 18, 2001

References

1. Nyström, T., and Neidhardt, F.C. (1992). Cloning, mapping and nucleotide sequencing of a gene encoding a universal stress protein in *Escherichia coli*. Mol. Microbiol. *6*, 3187–3198.
2. Nyström, T., and Neidhardt, F.C. (1994). Expression and role of the universal stress protein, UspA, of *Escherichia coli* during growth arrest. Mol. Microbiol. *11*, 537–544.
3. Diez, A., Gustavsson, N., and Nyström, T. (2000). The universal stress protein A of *Escherichia coli* is required for resistance to DNA damaging agents and is regulated by a RecA/FtsK-dependent regulatory pathway. Mol. Microbiol. *36*, 1494–1503.
4. Freestone, P., Nyström, T., Trinei, M., and Norris, V. (1997). The universal stress protein, UspA, of *Escherichia coli* is phosphorylated in response to stasis. J. Mol. Biol. *274*, 318–324.
5. Bateman, A., Birney, E., Durbin, R., Eddy, S.R., Howe, K.L., and Sonnhammer, E.L. (2000). The Pfam protein families database. Nucleic Acids Res. *28*, 263–266.
6. Zarembinski, T.I., Hung, L.W., Mueller-Dieckmann, H.J., Kim, K.K., Yokota, H., Kim, R., and Kim, S.H. (1998). Structure-based assignment of the biochemical function of a hypothetical protein: a test case of structural genomics. Proc. Natl. Acad. Sci. USA *95*, 15189–15193.
7. Pellegrini, L., Tan, S., and Richmond, T.J. (1995). Structure of serum response factor core bound to DNA. Nature *376*, 490–498.
8. Huang, K., Louis, J.M., Donaldson, L., Lim, F.L., Sharrocks, A.D., and Clore, G.M. (2000). Solution structure of the MEF2A-DNA complex: structural basis for the modulation of DNA bending and specificity by MADS-box transcription factors. EMBO J. *19*, 2615–2628.
9. Mushegian, A.R., and Koonin, E.V. (1996). Sequence analysis of eukaryotic developmental proteins: ancient and novel domains. Genetics *144*, 817–828.
10. Barcak, G.J., Chandler, M.S., Redfield, R.J., and Tomb, J.F. (1991). Genetic systems in *Haemophilus influenzae*. Methods Enzymol. *204*, 321–342.
11. Fleischmann, R.D., et al., and Venter, J.C. (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science *269*, 496–512.
12. Yu, R.C., Jahn, R., and Brunger, A.T. (1999). NSF N-terminal domain crystal structure: models of NSF function. Mol. Cell *4*, 97–107.
13. Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. In Methods in Enzymology, Volume 276, C.W. Carter, Jr. and R.M. Sweet, eds. (New York: Academic Press), pp. 307–326.
14. Brünger, A.T., et al., and Warren, G.L. (1998). Crystallography & NMR System: a new software suite for macromolecular structure determination. Acta Crystallogr. D *54*, 905–921.
15. Jones, A. (1978). A graphics model building and refinement system for macromolecules. J. Appl. Cryst. *11*, 268–272.
16. Jones, T.A., Zhou, J.Y., Cowan, S.W., and Kjeldgaard, M. (1991). Improved methods for the building of protein models in electron density maps and the location of errors in these models. Acta Crystallogr. A *47*, 110–119.
17. Kraulis, P. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J. Appl. Cryst. *24*, 946–950.
18. Merritt, E.A., and Bacon, D.J. (1997). Raster3D: photorealistic molecular graphics. In Methods in Enzymology, Volume 277, C.W. Carter, Jr. and R.M. Sweet, eds. (New York: Academic Press), pp. 505–524.
19. Kleywegt, G.J., and Jones, T.A. (1997). Detecting folding motifs and similarities in protein structures. In Methods in Enzymology, Volume 277, C.W. Carter, Jr. and R.M. Sweet, eds. (New York: Academic Press), pp. 525–545.
20. Esnouf, R.M. (1997). An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. J. Mol. Graph. Model. *15*, 132–134.
21. Esnouf, R.M. (1999). Further additions to MolScript version 1.4, including reading and contouring of electron-density maps. Acta Crystallogr. D *55*, 938–940.

Accession numbers

The UspA coordinates have been deposited in the Protein Data Bank under the accession code 1JMV.