
Surbhi Gupta, Shashi Bhushan Pandit, Narayanaswamy Srinivasan and Dipankar Chatterji

Molecular Biophysics Unit, Indian Institute of Science, Bangalore-560012, India

1To whom correspondence should be addressed.
E-mail: dipankar@mbu.iisc.ernet.in

S.G. and S.B.P. contributed equally to this work

*Mycobacterium tuberculosis* is a globally successful pathogen, infecting more than one third of total world’s population. These bacteria have the remarkable ability to persist in the host for long periods of time unrecognized by the immune system and then to re-emerge later in life causing the disease. The physiology of such persistent or dormant bacilli is not very well characterized. Some evidence suggests that the dormant bacilli survive in a nutrient-deprived state that is similar to the stationary phase of the bacteria with respect to gene expression and physiology. Under this assumption we have studied the survival of *Mycobacterium smegmatis* in carbon starvation conditions as a model for mycobacterial persistence. *M.smegmatis*, being a fast-growing strain, serves as a good model to study starvation responses. Using the two-dimensional electrophoresis-based proteomics approach, we identified a protein which was found to be expressed preferentially under starvation conditions. This protein is homologous to a family of proteins called Dps (DNA binding Protein from Starved cells) that are known to protect DNA under various kinds of environmental stresses and its existence has, so far, not been reported in mycobacteria. Upon expression and purification of this protein, we observed that it has non-specific DNA-binding ability. Formation of a cage-like dodecamer structure is a characteristic feature of Dps. Using comparative modelling we were able to show that Dps from *M.smegmatis* could form a dodecamer structure similar to the crystal structure of Dps from *Escherichia coli*.

**Keywords:** DNA protection/Dps/mycobacteria/starvation/two-dimensional electrophoresis

**Introduction**

Tuberculosis is currently the leading cause of death from a single infectious agent, killing three million people worldwide every year. The pathogen *Mycobacterium tuberculosis* is responsible for the disease. This pathogen has a remarkable ability to survive in the host without being recognized by the immune system. In this latent state bacteria may remain in the body for decades and then reappear later in life, causing the infection (Parrish *et al.*, 1998). Most cases of active tuberculosis are believed to arise by reactivation of such dormant bacteria by several factors (Bloom and Murray, 1992). These dormant bacilli or ‘persistors’ are resistant to many conventional drugs. BCG vaccination also has little effect in blocking reactivation of the bacteria. Thus, to reduce the prevalence of tuberculosis infection it is necessary to understand the physiology of latent mycobacteria, but an efficient model in this respect is lacking.

The granulomas in the host where the bacteria reside present a nutritionally deprived and anoxic environment to these persistors. Therefore, anaerobic and starved cultures of mycobacteria are often used as models to study the molecular basis of latency (Wayne and Hayes, 1996; Cunningham and Spreadbury, 1998; Dick *et al.*, 1998). There have been a number of important observations demonstrating morphological and physiological similarities between latent bacilli and *in vitro* starved cultures of various species of mycobacteria (Nyka, 1974). Such observations give a clue that latent bacilli residing in granulomas might be nutritionally starved. Adaptation of bacteria to nutritional stress occurs predominantly through a mechanism called the stringent response. The hallmark of stringent response is the accumulation of a nucleotide tetraphosphate, ppGpp, and two enzymes, RelA and SpoT, that are known to maintain the level of this unusual nucleotide (Cashel *et al.*, 1996; Chatterji and Ojha, 2001). Accumulation of ppGpp was shown to occur in *Mycobacterium tuberculosis* (Avarbock *et al.*, 1999) and *Mycobacterium smegmatis* while under carbon starvation (Ojha *et al.*, 2000). As *M.smegmatis* is a fast-growing mycobacterium, we used it as a model for starvation and latency. During the course of investigations, we also observed that overexpression of *relA*, the gene responsible for the synthesis of ppGpp, renders coccoid morphology to *M.smegmatis* similar to that of persistors (Ojha *et al.*, 2000).

Based on the available information, we opted to search for proteins that become differentially expressed during starvation, since such proteins would be important for the survival of latent bacilli. With the estimated number of genes from the genome sequence of *M.tuberculosis* being 4000 (Cole *et al.*, 1998), it is possible to aim for the separation of the whole proteome of the bacteria by two-dimensional electrophoresis. This high-resolution technique allows the separation of up to 10,000 protein species in one electrophoretic run (O’Farrell, 1975).

We report here the identification of a novel protein in *M.smegmatis*, which was found to be preferentially expressed under nutritional stress. This protein is highly homologous to a protein designated Dps (DNA-binding Protein from Starved cells) in *Escherichia coli*, which is expressed under starvation conditions (Almiron *et al.*, 1992). Dps binds DNA without apparent sequence specificity, forming extremely stable complexes. Mutants lacking Dps are very sensitive to H2O2 in the stationary phase and show an altered pattern of protein synthesis (Almiron *et al.*, 1992). Dps is critical for the survival during oxidative and nutritional stress, independent of whether cells are actively growing or not (Matin, 1991; Altuvia *et al.*, 1994; Lomovskaya *et al.*, 1994). Ultrastructural studies have indicated aggregation of dodecameric Dps protein upon the addition of DNA (Almiron *et al.*, 1992). In the stationary state of the
bacteria, the abundant Dps protects DNA through Dps–DNA co-crystallization (Wolf et al., 1999). The availability of the high-resolution crystal structure of Dps protein from \textit{E.coli} (Grant et al., 1998) and \textit{Listeria innocua} (Ilari et al., 2000) enabled us to generate a three-dimensional structural model of Dps from \textit{M.smegmatis}. The tertiary and quaternary structural models suggest similarities in the structural properties of \textit{M.smegmatis} Dps with \textit{E.coli} Dps, indicating a strong possibility of its binding to DNA. The DNA binding ability was then experimentally demonstrated by gel retardation assay using the purified Ms-Dps.

**Materials and methods**

**Strains, media and growth**

\textit{M.smegmatis}, strain mc\textsuperscript{2}155, was grown in MB7H9 (Difco) with 2% glucose and 0.05% Tween-80 for the enriched culture.

In the carbon-starved medium the glucose concentration was reduced to 0.02%. For the plate culture the same composition was used with 1.5% agar.

**Two-dimensional electrophoresis**

Two-dimensional (2-D) electrophoresis was carried out using a Hoeffer SE-600 unit (Amersham Pharmacia) following the procedure described earlier (O’Farrell, 1975). In brief, \textit{M.smegmatis} was grown in both enriched and carbon-starved medium until mid-log phase. The culture broth was centrifuged for 5 min at 4000 r.p.m. and 4°C. The pellet was washed once with STE solution (150 mM NaCl, 10 mM Tris–HCl, 1 mM EDTA, pH 8.0) and was resuspended in lysis buffer (9.5 M urea, 2% ampholytes pH 5–7, 0.4% ampholytes pH 3.5–10, 5% β-mercaptoethanol, 40 mM PMSF, 5 mM EDTA pH 8.0, 10 µg/ml antipain, 10 µg/ml trypsin inhibitor). Cells were stirred in this lysis buffer for 1 h at 4°C, following which sonication was applied until the solution became clear. Soluble proteins were obtained by centrifugation of the cell extract at 12 000 r.p.m. for 60 min at 4°C. A rough protein estimation was made by taking the absorbance at 280 nm. Prefocusing of the isoelectric focusing (IEF) tube gel (9 M urea, 2% ampholytes pH 4–6, 3.3% T acrylamide, 2% NP-40) was applied for 1 h (400 V for 15 min, 500 V for 15 min and 600 V for 30 min). A 700 µg amount of total protein in a volume of not more then 20 µl was then carefully loaded on to the prefocused tube gel and run at 600 V for 7 h followed by 1000 V for 1 h. After the run, the tube gels were extruded out with a syringe and the gels were equilibrated in SDS buffer (10% glycerol, 4.9 mM DTT, 2% SDS, 0.004% bromophenol blue, 0.125 M Tris base) for 5 min. The second dimension was run on a standard 12.5% SDS gel prepared by

---

**Table I. Peptide sequences obtained upon tryptic digestion of spot 1**

<table>
<thead>
<tr>
<th>Peptide No.</th>
<th>Measured $M_r$ (M + H$^+$, Da)</th>
<th>Peptide sequence by ESI–MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>602.5 +2</td>
<td>XATXGK</td>
</tr>
<tr>
<td>2</td>
<td>882.4 +2</td>
<td>FQWVFVR</td>
</tr>
<tr>
<td>3</td>
<td>1009.4 +2</td>
<td>GYADEVAER</td>
</tr>
<tr>
<td>4</td>
<td>1059.6 +2</td>
<td>ASDVADXXQK</td>
</tr>
<tr>
<td>5</td>
<td>1170.6 +2</td>
<td>TWDDYSVER</td>
</tr>
<tr>
<td>6</td>
<td>1335.8 +2</td>
<td>- - - FTX - - -</td>
</tr>
<tr>
<td>7</td>
<td>1546.7 +3</td>
<td>QXSTYDDXHTXK</td>
</tr>
<tr>
<td>8</td>
<td>1980.2 +4</td>
<td>AHXESAGGGXTHEGQSTEK</td>
</tr>
<tr>
<td>9</td>
<td>2321.1 +3</td>
<td>XEDXDSQDXXAHADXEK</td>
</tr>
</tbody>
</table>

---

**Fig. 2.** Enlarged views of the region containing spot 1 (arrows show the position of the spot of interest), which was subsequently analysed. Left, 0.02% glucose; right, 2% glucose.
Carbon starvation in *Mycobacterium smegmatis*

Fig. 3. (a) Frames output of the predicted ORF. (b): Codon Preference output of the predicted ORF. Arrows show the start and stop of the predicted ORF.

Table II. Comparison of the molecular mass of the tryptic digest of the peptide obtained by ESI–MS of induced protein and translated protein

<table>
<thead>
<tr>
<th>Peptide No.</th>
<th>Molecular mass of peptide (Da)</th>
<th>ESI–MS</th>
<th>Translated protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>602.5</td>
<td>601.7</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>882.4</td>
<td>882.0</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1009.4</td>
<td>1009.0</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1059.6</td>
<td>1059.2</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1170.6</td>
<td>1170.2</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1335.8</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1546.7</td>
<td>1545.7</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1980.2</td>
<td>1980.1</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>2321.1</td>
<td>2321.6</td>
<td></td>
</tr>
</tbody>
</table>

*The peptide whose sequence was ambiguous and also did not match with the tryptic digest of the translated protein.*

the Laemmli protocol (Lammeli, 1970). The gels were then silver stained and compared visually.

**Peptide mass fingerprinting**

Peptide mass fingerprinting and sequencing of the peptides were carried out by a commercial network situated at the University of Virginia, USA.

**Sequence analysis**

A search for the sequences of peptides identified from *M.smeqmati* was carried out against the unannotated partially sequenced genome of *M.smeqmati* and *M.avium* available at TIGR using the tblASTN (protein query searched against the nucleotide database). The software Frames and Codon Preference for ORF prediction, PeptideSort for tryptic digest of protein sequence and Translate to translate DNA sequence were used from the GCG package. MALIGN (Johnson et al., 1993) was used to perform multiple sequence alignment of amino acid sequences.

**Comparative modelling of Dps from M.smeqmati**

Three-dimensional (3-D) structures of Dps from *E.coli* (Grant et al., 1998) and the homologue from *L.innocua* (Ilari et al., 2000) determined by X-ray analysis are available in the RCSB protein databank (Berman et al., 2000). The sequences of homologous proteins were aligned on the basis of their structural features and relationships, using the program COMPARER (Sali and Blundell, 1990; Zhu et al., 1992). The sequence of Dps from *M.smeqmati* (MS-DPs) was then aligned with the known structures and both of the structures were initially considered as a basis for building the model of MS-Dps, since the use of many basis structures ensures a more accurate model (Srinivasan and Blundell, 1993). However, the sequence of Dps from *E.coli* is more closely related to MS-Dps (sequence identity 35%) than to Dps from *L.innocua* (sequence identity 25%) and there are more insertions and deletions in the sequence of *L.innocua* variant with respect to MS-Dps. Hence we modelled MS-Dps on the Dps from *E.coli* only (protein databank code: 1dps).

The suite of programs encoded in COMPOSER (Srinivasan and Blundell, 1993) and incorporated in SYBYL (Tripos, St. Louis, MO) was used to generate the 3-D model of MS-Dps. The structures of the conserved regions of 1dps were extrapolated to the equivalent regions of MS-Dps. The variable regions were modelled by identifying a suitable segment from a known structure in the databank. A search was made for a segment having the desired number of residues and the proper end-to-end distances across the three ‘anchor’ Cα at the either side of the putative loop such that the loop can be fitted joining the contiguous conserved regions. A template matching approach (Topham et al., 1993) to rank the candidate loops was also used. The best ranking loop with no short contacts were then saved. The rest of the protein was fitted using the ring-closure procedure of F.Eisenmenger (unpublished results). Side chains were modelled either by extrapolating from the equivalent positions in the basis structure where appropriate or by using rules derived from the analysis of known protein structures (Sutcliffe et al., 1987).

**Energy minimization**

The COMPOSER-generated model was energy minimized in SYBYL using the AMBER force-field (Weiner et al., 1984). During the initial cycles of energy minimization the backbone
Fig. 5. Sequence alignment of *M. smegmatis* (Ms-Dps) and *M. avium* (Mv-Dps) with the bona fide members of Dps family from the Pfam Database. Residues forming the ‘DNA binding signature’ are shown in bold.

was kept rigid and side chains alone were moved. Subsequently all atoms in the structure were allowed to move during minimization. This approach kept the disturbance of the backbone structure to the minimum. Energy minimization was performed till all short contacts and inconsistencies in geometry were rectified. During the initial stages of minimization, the electrostatic term was not included, as the initial main objective was to relieve steric clashes and to rectify bad geometry.

**Modelling of the quaternary structure**

Based on the alignment of amino acid sequence shown in Figure 6, residues at the inter-protomer interface of the dodecameric structure *E. coli* Dps (1dps) were compared with the equivalent residues of Ms-Dps. The high similarity of the residues forming inter-protomer contacts in 1dps with the aligned residues in Ms-Dps suggests that Ms-Dps could form a dodecameric structure (see below).

The tertiary structural model of Ms-Dps was superimposed with each of the 12 subunits in 1dps without altering the coordinate system of atomic positions in 1dps. A dimeric form was energy minimized to ensure the proper inter-protomer interactions. The dodecameric model of Ms-Dps was analysed for the surface properties and explored for the possibility of DNA binding.

**Cloning and sequencing of Ms-Dps**

A DNA fragment containing the Dps gene was amplified from genomic DNA of *M. smegmatis* mc2155 by PCR using a pair of oligonucleotide primers carrying NdeI and HindIII sites. The primers (5'-H11032 AAGGAGCACATATGACCTCATTCACCATCCC 3', sense primer and 5'-H11032 GTTCTAAGCTTGGCAGACTTGCGGCGCGCC 3', antisense primer) were designed using the sequence of contig 3021 of *M. smegmatis* available at the TIGR database and synthesized (Microsynth, Switzerland).
Carbon starvation in Mycobacterium smegmatis

Fig. 6. Structure-based sequence alignment of *E. coli* and *L. innocua* Dps also aligned with MS-Dps using COMPERER (Šali and Blundell, 1990; Zhu et al., 1992). The alignment is represented using JOY (Mizuguchi et al., 1998). The amino acid code is the standard one-letter code formatted by using the following convention: *italic* for positive φ angle; *UPPER CASE* for solvent-inaccessible residues; lower case for solvent accessible residues; *bold* for hydrogen bonds to the main chain amide; *underline* for hydrogen bonds to the main carbonyl oxygen; and tilde (~) for side chain to side chain hydrogen bonds.

The resulting PCR fragment was cloned into pGEM-T Easy vector (Promega). The insert was then subcloned into pET 21b using the enzymes NdeI and HindIII to produce pET-Dps. The nucleotide sequence of the 552 bp DNA was then confirmed (Microsynth).

**Results**

**Protein separation**

The total protein profile of *M. smegmatis* grown in 2% C source and 0.02% C source showed differential expression of few proteins. It should be mentioned that we could detect only a small fraction of the total protein component of the bacteria for various reasons, such as narrow pH range, loss during 30 min at 30°C. The complex was resolved on 1.2% agarose gel run in 0.5 × TBE buffer and stained with ethidium bromide.

**Nucleotide sequence accession number**

The sequence of the cloned Ms-Dps has been deposited in Genbank (accession number AY065628).

**Gel retardation assays**

Supercoiled pUC19 DNA (50 ng) was incubated with 2–4 μg of protein in 10 mM Tris–HCl (pH 7.9), 10 mM NaCl for
transfer from 1-D to 2-D gel, etc., so the 2-D gels reported here only represent a detectable limit of proteins in our assay system. We focused our attention on the appearance of the protein spots under C-starvation in comparison with C-fed conditions. Many protein spots were found to be absent in C-starved conditions, as can be seen in Figure 1a and b. Such differential expression of proteins can be explained from the fact that starved bacteria shut down the synthesis of proteins that are not absolutely essential for survival during stress. In contrast, the protein spots specifically present in starved conditions are interesting to look at since these could be the proteins that would be helpful for the latent bacteria to survive in adverse situations. It appears that in spite of the low nutrient environment, the limited energy source is being used for the synthesis of these proteins. Such a protein spot visible in carbon-starved culture but not in carbon-enriched culture (Figure 2) was used for further investigation. This protein spot had shown consistent differential expression from three different batches of low- and high-glucose containing bacterial cultures.

The masses of peptides obtained upon in-gel tryptic digestion of the protein spot and the sequences of the peptide are shown in Table I.

**Genome data analysis**

The peptide masses were used to search the database using the software available at Prospector (prospector.ucsf.edu), MOWSE (www.mann.embl-heidelberg.de) and Proteometrics (www.proteometrics.com). No significant match was obtained with the peptide molecular masses. It should be noted that
translated proteins from partially sequenced genomes are not included in the databases available at these sites. These peptide sequences were then used to search in *M. smegmatis* unannotated partially sequenced genome available at TIGR (www.tigr.org).

The tBLASTN against *M. smegmatis* unannotated partially sequenced genome gave one hit with the contig number 3021. This contig DNA sequence was used for the prediction of open reading frame (ORF) using Frames and Codon Preference programs available in the GCG package, using the *M. smegmatis* codon usage table for the latter. Both of the ORF prediction methods gave ORF from nucleotide 731 to nucleotide 1283, as shown in Figure 3a and b. Upon translation, this putative ORF gave 183 amino acid long protein. Molecular masses of the peptides obtained on theoretical tryptic digestion of the translated protein using Peptide sort matched those obtained by electrospray ionization mass spectrometry (ESI-MS) within ± 1Da except for a peptide of molecular mass 1335.8 Da whose sequence was ambiguous. The masses of the peptides obtained upon tryptic digestion experimentally and theoretically are shown in Table II.

These results clearly indicated that the identified ORF codes for the protein in *M. smegmatis* induced under carbon starvation.

There was ambiguity in the sequence obtained by ESI-MS and translated ORF at three residues as shown in Figure 4. The translated protein sequence was searched against the family domain database Pfam (Bateman et al., 2000) and it showed homology to the Dps family of proteins with an E value <<10^{-10}. This family of proteins is known to be induced under oxidative stress in *E. coli*.

The tBLASTN of translated protein against *M. avium* unannotated partially sequenced genome at TIGR gave one hit with contig number 4. The putative ORF was predicted as before, from nucleotide 1001 to nucleotide 1549 (data not shown). This translated protein also showed homology to the Dps family of proteins in the Pfam database.

Similar searches in *M. bovis*, *M. leprae* and *M. tuberculosis* genomes did not pick up any protein homologous to the predicted Dps in *M. smegmatis*. The use of PSI-BLAST (Altschul et al., 1997) and IMPALA (Schaffer et al., 1999) searches also did not identify any protein homologous to the Dps. The searches using the signature domain of Dps proteins also did not locate a close homologue of Dps.

The SCOP database (Murzin et al., 1995) suggests that Dps and bacterial ferritins form a superfamily, implying that these families of proteins have evolved from a common ancestor.

The amino acid sequences of putative Dps proteins from *M. smegmatis* (Ms-Dps) and *M. avium* (Mv-Dps) were aligned with the bona fide members of the Dps protein family available in the Pfam database using MALIGN and the alignment is shown in Figure 5. The group of 15 highly conserved amino acids that form the ‘DNA binding signature’ (Ilari et al., 2000) are found to be conserved in both the predicted Dps. The *M. smegmatis* Dps and *M. avium* Dps showed 78% sequence identity.

**Tertiary fold**

The alignment between the amino acid sequences of Dps homologues of known structures from *E. coli* and *L. innocua* was performed using the COMPARER (Sali and Blundell, 1990; Zhu et al., 1992). The amino acid sequences of the putative Dps from *M. smegmatis* was aligned with the known structures by considering the structural features at various residue positions and the probabilities of various amino acids occurring in a given structural environment. The alignment decorated with the representation of structural features using the program JOY (Mizuguchi et al., 1998) is shown in Figure 6. As mentioned earlier, the Dps homologue from *E. coli* (1dps) was used as the template to model the 3-D structure of the putative Dps from *M. smegmatis*. The alignment between MS-Dps and the 1dps showed an insertion of two residues in MS-Dps that corresponded to one of the loop regions of 1dps. Apart from this, no insertion or deletion is present in the two sequences, indicating that although the sequence identity is only ~35%, these two sequences are fairly closely related to each other, which will enable one to model a reasonably accurate 3-D structure. The only structurally variable region defined during the use of COMPOSER corresponds to a segment which includes the two residue insertion in MS-Dps. The progress made in the energy minimization ensured the overall good stereochemical quality of the model.

As substitution of a non-polar amino acid residue in the crystal structure by a polar residue in MS-Dps could cause some instability in the structure of MS-Dps, we particularly investigated this aspect. Most of the buried hydrophobic residues in 1dps are conserved, in terms of hydrophobic
nature, in MS-Dps. However, in four positions, corresponding to residue numbers 69, 132, 135 and 151 in 1dps, the buried hydrophobic residues are substituted by polar residues. Interestingly, Val132 (1dps) → Thr (MS-Dps) and Ala135 (1dps) → Ser (MS-Dps) correspond to $i$ and $i + 3$ positions of an α-helical segment and hence are spatially proximal. It is interesting that the hydrophobic interaction in 1dps is replaced by the interaction between polar side chains in MS-Dps. In fact Ala151 (1dps) → His (MS-Dps) which is present in the following helix is spatially proximal to the residues in positions 132 and 135 of 1dps and hence a hydrophobic cluster in 1dps is substituted by a cluster of polar groups which are interacting with one another. Phe69 of 1dps is buried and it is replaced by Gln in MS-Dps. An examination of the model suggests that the side chain of Gln could form hydrogen bonds with main chain polar atoms in the vicinity. In a few positions in the alignment, the buried polar residues (which are generally uncommon) of 1dps are replaced by hydrophobic residues and this might result in a better hydrophobic core of MS-Dps compared with 1dps. In most positions of the alignment, the acidic and basic residues are conserved in terms of the charged nature of the residues. There are three residue positions in 1dps (56, 58 and 88) corresponding to the positive φ angle that is usually preferred by Gly and Asn residues. In fact, two of these three residues are Gly and the other is Asn. All these residues are conserved in MS-Dps. In general, the residues in MS-Dps are comfortably accommodated in the fold of 1dps and the compatibility of the sequence of MS-Dps to the Dps fold is excellent.

Comparisons between inter-protomer regions of E.coli Dps and M.smegmatis Dps

One of the essential features of Dps protein appears to be the formation of a dodecamer. We specifically assessed the possibility of MS-Dps forming a dodecameric structure. We analysed the inter-protomer region in E.coli Dps and compared the inter-protomer residues with the equivalent residues in M.smegmatis Dps. There are 20 residues in the inter-protomer interface of E.coli Dps that could be involved in interaction between the protomers. Of these, five residues are conserved in M.smegmatis Dps. Among the 15 residues, which are different in the two Dps proteins, nine residues are substituted by highly similar residues and these changes are not expected to affect the nature of quaternary structure. We analysed the six changes that correspond to replacement of a residue in E.coli Dps by a very different residue in MS-Dps. These changes appear to be congenial for the inter-protomer interaction. For example, the changes Arg70 (1dps) → Val (MS-Dps) and Asn99 (1dps) → Ile (MS-Dps) lie in a hydrophobic environment. The residues involved in the changes Ser100 (1dps) → Lys (MS-Dps) and Ile111 (1dps) → Arg (MS-Dps) are not involved in the interactions between the protomers although they are located in the inter-protomer interface. These residues are solvent exposed in both the crystal structure and the model.

We also analysed the dimer–dimer interface in the crystal structure. There are 21 residues at the interface that could be involved in the dimer–dimer interaction. Of these, 11 are conserved between E.coli and M.smegmatis Dps and three residues in E.coli are replaced by similar residues in M.smegmatis Dps. We analysed the seven drastic changes in M.smegmatis compared with E.coli and found that most of the changes could be accommodated at the interface. For example, in Leu15 (1dps) → Thr (MS-Dps) and Leu16 (1dps) → Ser (MS-Dps) changes, Thr and Ser are solvent exposed whereas Arg17 (1dps) → Ile (MS-Dps) and Asn18 (1dps) → Pro (MS-Dps) contribute to the hydrophobic environment at the interface. In the cases of Asp142 (1dps) → Leu (MS-Dps) and Asp143 (1dps) → Val (MS-Dps), the changes are such that the residues contribute to the hydrophobic interaction at the dimer–dimer interface of M.smegmatis Dps.

Quaternary structure

A series of dimers of E.coli variant make a cage-shaped dodecameric form. An analysis of the inter-protomer residues of 1dps and comparison with the topologically equivalent residues of MS-Dps suggest that MS-Dps could form a dodecamer similar to 1dps. The dimer model of MS-Dps generated by superimposing a tertiary structural model on the protomers of 1dps was energy minimized and the dodecameric model of MS-Dps was generated using a series of dimer models. The inter-protomer interface of the dodecameric MS-Dps model is devoid of any unusual features such as short contacts or bad geometry. The dodecamer structures of MS-Dps and E.coli Dps are shown in Figure 7a and b, respectively.

The overall solvent-accessible surface area of the dodecameric model is 18 185 Å$^2$ which is larger than that of 1dps (16 930 Å$^2$). Despite this difference, the area buried in the inter-protomer interface as a consequence of dodecamer formation is of the order of 1000–1100 Å$^2$ for the model and crystal structures. However, the non-polar area buried in the protomers of the model and crystal structure are about 800 and 700 Å$^2$, respectively. The polar areas buried due to dodecamer formation are 330 and 250 Å$^2$ for the crystal structure and model, respectively. This means that in the MS-Dps model the inter-protomer interface is made-up of more non-polar groups than in 1dps, indicating that the dodecameric form of MS-Dps is likely to be more stable than that of 1dps. If the tertiary structure of MS-Dps is exposed to the surroundings it would result in the solvent exposure of more number of non-polar residues than in the case of 1dps, rendering the dodecameric form of MS-Dps very stable.

The surface properties of the MS-Dps model and the crystal structure 1dps were analysed using the program GRASP (Nicholls et al., 1991). Both molecules showed a preponderance of negative charge in the structure of the dodecamers (Figure 8a and b, respectively). It is proposed that a series of dodecameric structures of Dps will bind non-specifically around DNA during starvation.

DNA binding ability

The gene corresponding to Ms-dps was cloned from M.smegmatis genomic DNA and protein was purified to homogeneity (Figure 9a). This purified protein was used for DNA binding assays in vitro. Since Dps is a non-specific DNA binding protein, supercoiled pUC19, an E.coli plasmid, was used as a template. The change in the mobility of the DNA–Dps complex with respect to DNA alone could be visualized on ethidium bromide-stained agarose gels (Figure 9b).

Discussion

Growing cells respond to specific growth conditions by induction of stress-specific proteins to prevent cellular damage. Bacteria spend most of their life in a non-growing or nutrient-limited state which requires them to develop special strategies to ensure survival for longer periods of time, regardless of
the nature of the growth-restricting factor. Adaptability to environmental conditions is the hallmark for bacterial gene regulation. The most important molecule that requires protection during stress is the DNA of the cell. Any damage done to the DNA would mean a loss of genetic information. Such DNA lesions if left unrepaired can cause lethal effects even at low frequency. In bacterial natural habitats, exposure to DNA-damaging factors such as oxidizing and alkylating agents, radicals and UV radiation is often accompanied by nutrient depletion. The outcome of DNA lesions sustained during starvation can be very severe. One of the mechanisms by which bacterial cells protect their DNA from such assaults is by overproduction of a non-specific DNA-binding protein called Dps (Almiron et al., 1992).

The presence of Dps family members is known in a number of distantly related bacteria (Chen and Helmann, 1995; Pena and Bullerjahn, 1995). Now, the identification of a Dps homologue in mycobacteria suggests that the protection afforded by Dps to DNA is probably a global phenomenon. Since it is known that the in vivo environment encountered by mycobacteria in the host is stressful, the presence of Dps in mycobacteria could be important for its survival inside the host.

It is interesting that we were unable to identify any ‘Dps-like’ protein in M. tuberculosis genome by many sequence- and structure-based searches, although a putative sequence was found to be present in M. avium. It could be possible that the sequence of Dps has diverged among different species of mycobacteria to the extent that the similarity level falls below the threshold used in our searches. However, we did not pursue this further.

The results presented here also show that Dps isolated from M. smegmatis has DNA-binding ability, as expected, and it would be interesting to see the arrangement of this protein around the naked DNA.

Acknowledgements

The authors thank the Department of Biotechnology, Government of India, for financial help and the W.M. Keck Biomedical Mass Spectrometry Laboratory, University of Virginia, USA for LC–MS–MS analysis. N.Srinivasan is supported by a Senior Fellowship from the Wellcome Trust, London.

References


Received June 8, 2001; revised December 21, 2001; accepted March 8, 2002

Carbon starvation in Mycobacterium smegmatis