Functional insights by comparison of modeled structures of 18kDa small heat shock protein and its mutant in *Mycobacterium leprae*

Elengikal Abdul Azeez Rehna1, Sanjeev Kumar Singh2,* and Kuppamuthu Dharmalingam1

1Department of Genetic Engineering, School of Biotechnology, Madurai Kamaraj University, Madurai-625021, Tamil Nadu, India; 2Centre of Excellence in Bioinformatics, School of Biotechnology, Madurai Kamaraj University, Madurai-625021, Tamil Nadu, India; Sanjeev K. Singh* - Email: skysanjee@gmail.com; * Corresponding author

Abstract: In this work we are proposing Homology modeled structures of *Mycobacterium leprae* 18kDa heat shock protein and its mutant. The more closely related structure of the small heat shock protein (sHSP) belonging to the eukaryotic species from wheat sHSP16.9 and 16.3kDa ACR1 protein from *Mycobacterium tuberculosis* were used as template structures. Each model contains an N-terminal domain, alpha-crystalline domain and a C-terminal tail. The models showed that a single point mutation from serine to proline at 52nd position causes structural changes. The structural changes are observed in N-terminal region and alpha-crystalline domains. Serine in 52nd position is observed in β4 strand and Proline in 52nd position is observed in loop. The number of residues contributing α helix at N-terminal region varies in both models. In 185 more number of residues is present in α helix when compared to 18P. The loop regions between β3 and β4 strands of both models vary in number of residues present in it. Number of residues contributing β4 strand in both models vary. β6 strand is absent in both models. Major functional peptide region of alpha crystalline domains of both models varies. These differences observed in secondary structures support their distinct functional roles. It also emphasizes that a point mutation can cause structural variation.

Keywords: *Mycobacterium leprae*, *Mycobacterium tuberculosis*; heat shock protein; homology modeling

Background: Leprosy is a chronic infectious disease caused by *Mycobacterium leprae*. *M. leprae* remains one of the major pathogenic bacteria causing health problems worldwide particularly in developing countries. *M. leprae* can not be cultivated in vitro; however, it survives and proliferates within host macrophage cells by escaping its bactericidal activities, as well as in other cells. In order to understand the immunopathological mechanism of the pathogen and to develop effective vaccine candidates, many molecular biological studies have been undertaken to identify and characterize the immunodominant antigenic proteins of *M. leprae* [1-3]. Among them, an 18kDa antigen, a member of the small heat shock proteins, is known to be specific to *M. leprae*. The 18-kDa gene is transcriptionally activated during intracellular growth in macrophages and might be involved in the survival of *M. leprae* within the macrophages [4].

Traditionally, sHSPs have been grouped into five major families. They were designated HSP 100, HSP 90, HSP 70, HSP 60 and small HSPs according to their molecular masses [5, 6]. Small heat shock proteins are a ubiquitous and diverse family of stress proteins that have in common an alpha crystallin domain. They form large homologeric complexes and often exhibit a high degree of dynamic subunit exchange, which might be involved in their chaperone function [7, 8]. Previously it is shown that residues 70-88 in alphaA-crystallin can function like a molecular chaperone by preventing the aggregation and precipitation of denaturing substrate proteins. The peptide sequence corresponds to β3 and β4 region is present in the alpha crystallin domain of sHSP16.5 [9]. The crystallin subunits and mini-alphaA crystallin were able to suppress thermal aggregation of citrate synthase at 43°C [10]. Residues 73-92 (DRESVNLDVKSFPEELKVK), is the functional chaperone site of alphaB-crystallin which is known as mini-alphaB-crystallin [11]. Small heat shock proteins (sHSPs) are a superfAMILY of proteins with a molecular weight <40 kDa ubiquitously found in a variety of organisms [12]. Like all sHSPs, sHSP18 also share a conserved central domain of ~90 aminoacids called alpha-crystallin domain and have N terminal region and C terminal extension. *M. leprae* 18kDa heat shock protein gene is polymorphic. A single nucleotide polymorphism was detected at 15α position in this secreted antigen gene. In this gene, codon 52 exists as TCA in about 60% of the samples and CCA in rest of the leprosy cases. Armadillo derived *M. leprae* sHSP18 gene has TCA at the 52nd position.

The sequence has been deposited at the NCBI databank with the accession number M19058 [13]. 18kDa gene encodes 148 amino acids polypeptide. It shows a point mutation in 52nd position of its amino acid sequence. Serine at 52nd position mutates to Proline. Both types of gene are distributed equally in *M. leprae* biopsy samples. Reason for this mutation is unknown. Even though it shows sequence similarity to small heat shock proteins, its
biological function and crystal structure is unknown. The C-terminal domain is a common structural core across the small heat shock protein super family and the common sequence characteristics are identified in a stretch of 80-100 amino acid residues generally located in the C terminal part of the sequence and referred to as the alpha crystallin domain [14, 15]. Alpha-crystallin constitutes one of the three major classes of structural protein of the eye lens crystallins. They are associated with chaperone like function [16]. A recent review dealt with some of the unique properties of alpha crystallins emphasizing aspects that we still do not know of the structure and function [17].

Three dimensional models corresponding to the C-terminal domain of human alpha A crystallin [18] and full length human alpha B crystallin have been proposed [19, 20]. The domain in alpha A crystallin was demonstrated to comprise an immunoglobulin like fold as originally proposed by Bork P et. al.; Mormon JP et. al. [21, 22] in which 2 beta sheets, one consisting 3 beta strands and the other consisting 4 beta strands pack face to face to form an aligned beta sandwich. Template wheat sHSP16.9 has 3 a-helices and 2 beta-sheets having 9 beta strands. The core of sHSP16.9 adopts an immunoglobulin like fold consisting of two beta-sheets that are packed as parallel layers. beta5 and beta4 form one beta-sheet and beta2, beta3, beta8 and beta9 together with beta6 of a neighboring subunit, form the other beta-sheet. The donated strand is located in the center of the a-crystallin domain. Each subunit in the complex makes extensive contacts with other subunits via hydrogen bonds as well as hydrophobic and ionic interactions. The short C terminal extension is oriented toward the outside of the shell and interacts with beta4 and beta8 of a neighboring subunit. Template M.tuberculosis shSP16.9 has 2 beta sheets having 8 beta strands. ACR1 is a 16.3kDa protein, which is one of two members of the shSP family found in M.tuberculosis. ACR1 is the most abundant protein in M.tuberculosis during its dormant, non-replicative phase but not present under the condition of logarithmic growth [23]. sHSP16.3 (ACR 1) is not heat shock responsive but accumulates in the transition to stationary phase, during hypoxia and infection of macrophages [24]. We intended to evaluate the secondary structures present in both models of sHSP18 and compare with corresponding secondary structures present in wheat sHSP16.9.

Methodology:
The crystal structure coordinates corresponding to the shSPs i.e shSP16.9 (PDB ID: 1GME_A & 1GME_C) and 16.3kDa (PDB ID: 2BYY_E) were obtained from the Protein Data Bank [25]. The sequences corresponding to the shSP18 were obtained from the SWISSPROT database [26, 27]. We used 29 N terminal aminoacids, 88 amino acids in alpha crystallin domain region and 31 C terminal amino acids for Modeling. The sequence alignment was based on Identity. The schematic pictures representing the three dimensional models were obtained using Insight II. Quality of models evaluated using PROCHECK program. Templates used were eukaryotic small heat shock protein shSP16.9 dimers [28] and M.tuberculosis ACR1 protein of 16.3kDa [29]. The shSP18 and shSP16.9 have 41.7% sequence similarity and 27.4% sequence identity. shSP18 and ACR1 16.3kDa have 36.2% sequence similarity and 23% sequence identity. Template sequences were aligned by superposition of their Ca carbons.

Computational details
We performed Homology modeling using the Insight II software (Accelrys Inc.) [30] and same were used to visualize, model, modify, manipulate, analyse molecular systems and related molecular data. A Silicon Graphics O2 workstation with an R12k processor running at 150MHz in an Irix6.5 operating system was used for all computational requirements. 3 modules were used for Homology modeling. The commands in the Biopolymer module facilitate the building and modification of peptides, proteins, polysaccharides and nucleic acids. In homology module sequences were extracted and assigned coordinates for both Structurally Conserved Region (SCR) and loops. For both sHSP18S and sHSP18P we assigned the same SCR. Multiple sequence alignment based on MULTALIN [31] (Figure 1 Multiple sequence alignment in MULTALIN) was done manually by homology module. The Discover program is accessed from within Insight II is used to minimize the energy in the structure. The program performs energy minimization, template forcing, torsion forcing, and dynamic trajectories and calculates properties such as interaction energies, derivatives, mean square displacements and vibrational frequencies. It provides tools for performing simulations under various conditions including constant temperature, constant pressure, constant stress, periodic boundaries and fixed and restrained atoms. The potential energy of a biomolecule can be plotted as a multi-dimensional grid, which can be considered more simply as a two-dimensional topographic map. While minimizing the molecule’s potential energy, it reaches the nearest minimum. Minimization algorithms calculate the derivative of the current point on the map, and then determine which way to “move” (i.e., move the atoms) to reach the minimum. The steepest descent algorithm (with line searches) is the most basic algorithm. Using this algorithm first, especially if the molecule is far from the minimum and until the derivative<=0.1. Then switch to another algorithm. A distance dependent dielectric constant, Morse potential and cross terms is inactive and stresses, periodic boundaries and fixed and restrained atoms.

Discussion:
Secondary structure elements in shSP18 were compared with secondary structure elements in wheat shSP16.9 (Figure 2 Secondary structure and secondary structure elements in wheat shSP16.9 and corresponding residues in shSP 18S and shSP 18P). beta6 strand present in wheat shSP16.9 is absent in both models of shSP18. This result is consistent with the earlier observation (The region around beta6, which is intimately involved in monomer interactions, is either extremely variable or even absent among α-shSPs, depending on the gap positioning in a given sequence alignment [33]). N terminal region of shSP18 contains one α helix. N terminal region is required for chaperone activity [34]. N and C-terminal motifs in human alpha B-crystallin play an important role in the recognition, selection and solubilization of substrates [35].
N-terminal region of sHSP16.3 has dual function of self-oligomerisation and substrate binding [36]. Alpha crystallin domain of sHSP18 has two β-sheets, one consisting three β strands and the other consisting four β strands pack face to face to form an aligned beta sandwich. This result is consistent with earlier observations [22].

![Multiple sequence alignment in MULTALIN](image1)

**Figure 1:** Multiple sequence alignment in MULTALIN [Red color represents highly conserved residue; Blue color represents weakly conserved residue; A position with no conserved residue is represented by a dot in the consensus line; Symbols:! represents either I or V;# represents any one residue of NDQEBZ].

![Secondary structure and secondary structure elements in wheat sHSP16.9 and corresponding residues in sHSP 18S and sHSP 18P](image2)

**Figure 2:** Secondary structure and secondary structure elements in wheat sHSP16.9 and corresponding residues in sHSP 18S and sHSP 18P.
Structural differences were observed in sHSP18 due to single point mutation from serine (18S) to proline (18P) in 52\textsuperscript{nd} position of its amino acid sequence. In sHSP 18S, three-residue turn is present in the beginning of N-terminal region (3\textsuperscript{rd} to 5\textsuperscript{th} positions). This is absent in sHSP18P. Number of amino acid residues contributing the alpha helix at N terminal region varies in both. In sHSP 18S, 12 residues (7\textsuperscript{th} to 18\textsuperscript{th}) are present, where as in sHSP 18P, eight residues (7\textsuperscript{th} to 14\textsuperscript{th}) are present. In sHSP 18S next to alpha helix turn with two residues (21-22) is present. In sHSP 18P next to alpha helix, turn with three residues (15-17) is present. Serine-52 is present in 52-58 positions. Where as, in 18P \beta strand is present in 55-57 positions. In sHSP 18S, two-residue turn is present in 59-60 position, where as in sHSP 18P this is absent. In sHSP 18S, \beta strand is present in 61-67 position, where as in sHSP 18P \beta strand is present in 61-65 positions. Structural overlay of both the models were shown in (Figure 3 Structural Overlay of sHSP18P on sHSP18S).

The number of residues contributing \alpha helix at N-terminal region varies in both models. In sHSP 18S more number of residues are present when compared to sHSP 18P. An earlier observation that alpha B-crystallin has a greater content of \alpha-helices and is more hydrophobic than alpha A-crystallin \cite{37,38} indicating that hydrophobic interactions play an important role in substrate interaction. The loop regions between \beta and \beta strands of both models varies in number of residues present in the loop. This region is important in substrate binding. The deduced substrate-binding site of sHSP16.5 maps in a loop that links \beta3 and \beta4. The crystal structure indicates that this loop is surface exposed and therefore well suited for protein-protein interactions. Several residues in this loop are involved in inter subunit contacts \cite{33,39}. Number of residues contributing \beta strand in both models varies. In sHSP 18S model serine-52 is present in \beta4 strand. \beta4 strand is significant since in the crystal structures of Methanococcus jannaschii sHSP16.5 and in wheat sHSP16.9 the \beta4 and \beta8 strands provide an interface on the surface of the alpha-crystallin domain for self association into complexes \cite{27,33}. Also \beta4-\beta8 groove is an ATP interactive site in the alpha-crystallin core domain of the small heat shock protein, human alpha B-crystallin .The functional peptide sequence in alpha crystallin corresponds to \beta3 and \beta4 region present in the alpha crystallin domain of sHSP16.5 \cite{9}. In our model the structure with serine in 52\textsuperscript{nd} position corresponds to \beta3 and \beta4 region present in the mini alpha peptide sequence, where as structure with proline in 52\textsuperscript{nd} position corresponds to \beta3 region present in mini alpha peptide sequence but \beta4 region varies in number of residues. Variation is illustrated (Table 1 in supplementary material). Quality of models evaluated using PROCHECK program

Conclusion:
N-terminal region and alpha crystallin domain of 18 sHSP models varies due to single point mutation of serine to proline at 52\textsuperscript{nd} position. Major functional peptide region of \alpha crystallin domain of both the models also varies, i.e. sHSP18S is having similarity to human \alphaA crystallin where as sHSP18P differs from it. These studies will allow us to explore the biological significance of this protein in the process of pathogenesis since it is highly immunogenic and is produced early in infection.

References:
Table 1: Mini alpha peptide sequence in human alpha crystallin and corresponding region in sHSP16.9, sHSP18S and sHSP18P.