

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

Elegikil Abdul Azeez Rehna<sup>1</sup>, Sanjeev Kumar Singh<sup>2,\*</sup> and Kuppamuthu Dharmalingam<sup>1</sup>

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

received August 20, 2008; revised November 11, 2008; accepted December 16, 2008; published December 31, 2008

## 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 ACRI 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 52<sup>nd</sup> position causes structural changes. The structural changes are observed in N-terminal region and alpha-crystalline domains. Serine in 52<sup>nd</sup> position is observed in  $\beta$ 4 strand and Proline in 52<sup>nd</sup> position is observed in loop. The number of residues contributing  $\alpha$  helix at N-terminal region varies in both models. In 18S more number of residues is present in  $\alpha$  helix when compared to 18P. The loop regions between  $\beta$ 3 and  $\beta$ 4 strands of both models vary in number of residues present in it. Number of residues contributing  $\beta$ 4 strand in both models vary.  $\beta$ 6 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 homo oligomeric 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  $\beta$ 3 and  $\beta$ 4 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 (DRFSVNLVDVKHFSPEELKVK), 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 154<sup>th</sup> 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 52<sup>nd</sup> 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 52<sup>nd</sup> position of its amino acid sequence. Serine at 52<sup>nd</sup> 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 deals 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.; Mornon 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  $\alpha$  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.  $\beta$ 7,  $\beta$ 5 and  $\beta$ 4 form one  $\beta$ -sheet and  $\beta$ 2,  $\beta$ 3,  $\beta$ 8 and  $\beta$ 9 together with  $\beta$ 6 of a neighboring subunit, form the other  $\beta$ -sheet. The donated strand is located in the center of the  $\alpha$ -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  $\beta$ 4 and  $\beta$ 8 of a neighboring subunit. Template *M.tuberculosis* sHSP 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; 2BYU\_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

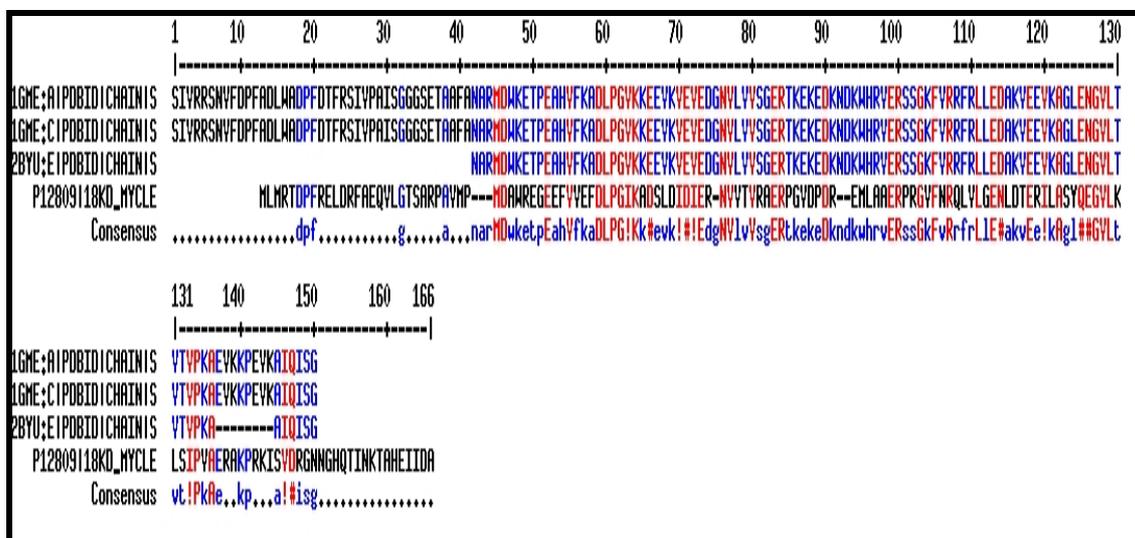
We performed Homology modeling using the Insight II software (Accelrys Inc.) [30] and same were used to visualize, model, modify, manipulate, analyse molecular systems and related molecular data. A Silicon Graphics O2 workstation with an R12k processor running at 150MHz in an Irix6.5 operating system 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 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 charges in active condition were used for minimization with steepest descent method for 100 runs and then conjugate-gradient energy minimization steps were used for further 2000 runs. Force field used was CVFF [32].

### Discussion:

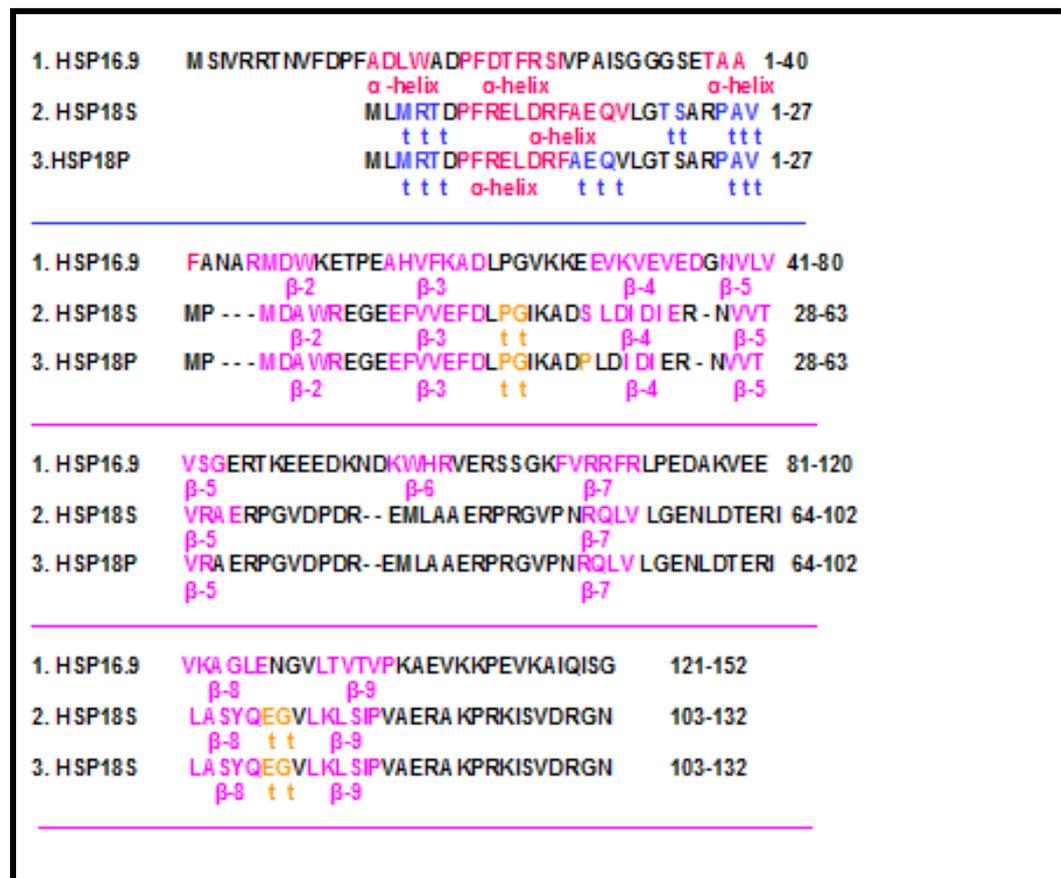
Secondary structure elements in sHsp18 were compared with secondary structure elements in wheat sHSP16.9 (Figure 2 Secondary structure and secondary structure elements in wheat sHSP16.9 and corresponding residues in sHSP 18S and sHSP 18P).  $\beta$ 6 strand present in wheat sHSP16.9 is absent in both models of sHSP18. This result is consistent with the earlier observation (The region around  $\beta$ 6, which is intimately involved in monomer interactions, is either extremely variable or even absent among  $\alpha$ -sHSPs, depending on the gap positioning in a given sequence alignment [33]). N terminal region of sHSP18 contains one  $\alpha$  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  $\beta$  sheets, one

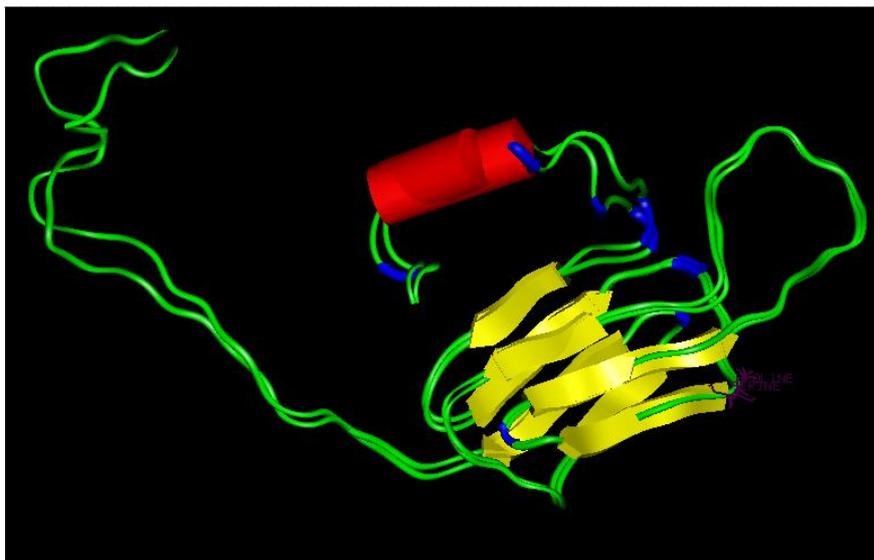
consisting three  $\beta$  strands and the other consisting four  $\beta$  strands pack face to face to form an aligned beta sandwich. This result is consistent with earlier observations [22].



**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].



**Figure 2:** Secondary structure and secondary structure elements in wheat sHSP16.9 and corresponding residues in sHSP 18S and sHSP 18P.



**Figure 3:** Structural Overlay of sHSP18P on sHSP18S.

Structural differences were observed in sHSP18 due to single point mutation from serine (18S) to proline (18P) in 52<sup>nd</sup> position of its amino acid sequence. In sHSP 18S, three-residue turn is present in the beginning of N-terminal region (3<sup>rd</sup> to 5<sup>th</sup> 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<sup>th</sup> to 18<sup>th</sup>) are present, where as in sHSP 18P, eight residues (7<sup>th</sup> to 14<sup>th</sup> positions) 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  $\beta$ 4 strand. Proline-52 is present in loop. In sHSP 18S  $\beta$  strand 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 [37, 38] indicating that hydrophobic interactions play an important role in substrate interaction. The loop regions between  $\beta$ 3 and  $\beta$ 4 strands of both models vary 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  $\beta$ 3 and  $\beta$ 4. 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 [33, 39]. Number of residues contributing  $\beta$ 4 strand in both models varies. In sHSP 18S model serine-52 is present in  $\beta$ 4 strand.  $\beta$ 4 strand is significant since in the crystal structures of *Methanococcus jannaschii* sHSP16.5 and in wheat sHSP16.9 the  $\beta$ 4 and  $\beta$ 8

strands provide an interface on the surface of the alpha-crystallin domain for self association into complexes [27, 33]. Also  $\beta$ 4- $\beta$ 8 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  $\beta$ 3 and  $\beta$ 4 region present in the alpha crystallin domain of sHSP16.5 [9]. In our model the structure with serine in 52<sup>nd</sup> position corresponds to  $\beta$ 3 and  $\beta$ 4 region present in the mini alpha peptide sequence, where as structure with proline in 52<sup>nd</sup> position corresponds to  $\beta$ 3 region present in mini alpha peptide sequence but  $\beta$ 4 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<sup>nd</sup> position. Major functional peptide region of  $\alpha$  crystallin domain of both the models also varies, i.e. sHSP18S is having similarity to human  $\alpha$ A 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:

- [01] S. W. Hunter *et al.*, *J Biol Chem.*, 265: 14065 (1990) [PMID: 2201679]
- [02] B. Rivoire *et al.*, *Infection and Immunity*, 62: 2417 (1994) [PMID: 7910593]
- [03] M. C. V. Pessolani and P. J. Brennan, *Infection and Immunity*, 64: 5425 (1996) [PMID: 8945601]
- [04] O. A. Dellagostin *et al.*, *Microbiology*, 141: 1785 (1995) [PMID: 7551043]
- [05] J. Buchner, *Federation of American Societies for Experimental Biology Journal*, (1996) 10: 10 [PMID: 8566529]

- [06] B. Bukau and A. L. Horwich, *Cell*, 92: 351 (1998) [PMID: 9476895]
- [07] R. van Montfort *et al.*, *Advances in Protein Chemistry*, 59: 105 (2001) [PMID: 11868270]
- [08] F. Sobott *et al.*, *J Biol Chem.*, 277: 38921 (2002) [PMID: 12138169]
- [09] K. K. Sharma *et al.*, *J Biol Chem.*, (2000) 275: 3767 [PMID: 10660525]
- [10] P. Santhosh Kumar and K. K. Sharma *et al.*, *Molecular Vision*, 7: 172 (2001) [PMID: 11483893]
- [11] J. Bhattacharya *et al.*, *Biochemistry*, 45: 3069 (2006) [PMID: 16503662]
- [12] G. Wistow *et al.*, *FEBS Letters*, 181: 1 (1985) [PMID: 3972098]
- [13] R. J. Booth *et al.*, *J. Immunology*, 140: 597 (1988) [PMID: 2447183]
- [14] W. W. de Jong *et al.*, *Molecular Biology and Evolution*, 10: 103 (1993) [PMID: 8450753]
- [15] W. W. de Jong *et al.*, *International Journal of Biological Macromolecules*, 22: 151 (1998) [PMID: 9650070]
- [16] J. Horwitz *et al.*, *Methods in Enzymology*, 290: 365 (1998) [PMID: 9534176]
- [17] J. Horwitz *et al.*, *Experimental Eye Research*, 76: 145 (2003) [PMID: 12565801]
- [18] H. A. Koteiche and H.S. Mehaorab, *J Mol Biol.* 294: 561 (1999) [PMID: 10610780]
- [19] P. J. Muchowski *et al.*, *J Mol Biol.* 289: 397 (1999) [PMID: 10366513]
- [20] K. Guruprasad and K. Kumari, *International Journal of Biological Macromolecules*, 33: 107 (2003) [PMID: 14599592]
- [21] P. Bork *et al.*, *J Mol Biol.* 242: 309 (1994) [PMID: 7932691]
- [22] J. P. Mornon *et al.*, *International Journal of Biological Macromolecules*, 22: 219 (1998) [PMID: 9650076]
- [23] Y. Yuan *et al.*, *Proceedings of the National Academy of Sciences USA*, 95: 9578 (1998) [PMID: 9689123]
- [24] Y. Yuan *et al.*, *J Bac.*, 178: 4484 (1996) [PMID: 8755875]
- [25] H. M. Berman *et al.*, *Nucleic Acids Research*, 28: 235 (2000) [PMID: 10592235]
- [26] A. H. Nerland *et al.*, *J Bac.*, 170: 5919 (1988) [PMID: 3056923]
- [27] B. Boeckmann *et al.*, *Nucleic Acids Research*, 31: 365 (2003) [PMID: 12520024]
- [28] R. L. van Montfort *et al.*, *Nat Struct Biol.*, 8: 1025 (2001) [PMID: 11702068]
- [29] K. Christopher *et al.*, *J Biol Chemistr.* (2005) 280: 33419 [PMID: 16046399]
- [30] Insight II, version 2.3.0. *Accelrys Inc., San Diego*, (1993)
- [31] Corpet *et al.*, *Nucleic Acids Research*, (1988) 16: 10881 [PMID: 2849754]
- [32] P. Dauber-Osguthorpe *et al.*, *Proteins: Structure Function and Genetics*, 4: 31 (1988) [PMID: 3054871]
- [33] K. K. Kim *et al.*, *Nature*, 394: 595 (1998) [PMID: 9707123]
- [34] S. Studer *et al.*, *European Journal of Biochemistry*, 269: 3578 (2002) [PMID: 12135498]
- [35] J. Ghosh *et al.*, *Biochemistry*, 45: 13847 (2006) [PMID: 17105203]
- [36] X. Fu *et al.*, *J Biol Chemistry*, 280: 6337 (2005) [PMID: 15545279]
- [37] S. A. Datta and C. M. Rao, *J Biol Chemistry*, 274: 34773 (1999) [PMID: 10574947]
- [38] T. X. Sun *et al.*, *J Biol Chemistry*, 272: 6220 (1997) [PMID: 9045637]
- [39] P. N. Farnsworth and K. Singh, *FEBS Letters*, 482: 175 (2000) [PMID: 11024455]

Edited by P. Kangueane

Citation: Rehna *et al.*, *Bioinformatics* 3(5): 230-234 (2008)

**License statement:** This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited.

### Supplementary material

Small heat shock proteins	Peptide sequence
sHSP 16.9	AHVFKADLPGVKKE <b>EVKVE</b> β3 β4
sHSP 18S	<b>EFVVEFD</b> LPGIKAD <b>SLDID</b> β3 β4
sHSP 18P	<b>EFVVEFD</b> LPGIKADPLD <b>ID</b> β3 β4
Human alpha A	<b>KFVIFLD</b> VKHFSPED <b>DLTVK</b> β3 β4

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