Journal of Proteomics & Bioinformatics - Open Access
www.omicsonline.com *Regular article* JPB/Vol.2/Special Issue 2008

In silico Analysis of Alkaline Shock Proteins in Enterobacteria

Seetharaaman Balaji [1,*] and V. Murali Krishnan [2]

[1] Lecturer, Department of Biotechnology,
Manipal Institute of Technology, Manipal University,
Manipal - 576104, India

[2] Lecturer, Department of Biochemistry,
PSG College of Arts and Science,
Coimbatore, India

[*]**Correspondence to:** S.Balaji, Lecturer, Department of Biotechnology, Manipal Institute of Technology, Manipal University, Manipal - 576104, India, Email: blast_balaji@rediffmail.com

In silico Analysis of Alkaline Shock Proteins in Enterobacteria

Abstract

Alkaline shock proteins (ASPs) adapt to atypical conditions, which are categorized into a family of small proteins. Alkaline shock proteins are identified in many bacteria and they are involved in stress response. The molecular basis for the survival of bacteria under extreme conditions in which growth is inhibited is a question of great current interest. A preliminary study was carried out to determine residue pattern conservation among the alkaline shock proteins of enteric bacteria, responsible for extreme alkaline tolerance especially in Staphylococcus and Streptococcus. To decipher, is there any secret hidden in the alkaline shock proteins? Bioinformatics approach was used and the molecular evidence proved the relationship between Staphylococcus and Streptococcus with respect to ASP. The sequence, structure and phylogenetic analyses inferred the relationships of various bacteria with respect to the conserved motif (VDNNKAK) of ASPs. Automated subsystem functional annotation of 172 homologous ASPs was done for various bacteria. Currently the structure of the ASP is not available in the Protein Data Bank (PDB). Since the Staphylococcus species formed the root of the phylogenetic tree, the structure of the Staphylococcus aureus (strain bovine RF122) was modeled in order to understand further about the structure and mechanism.

Keywords: Alkaline Shock Protein (ASP), Staphylococcus, Streptococcus, Structure, Phylogeny.

Introduction

Bacteria can switch its gene expression to adapt to various environments. Albeit the physiological role of alkaline shock proteins is not clear, it plays a key role in alkaline pH tolerance. Though such alkaline resistance mechanism is found in some bacteria, it is not the characteristic feature of all microbes. The response to alkaline pH can be characterized as alkaline shock (exposure to alkaline pH for up to 30 min) and alkaline adaptation (exposure to alkaline pH for periods of more than 60 min). Adaptation of Salmonella or E. coli to alkaline conditions is accompanied by induced thermotolerance, increased resistance to bile salts, and increased resistance to high pH (Taglicht et al., 1987; Flahaut et al., 1997; Goodson & Rowbury, 1990; Humphrey et al., 1991). Conversely, alkaline adaptation of S. enterica serovar enteritidis or E. coli sensitizes the cells to acid stress (Rowbury et al., 1993) and vice versa (Rowbury & Hussain, 1996). Staphylococcus aureus is a member of mammalian body surface normal flora and occasionally causes pyogenic infections, from simple skin suppuration to life-threating septicemia and may co-regulate the production of its virulence factors in response to environmental stresses such as heat shock, osmotic shock, and anaerobiosis (Mekalanos, 1992). Among these factors, pH is emerging as an important mean of regulating gene expression in S. aureus, because alkaline pH decreases expression of the Accessory Gene Regulator (agr) which affects the expression of numerous exoproteins, including E-hemolysin, toxic shock syndrome toxin 1, protein A, and the staphylococcal enterotoxins types B, C and D (Regassa & Betley, 1992). In order to clarify whether alkaline shock correlates with the expression of these exoproteins Kuroda et al. (1995) examined the effect of pH shift on the composition of cytosolic proteins in S. aureus. A protein of molecular mass of 23 KDa was remarkably enhanced by a pH up-shift from 7 to 10. This alkaline shock protein (ASP 23) was isolated and purified. The deduced primary sequence of ASP23 comprised 169 amino acids with a calculated molecular weight of 19,191. At least three sigma factors, sigA, (Deora & Misra, 1996) sigB, (Kullik & Giachino, 1997; Wu et al., 1996) and SA0492, (Morikawa et al., 2003) have been identified in S. aureus. Each sigma factor recognizes a different promoter sequence and allows the RNA polymerase to initiate site-specific transcription for a specific group of genes. Although the primary sigma factor, sigA, is constitutively maintained, the activity of the alternative sigma factor, sigB, depends on growth-phase and various environmental stresses, suggesting that staphylococcal sigB may regulate some stress responses (Bateman et al., 2001; Nair et al., 2003). The alkaline shock protein gene, asp23, is under the sole control of SigB and its expression is often used as an indicator of sigB activity (Giachino et al., 2001). The aim of this study is to understand the properties of alkaline shock proteins because the molecular mechanism of such pH tolerant properties should be elucidated because the production of the virulence factors was greatly affected by environmental pH (Kuroda et al., 1995). Since the protein structure for the ASP is not available in the Protein Data Bank (PDB) (Berman et al., 2000) and moreover currently no work has been done so far to predict the structure, it limits enthusiasm to understand the mechanism through the structure. Hence we took initiative and developed a structure for asp23 of Staphylococcus aureus, strain bovine RF122 (because of analogy to other bacterial pathogens) by prediction-based threading method.

Materials and methods

Sequence Data

The alkaline shock protein sequences were collected from UniProt Knowledgebase, SwissProt and TrEMBL (Table 1 under supplementary material).The key word 'alkaline shock protein' yielded 70 hits of protein sequences from SwissProt and TrEMBL (on the month of June, 2006) [URL http://www.expasy.org/]. There were only 9 sequences in SwissProt and 61 sequences in TrEMBL. In addition to that 175 putative proteins were also collected and annotated using National Microbial Pathogen Data Resource (NMPDR).

Table 1. ASP Sequence data used for analyses

Sl No	Gene/locus/ ORF name	Organism	Strain
1	Q5HE23	Staphylococcus aureus	COL
2	P0A0P6	Staphylococcus aureus	Mu50 / ATCC
3	P99157	Staphylococcus aureus	N315
4	Q6GEP7	Staphylococcus aureus	MRSA252
5	Q6G7D2	Staphylococcus aureus	MSSA476
6	P0A0P8	Staphylococcus aureus	
7	P0A0P7	Staphylococcus aureus	MW2
8	Q5HM47	Staphylococcus epidermidis	ATCC 35984 / RP62A
9	Q8CNG0	Staphylococcus epidermidis	ATCC 12228
10	A0NIZ2	Oenococcus oeni	ATCC BAA-1163
11	A0Q098	Clostridium novyi	NT
12	A0Q105	Clostridium novyi	NT
13	A1R1Z4	Arthrobacter aurescens	TC1
14	A3CQ96	Streptococcus sanguinis	SK36
15	A3CQR0	Streptococcus sanguinis	SK36
16	A3FQT7	Listeria monocytogenes	J0161
17	A3G3B4	Listeria monocytogenes	J2818
18	A3G9G5	Listeria monocytogenes	10403S
19	A4DFA3	Listeria monocytogenes	FSL N1-017
20	A4DMY2	Listeria monocytogenes	FSL N3-165
21	A4DVB7	Listeria monocytogenes	F6900
22	A4E3D7	Listeria monocytogenes	HPB2262
23	A4IM56	Geobacillus thermodenitrificans	NG80-2

23	A4IM56	Geobacillus thermodenitrificans	NG80-2
24	A5CMS9	Clavibacter michiganensis subsp. Michiganensis	NCPPB 382
25	A5CMX4	Clavibacter michiganensis subsp. Michiganensis	NCPPB 382
26	Q0SFG0	Rhodococcus sp.	strain RHA1
27	Q0SRZ9	Clostridium perfringens	SM101/Type A
28	Q0TPD2	Clostridium perfringens	ATCC 13124 / NCTC 8237 / Type A
29	Q1WUJ5	Lactobacillus salivarius subsp. Salivarius	UCC118
30	Q2BS41	Lactobacillus reuteri	JCM 1112
31	Q2BVJ2	Lactobacillus reuteri	JCM 1112
32	Q2FEV0	Staphylococcus aureus	USA300
33	Q2WHG9	Clostridium beijerincki	NCIMB 8052
34	Q2WHL5	Clostridium beijerincki	NCIMB 8052
35	Q2YXJ1	Staphylococcus aureus	bovine RF122
36	Q2YYG3	Staphylococcus aureus	bovine RF122
37	Q3DSG3	Streptococcus agalactiae	18RS21
38	Q3K0W3	Streptococcus agalactiae	serotype Ia
39	Q3K0W5	Streptococcus agalactiae	serotype Ia
40	Q48RF5	Streptococcus pyogenes	serotype M28
41	Q49ZC7	Staphylococcus saprophyticus subsp. Saprophyticus	ATCC 15305 / DSM 20229
42	Q4MFX1	Bacillus cereus	G9241
43	Q5FJG3	Lactobacillus acidophilus	
44	Q5FJI4	Lactobacillus acidophilus	
45	Q5LXU6	Streptococcus thermophilus	CNRZ 1066
46	Q5M2E8	Streptococcus thermophilus	ATCC BAA-250/LMG 18311

46	Q5M2E8	Streptococcus thermophilus	ATCC BAA-250/LMG 18311
47	Q5QGI3	Lactobacillus sakei	
48	Q5QGI4	Lactobacillus sakei	
49	Q5SLS8	Thermus thermophilus	HB8 / ATCC 27634 / DSM 579
50	Q67NA6	Symbiobacterium thermophilum	
51	Q67PP8	Symbiobacterium thermophilum	
52	Q7CER3	Streptococcus pyogenes	serotype M3
53	Q7CMS7	Streptococcus pyogenes	serotype M18
54	Q819V0	Bacillus cereus	ATCC 14579 / DSM 31
55	Q88WN0	Lactobacillus plantarum	
56	Q88Y60	Lactobacillus plantarum	
57	Q88Y61	Lactobacillus plantarum	
58	Q894G4	Clostridium tetani	
59	Q8CNG1	Staphylococcus epidermidis	ATCC 12228
60	Q8DR07	Streptococcus pneumoniae	ATCC BAA-255 / R6
61	Q8DW47 (putative)	Streptococcus mutans	
62	Q8XJD5 (probable)	Clostridium perfringens	
63	Q8XJM3 (probable)	Clostridium perfringens	

64	Q97HC7 B.subtilis ortholog	Clostridium acetobutylicum	
65	Q97IB6	Clostridium acetobutylicum	
66	Q99Y35	Streptococcus pyogenes	serotype M1
67	Q9R4V9 (fragment)	Campylobacter jejuni	
68	Q9R4W0(fragment)	Campylobacter jejuni	
69	Q9RSQ6	Deinococcus radiodurans	
70	Q5XA14	Streptococcus pyogenes	Serotype M6

Sequence Alignment

A multiple sequence alignment was done by using Clustal X Ver.1.83 (Thompson, et al., 1997), the gap opening was set at 10.00, the gap extension at 0.20 with 30% delay divergent sequences and Gonnet series weight matrix was used. From the multiple sequence alignment, the guide tree was derived. To justify the confidence of the clades, re-sampling method (bootstrap) was used with 10000 trails. Web logo (ver 2.8.2) was used to identify the conserved pattern in the ASPs of pathogenic bacteria. Alignments were analysed and phylogenetic relationships among the sequences were established using different procedures: Neighbour-Joining (NJ) (Saitou & Nei, 1987), Fast Minimum Evolution (FastME) (Desper & Gascuel, 2002), Unweighted Pair Group Method with Arithmetic Mean (UPGMA) (Sokal & Michener, 1958). The final tree was displayed by using MEGA 3.1 (Kumar, et al., 2004).

Phylogenetic Alignment

Alignments were analysed and phylogenetic relationships among sequences were established using different procedures: Neighbour-Joining (NJ), Fast Minimum Evolution (FastME), Unweighted Pair Group Method with Arithmetic Mean (UPGMA) and Fitch-Margoliash (FM). Trees and genetic distances were based on 10,000 replicates in order to assess the degree of confidence for each branch on the trees. Heuristic searches were completed with maximum parsimony. Absolute distances and pairwise distances were calculated for all pairwise combinations of operational taxonomic units (OTUs).

Protein Structure Prediction

Identification of structurally homologous proteins

In searching for structural homologues for alkaline shock protein (asp23) of Staphylococcus aureus (strain bovine RF122), the complete amino acid sequence of length 124aa (EMBL Protein: AJ938182 /TrEMBL: Q2YXJ1) was primarily submitted to Basic Local Alignment Tool (BLAST) (Altschul et al., 1990). Nevertheless, it does not yield any significant templates from Protein Data Bank (PDB). Therefore it was re-submitted to the PredictProtein server (Rost et al., 2004). This server returns a multiple sequence alignment and predictions of secondary structure, residue solvent accessibility, and the location of transmembrane helices. The secondary structure of asp was then threaded, using this information against proteins in the PDB. The PredictProtein program detects remote homologues (0–25% sequence identity) by a novel prediction-based threading method (Rost et al., 1997). To recognize folds by threading, the PredictProtein program evaluates the amino acid sequence of a protein and determines how well it fits into the 3D configuration of proteins whose structures are known. The goal is to detect similar motifs of secondary structure and accessibility between a sequence of unknown structure and a known fold. Proteins with known 3D structure and the highest degree of structural homology to asp23 were identified by Predict-Protein, which also provided summary information on these proteins from the server via e-mail.

Identification and alignment of structurally conserved regions (SCRs)

The multiple sequence alignment function in PredictProtein is automatically returned in the report from PredictProtein and is built up in two steps (Sander & Schneider, 1991). In sweep 1, sequences are aligned consecutively to the search sequence by a standard dynamic programming method. After each sequence has been added, a profile is compiled and used to align the next sequence. In sweep 2, after all sequences with significant structural homology have been selected from SWISSPROT, the profile is recompiled and the dynamic programming algorithm commences once again to align the sequences consecutively, this time using the conservation profile as derived after completion of sweep 1. The output consists of structurally homologous proteins with regions automatically aligned to asp23. In addition, the known and the predicted secondary structures of the PDB proteins and asp23 are shown. With this information, we manually highlighted areas of predicted secondary structure in asp23 that were identical to the known structural homologues.

Assignment of coordinates

DeepView (Swiss-PdbViewer) is tightly linked to SWISS-MODEL (Guex & Peitsch, 1997), an automated homology modeling server that was used in combination with the downloaded asp23 sequence. With these two programs it is possible to thread a protein primary sequence onto a 3D template and obtain an immediate feedback of how well the threaded protein will be accepted by the reference structure prior to submitting a request to build missing loops and refine side chain packing. The PDB files of the three best-fitting structural homologues of asp23 identified by PredictProtein were downloaded and individually manually aligned to asp23, according to the alignment suggested by PredictProtein. Secondary structures were predicted around the asp23 sequences by using nnPredict (Kneller et al., 1990) that was found to be structurally homologous to the other PredictProtein listed proteins. With Swiss-PdbViewer, manually and carefully adjusted the alignment and transmitted the project file to SWISS-MODEL through SwissModel optimise mode. Coordinates were then assigned based on the known reference protein structure (1IXH). All coordinates were transferred if the side chains of the reference and model proteins were at the same corresponding locations along the sequence of the structurally conserved region. However, if these locations differed, only the backbone coordinates were transferred and the side chain atoms were automatically replaced to preserve the asp23 model's residue types. These replaced residues were first aligned to the backbone of the original residue; the dihedral angles in common with the residue being replaced were also aligned. This allowed the conformation side chain to be preserved as much as possible.

Loop generation

Since the asp23 protein had structural homology to the known proteins, and gaps existed in between the alignment, loops had to be generated. This was done using the method described by Shenkin et al. (1987). Briefly, a conformational search with random settings of Phi and Chi angles was made in order to build a peptide backbone chain connecting two conserved peptide segments. A set of six distances was defined using two atoms in the start residue of the loop at the amino-terminal as well as two atoms at the carboxylterminal stop residue of the loop. These distances must meet specific criteria for the loop to be acceptably closed. The loops were generated by using the "Build Loop" option of the DeepView, which uses energy information (computed with a partial implementation of the GROMOS Force-Field (van Gunstern & Berendsen, 1977) and a mean force potential value (PP) computed from a "Sippl-like" mean force potential (Sippl, 1990). This process also used the "Scan Loop" option that gave the name of PDB files that contain a suitable loop, the chain identifier, the starting residue, the sequence of the possible fragment, and the resolution (in Å) at which the structure had been solved. The similarity score for the fragment was also computed from the PAM200 matrix. In addition to those, a clash score, and the number of residues from the source loop that have bad phi/psi angles were also obtained to sort the loops by energies or clashes to ease the process of identifying the best loop. Finally, an energy minimization was performed and the geometry of the loop was checked for proper chirality and steric overlap violations, accepting those conformations that close the loop.

Structure Validation

To assess the geometric correctness of the theoretical structure, the following programs were used; VERIFY 3D (Eisenberg *et al.*, 1997) for assessment of asp23 model with three-dimensional profiles, WHATIF (Vriend, 1990) to validate asp23 structure, PROCHECK (Laskowski *et al*, 1996) to check the stereochemical quality of asp23 and plots its overall and residue-by-residue geometry, WHATCHECK (Hooft *et al.*, 1996) to find out errors in asp23 structure. These programs checked the protein-specific bond lengths, angles, and torsions of the theoretical asp23 model. The parameters checked included phi-psi angles, chi1 dihedral angles, chi2 dihedral angles, main-chain parameters, side-chain parameters, residue properties, main-chain bond length and bond angle distributions, RMS distance from planarity and distorted geometry plots etc. This process not only assessed the geometric validity of the proposed structures, but also focused attention on problem areas in the structure.

Results and Discussion

Automated functional annotation was performed for 175 proteins by using the tools provided by the National Microbial Pathogen Data Resource (NMPDR) (McNeil, 2007) which was useful to study the comparative functional analysis of genomes and biological subsystems, with an emphasis on pathogenic species of *Campylobacter, Listeria, Staphylococcus, Streptococcus*, and *Vibrio*.

Graphical Representation of Protein Encoding Genes (PEGs)

In the graphical representation of the gene's structure, blue genes are located within 6 kb upstream or 6 kb downstream of the green gene of focus in at least four other species. Whether or not the gene of interest appears to be clustered in its genome, homologs of this gene may occur in clusters in other genomes. The organisms in which the homologous gene were clustered with others, ordered by the size of the cluster. The neighboring proteins were also predicted and the numbers were computed. The functional clustering scores, which are approximately equal to the number of different species (not strains), were also predicted. The graphical representation of the homologous chromosomal clusters (Fig. 1) makes easy understanding of the gene architecture and illustrates functional clustering.

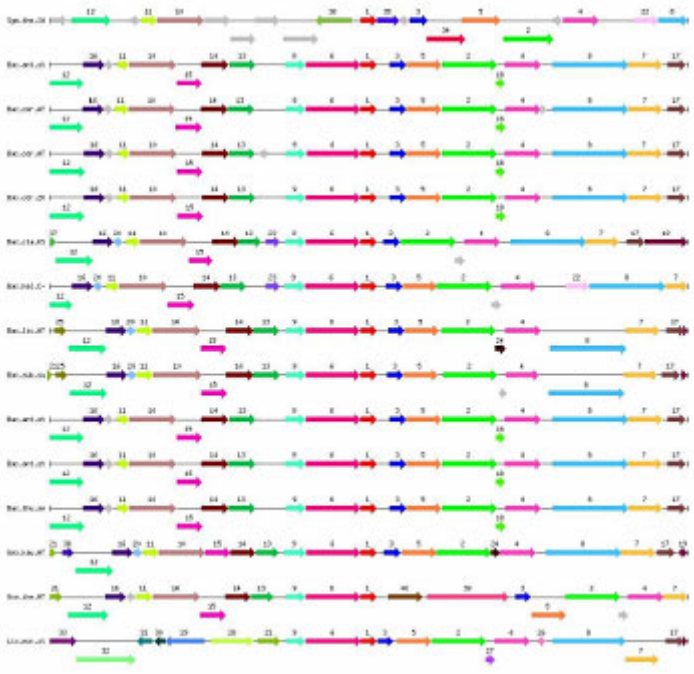


Fig.1 Homologous chromosomal clusters

Homologs of the focus peg are red, labeled 1, and aligned in the center of the page. All of the genes within about 8 kb of this central peg are shown. Numerical labels correspond to rank ordered frequency of co-localization with the focus protein-encoding gene (peg). Non-homologous proteins and RNAs are grey, although sometimes very small proteins that share real homology are grey because the homology score, which is a function of chain length, is below the cut-off imposed on this figure. The

homologous pegs were compared with UniProt and NMPDR annotations. The strategy for finding genes was based on chromosomal clustering and occurrence profiling. The abbreviations used to represent bacteria in the chromosomal clusters (Fig.1) are Sym.the.IA (Symbiobacterium thermophilum), Bac.ant.st (Bacillus anthracis), Bac.cer.ZK (Bacillus cereus strain ZK), Bac.cla (Bacillus clausii), Bac.hal (Bacillus halodurans), Bac.lic (Bacillus licheniformis), Bac.sub (Bacillus subtilis), Bac.thu (Bacillus thuringiensis), Geo.Kau (Geobacillus kaustophilus), Oce.ihe (Oceanobacillus iheyensis), Lis.mon (Listeria monocytogenes).

Sequence analysis

A preliminary multiple sequence alignment was carried out among all ASPs of all available sequence of gram positive bacteria and a few gram negative bacteria as well. Multiple sequence alignment of the ASPs showed many conserved residue patterns in a regular interval at the N-terminal region (Fig.2). As the alignment approached the C-terminal end, the number of conserved residues was gradually decreased (Fig. 3), indicating that the N-terminal region of this protein has a much active role when compared to the carboxyl terminal end. The motif VDNNKAK is considered to play an important role, sequence alignment analysis confirmed that the motif VDNNKAK was well conserved within the entire ASPs of different bacteria (at the amino terminal). It proved that the VDNNKAK motif is not only unique for Staphylococcus aureus but also for the other bacteria such as Arthrobacter sp., Streptococcus sp., Listeria sp., Rhodococcus sp., Clostridium sp., Lactobacillus sp., Clostridium sp., Bacillus sp., Lactobacillus sp., Symbiobacterium sp., Streptococcus sp., Lactobacillus sp., Clostridium sp., Staphylococcus sp., Streptococcus sp., Deinococcus radiodurans, Geobacillus thermodenitrificans, Clavibacter michiganensis and Oenococcus oeni.

We also extended our analysis to address the pattern conservation among the other ASPs involved in alkaline resistance mechanism and found that the pattern is still partially conserved. This pattern conservation also depicts that the function is highly dependant on the pattern used for the alkaline resistance. The strong motif conservation could be the reason for the extreme alkaline resistance of the analysis dataset. The pattern was also searched against the PROSITE database and found match with all Staphylococcus sp.

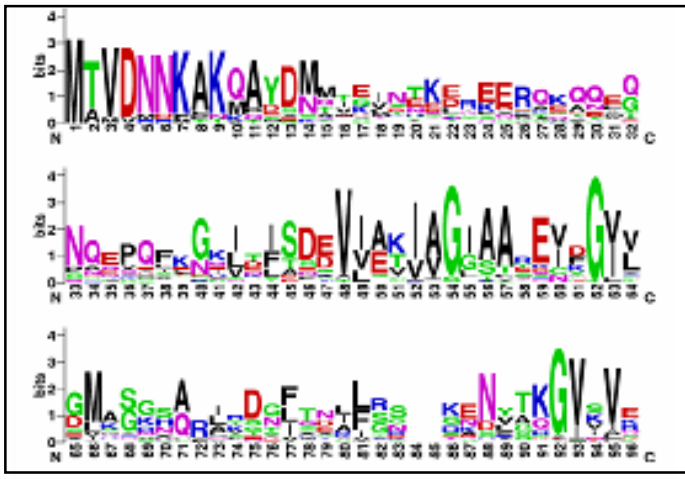


Fig.2

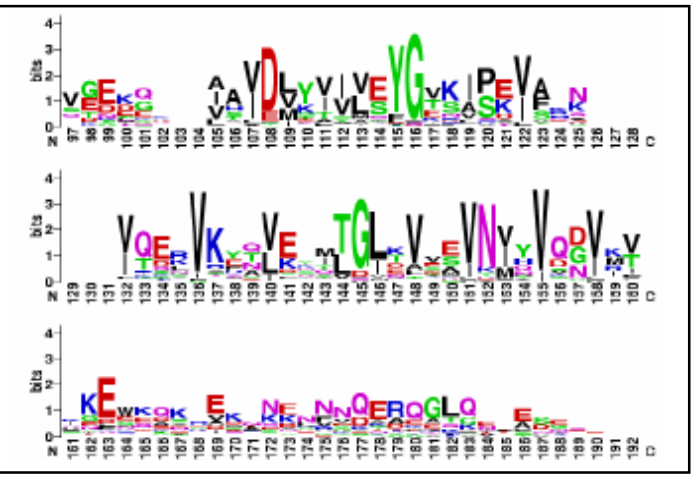


Fig.3

Phylogenetic Analysis

Phylogenetic trees were built based on the multiple sequence alignment. The trees were constructed with 10000 bootstrap trials (Fig. 4) shows that the ASPs of similar species form individual clusters which deviate from other ASPs. The similar trend was also observed in phylogenies obtained by using different methods (NJ, UPGMA and FastME). The ASP sequence of *Staphylococcus aureus* was found to be the root for all other bacterial ASPs. The convergence of different *Staphylococcus aureus* (P0A0P8) strains such as strain COL (Q5HE23), strain Mu50 / ATCC 700699 (P0A0P6), strain N315 (P99157), strain MRSA252 (Q6GEP7), strain MSSA476 (Q6G7D2), strain MW2 (P0A0P7), strain USA300 (Q2FEV0) and strain bovine RF122 (Q2YYG3) was also observed. From Fig. 5, it is clear that *Staphylococcus aureus* and their strains form the root of the tree (0.0) to all operational taxonomical units (OTUs). All other bacteria were dispersed (diverged) equally from the origin. Minimum degree of divergence from the root corresponds to the *Streptococcus* group (0.090910) such as *Streptococcus pyogenes* serotype M3 (Q7CER3), *Streptococcus mutans* (Q8DW47) and *Streptococcus pyogenes* serotype M28 Q48RF5. The species epidermidis (strain ATCC 35984 / RP62A and strain ATCC 12228) of the genus *Staphylococcus* (Q5HM47 and Q8CNG0) showed little divergence from the *Staphylococcus* group (0.156630). The species *saprophyticus* of the genus *Staphylococcus* (Q49ZC7), *Staphylococcus saprophyticus subsp. Saprophyticus* (strain ATCC 15305 / DSM 20229) was also deviated (0.224810) from the *Staphylococcus* group. The rest of all bacteria diverge from the root in the range of (0.6 to 0.94). The root (0.0) and the closely related cluster (0.09 - 0.22) have the conserved VDNNKAK motif. The conservation decreases slightly with respect to other clusters or distantly related /orthologous ASPs might be due to the interruption of indels in the ASP coding gene. The pattern analysis shows the strong relationship of *Staphylococcus* and *Streptococcus*, they are members of the group of potentially pathogenic bacteria (PPB). This finding is also supported by the phylogenetic analysis (Fig. 6) of ASP cluster which is congruent with the high degree of identity between *Staphylococcus* and *Streptococcus* groups, shows the close pathogenic association and the alkaline tolerance of the two genuses.

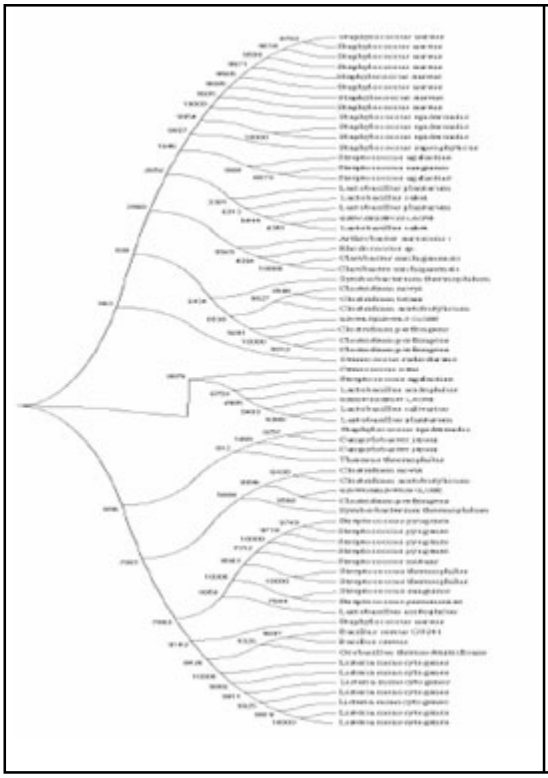


Fig.4

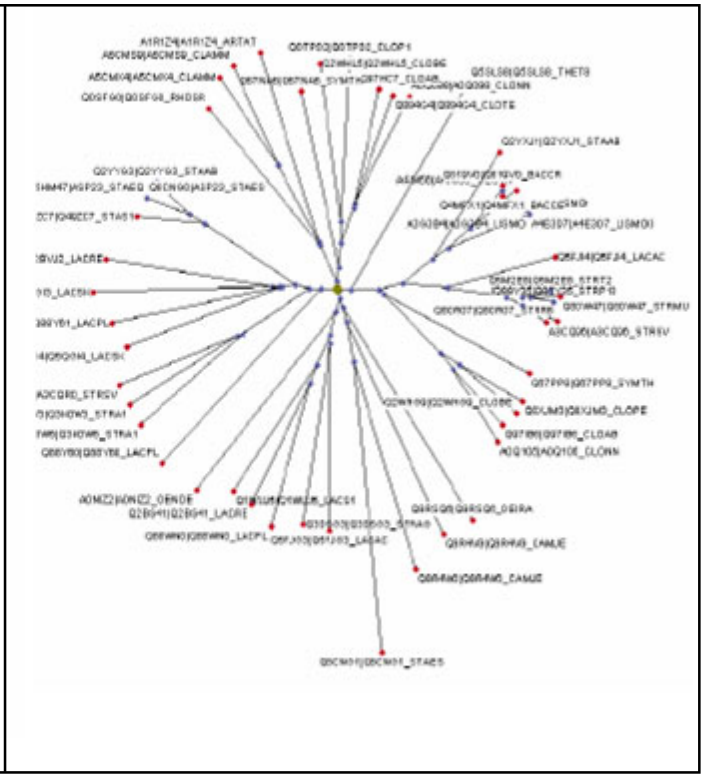


Fig.5

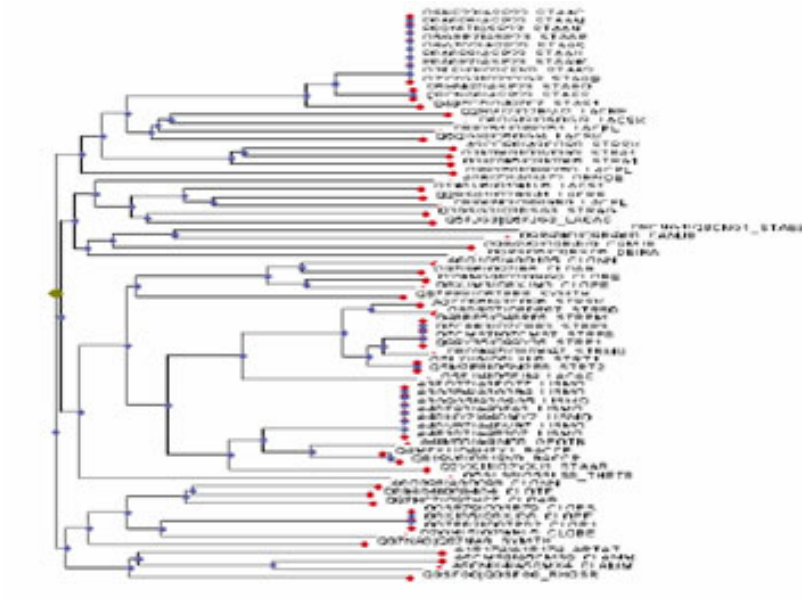


Fig. 6

The phylogenetic tree of alkaline shock proteins from various bacterial strains. The bootstrap values are shown at the branch points, 10000 replicates were done in order to access the confidence of the clades.

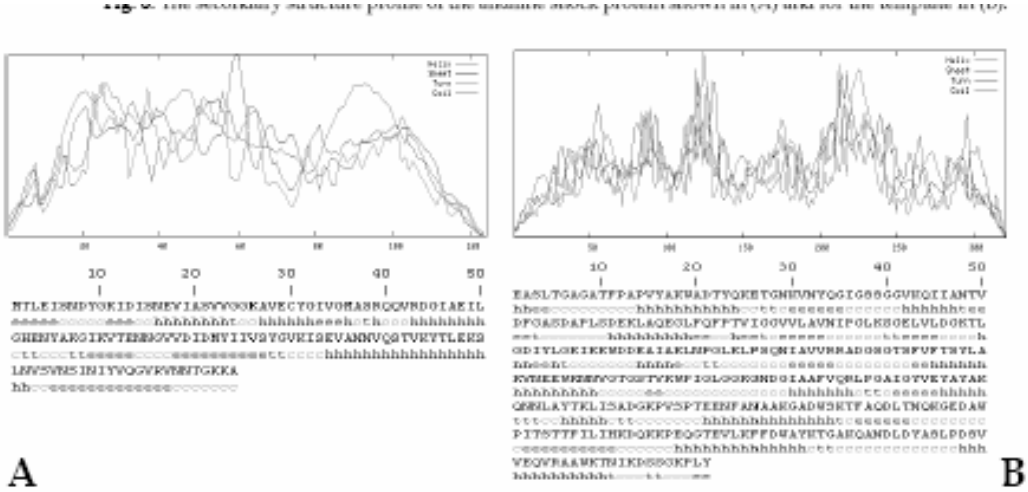
The Radial Phylogram of alkaline shock proteins from various bacterial strains which are dispersed (diverged) equally from the root (*Staphylococcus aureus*).

The Phylogram of alkaline shock proteins from vari-

ous bacterial strains. The branch length were evaluated with 10000 replicates shows the equal divergence of the clades.

Protein Structure Prediction

The phylogenetic analysis proved that the *Staphylococcus* serves as a root for all ASPs of the analysis data set. Hence the ASP of *Staphylococcus aureus* (strain bovine RF122) has been modeled by using prediction based threading, the complete amino acid sequence of length 124aa (EMBL Protein: AJ938182 / TrEMBL: Q2YXJ1) was primarily submitted to Basic Local Alignment Tool (BLAST). Nevertheless, it does not yield any significant templates from Protein Data Bank (PDB). Therefore it was re-submitted to the PredictProtein server. The secondary structure of the ASP was then threaded, using this information against proteins in the PDB. The PredictProtein program detects remote homologues by a novel prediction-based threading method. To recognize folds by threading, the PredictProtein program evaluates the amino acid sequence of a protein and determines how well it fits into the 3D configuration of proteins whose structures are known. The Predict Protein program identified 20 closest structural homologues from threading-based prediction TOPITS (Threading One-dimensional Predictions In to Three-Dimensional Structures) and provided a z score for each. This score is highly dependent on the similarity of characteristics including alignment length, compositions of secondary structure, and accessibility of amino acids between the protein of known 3D structure and the protein of interest. The higher the z score, the higher the probability that the first hit is correct. In general, an alignment z score (ZALI) with z > 3 is more reliable. Threading one-dimensional predictions in to three-dimensional structures (TOPITS) yielded twenty different templates. Nevertheless, based on the ZALI score first three templates were chosen. The first three ranks of protein templates to model ASP are Phosphate-Binding Protein [PDB: 1ixh], GTP cyclohydrolase [PDB: 1a8r], and OMPF porin from *E.coli* [PDB: 2omf]. Only the first template was considered as a better template based on the statistical significance. By using PIR pairwise alignment the template was found to have a Smith-Waterman score of 64 and have 32.231% identity with the template. The reliability was compared by ZALI score of the templates, and in addition a better template with higher ZALI score (2.51) among the scored templates was selected and used for this modeling procedure. In the absence of the best template, we have selected the first scored template [PDB: 1BK0]. Homology models were constructed using a combination of the Swiss-PDB viewer (Deep View) and the Swiss-Model online modeling server (project-optimize mode) for assigning atomic coordinates. Prediction of the secondary structure for ASP is based on the sequence and structure comparison taken into account. In order to validate our method to predict the secondary structure, we have first used the 1ixh (template) as a test sequence. Using the software SOPMA (Geourjon & Deleage, 1995) we obtain correctly predicted structural elements. Such prediction results provide confidence with respect to the reliability of the ASP structure prediction. The secondary structure elements of the template and the structural prediction of the ASP are shown in Fig. 7. The comparison between the secondary structure elements observed in the template crystal structure and the prediction made for ASP (Fig. 8) shows that beta strands and alpha helices are almost perfectly superimposed, though some residues of the template are not aligned because of the large sequence length. The alignment shows 32.231% sequence identity, with insertions and deletions (Indels) primarily localized in loops.



	Secondary structure	A	B
1	Alpha helix (Hh)	34.68%	34.27%
2	Extended strand (Ee)	31.45%	18.69%
3	Beta turn (Tt)	6.45%	7.17%
4	Random coil (Cc)	27.42%	39.88%

41

Fig.7

The secondary structure profile of the alkaline shock protein shown in (A) and for the template in (B).

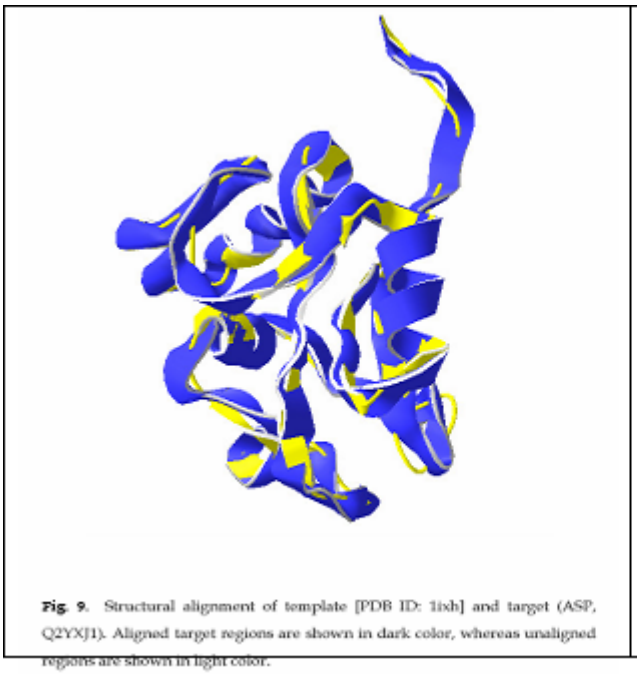


Fig. 9. Structural alignment of template [PDB ID: 1ixh] and target (ASP, Q2YXJ1). Aligned target regions are shown in dark color, whereas unaligned regions are shown in light color.

Fig.8.

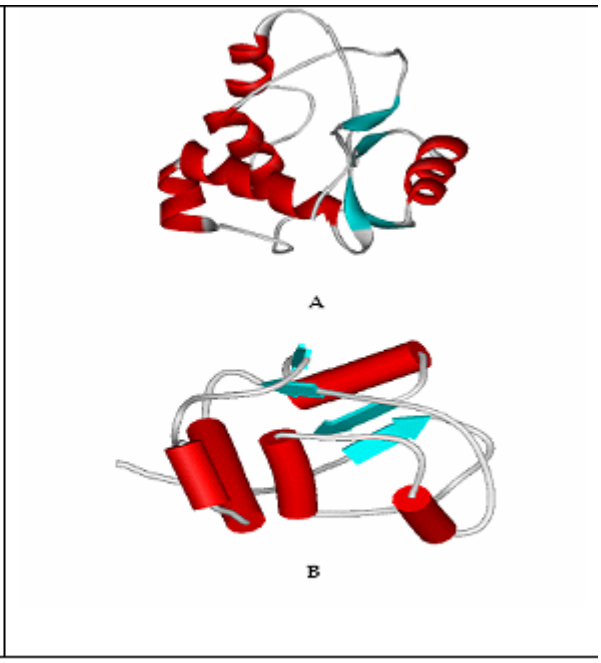
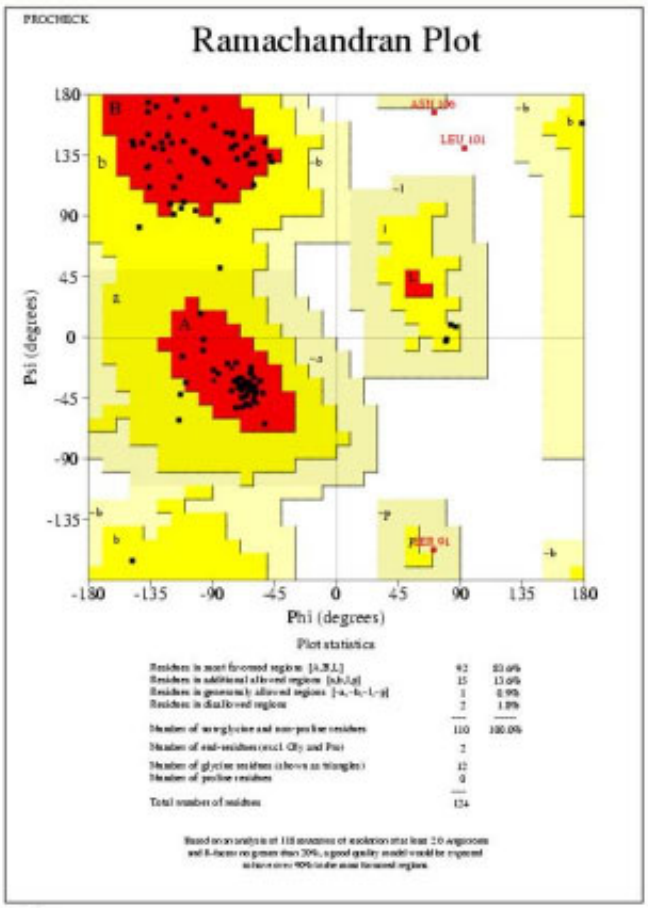


Fig.9.

The alignment of the better template [PDB: 1ixh] with ASP (target) was done and the positions of the experimentally observed secondary structure elements of the template and of the predicted secondary structure of ASP were superimposed on the aligned sequences. The structures are depicted in the Fig. 9. The aligned template [1ixh] with target [ASP] was used to build and refine the ASP model. The final model (Fig. 9) was iteratively minimized for energy and subjected to structure verification and evaluation. The Sasisekaran - Ramakrishnan - Ramachandran diagram (or simply "Ramachandran plot") of PROCHECK (Laskowski *et al*, 1996) showed 83.6% residues in most favored regions with 1.8% residues in disallowed regions (Table 2). The Ramachandran plot (Fig. 10) shows the phi-psi torsion angles for all residues in the structure (except those in chain termini).



Residues in most favoured regions [A,B,L]	92	83.6%
Residues in additional allowed regions [a,b,l,p]	15	13.6%
Residues in generously allowed regions [~a,~b,~l,~p]	1	0.9%
Residues in disallowed regions	2	1.8%
Number of non-glycine and non-proline residues	110	100.0%
Number of end-residues (excl Gly and Pro)	2	
Number of glycine residues(shown as triangles)	12	
Number of proline residues	0	
	124	

Table 2.

Glycine residues are separately identified by triangles. The shading on the plot represents the different regions described in Morris *et al.*, (1992). The darkest areas correspond to the "core" regions representing the most favourable combinations of phi-psi values. The modeled structure has 83.6% of the residues in the core region. The model can be compared with other bacterial ASP models (if available) in order to find the structural divergence especially the VDNNKAK motif.

Structural alignment of template [PDB ID: 1ixh] and target (ASP, Q2YXJ1). Aligned target regions are shown in blue color, whereas unaligned regions are shown in yellow color.

(A) The final model of the alkaline shock protein (Q2YXJ1) in ribbon model (B) Schematic representation.

The Ramachandran plot of the modeled alkaline shock protein

Conclusion

Multiple sequence alignment of all available ASPs showed many conserved residue patterns at regular intervals at the N-terminal region. It was observed that when the alignment approaches towards the C-terminal, there is a decline in the number of conserved residues, indicating that the N-terminal region of this protein has much active role when compared to the carboxyl terminal. The motif, VDNNKAK is well conserved in the entire ASP at the amino terminal. The motif is also partially conserved among other distantly related bacteria but involved in alkaline resistance mechanism. Phylogenetic cluster analysis proves the relationship of *Staphylococcus* and *Streptococcus* with other bacteria. The model presented here shows for the first time details of the unique structural features of ASP. This model should be used with caution, because of the lack of x-ray diffraction data. Moreover in our predicted protein, positions of atoms are far from precise, and geometries are often non-ideal, so an x-ray structure of the ASP with high resolution may reveal new, unexpected features of ASP. The predicted model will be compared with the actual crystal data when available. The characterization of the ASP binding site could shed light and insight into an interesting perspective to predict mechanism of interaction with other biomolecules. The production of the virulence factors was greatly affected by environmental pH. Inhibitors of this regulatory protein are not currently available. Therefore, this model would be helpful to initiate structure related studies, which provide guidance for rational drug discovery.

Acknowledgments:

This work was supported by Bharathidasan University, Trichirappalli, Tamilnadu, India, as a part of my M.Phil in Bioinformatics.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990). Basic local alignment search tool. *J. Mol. Biol.*, 215:403.

Bateman, B. T., Donegan, N. P., Jarry, T. M. et al. (2001). Evaluation of a tetracycline-inducible promoter in Staphylococcus aureus in vitro and in vivo and its application in demonstrating the role of sigB in microcolony formation. Infection and Immunity 69, 7851–7.

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000). The Protein Data Bank. Nucleic *Acids Research*, 28:235

Deora, R. & Misra, T. K. (1996). Characterization of the primary sigma factor of Staphylococcus aureus. Journal of Biological Chemistry 271, 21828–34.

Desper R. & Gascuel O. (2002). Fast and accurate phylogeny reconstruction algorithms based on the minimum-evolution principle. *Journal of Computational Biology*, 9:687

Eisenberg D, Lüthy R, Bowie JU (1997). VERIFY3D: assessment of protein models with three-dimensional profiles. *Methods Enzymol.*, 277:396-404.

Flahaut, S., A. Hartke, J.-C. Giard, and Y. Auffray. 1997. Alkaline stress response in Enterococcus faecalis: adaptation, cross-protection, and changes in protein synthesis. Appl. Environ. Microbiol. 63:812–814.

Geourjon C & Deleage G (1995). SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Comput. Appl.Biosci.*, 11:681.

Giachino, P., Engelmann, S. & Bischoff, M. (2001). [B activity depends on RsbU in Staphylococcus aureus. Journal of Bacteriology 183, 1843–52.

Goodson, M., and R. J. Rowbury. 1990. Habituation of alkali and increased UV resistance in DNA repair-proficient and -deficient strains of Escherichia coli grown at pH 9.0. Lett. Appl. Microbiol. 11:123–125.

Guex N. & Peitsch M.C. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis*, 18:2714.

Hooft RW, Vriend G, Sander C, Abola EE (1996). Errors in protein structures. *Nature*, 381:272.

Humphrey, T. J., N. P. Richardson, A. H. L. Gawler, and M. J. Allen. 1991. Heat resistance of Salmonella enteritidis PT4: the influence of prior exposure to alkaline conditions. Lett. Appl. Microbiol. 12:258–260.

Kneller DG, Cohen FE, Langridge R (1990). Improvements in protein secondary structure prediction by an enhanced neural network. *J. Mol. Biol.*, 214:171

Kullik, I. I. & Giachino, P. (1997). The alternative sigma factor sigmaB in Staphylococcus aureus: regulation of the sigB operon in response to growth phase and heat shock. Archives of Microbiology 167, 151–9.

Kumar S, Tamura K, Nei M (2004). MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform.*, 5: 150

Kuroda, M., Ohta, T. and Hayashi, H. (1995). Isolation and gene cloning of an alkaline shock protein in methicillin resistant Staphylococcus aureus. Biochem. Biophys. Res. Commun. 207, 978–984.

Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol.NMR.*, 8:477.

McNeil LK, Reich C, Aziz RK, Bartels D, Cohoon M, Disz T, et al. (2007) The National Microbial Pathogen Database Resource (NMPDR): A genomics platform based on subsystem annotation. Nucleic Acids Res. 35: D347-D353. Epub 2006 Dec 1.

Mekalanos, J.J. (1992) Environmental signals controlling expression of virulence determinants in bacteria. J. Bacteriol. 174 (1), 1-7.

Morikawa, K., Inose, Y., Okamura, H. et al. (2003). A new staphylococcal sigma factor in the conserved gene cassette: functional significance and implication for the evolutionary processes. Genes to Cells 8, 699–712.

Nair, S. P., Bischoff, M., Senn, M. M. et al. (2003). The sigma B regulon influences internalization of Staphylococcus aureus by osteoblasts. Infection and Immunity 71, 4167–70.

Regassa L B and Betley M J (1992). Alkaline pH decreases expression of the accessory gene regulator (agr) in Staphylococcus aureus. J Bacteriol. 174 (15): 5095–5100.

Rost B, Schneider R, Sander C (1997). Protein fold recognition by prediction-based threading. *J. Mol. Biol.*, 270:471.

Rost B, Yachdav G, Liu J. (2004). The PredictProtein server. *Nucleic Acids Research*, 32 (Web Server issue):W321

Rowbury, R. J., and N. H. Hussain. 1996. Exposure of Escherichia coli to acid habituation conditions sensitizes it to alkaline stress. Lett. Appl. Microbiol. 22:57–61.

Rowbury, R. J., M. Goodson, and T. J. Humphrey. 1993. Novel acid sensitivity induced in Escherichia coli at alkaline pH. Lett. Appl. Microbiol. 16:223–227.

Saitou N and Nei M. (1987).The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4:406.

Sander C & Schneider R. (1991). Database of homology-derived protein structures and the structural meaning of sequence alignment. *Proteins*, 9:56.

Shenkin PS, Yarmush DL, Fine RM, Wang HJ, Levinthal C (1987). Predicting antibody hypervariable loop conformation. I. Ensembles of random conformations for ringlike structures. *Biopolymers*, 26:2053

Sippl J.M. (1990). Calculation of conformational ensembles from potentials of mean force. An approach to the knowledge-based prediction of local structures in globular proteins. *J. Mol. Biol.*, 213:859-883.

Sokal, R.R. & C.D. Michener (1958). A statistical method for evaluating systematic relationships. Univ. of Kansas Scientific Bulletin. 38: 1409–38

Taglicht, D., E. Padan, A. B. Oppenheim, and S. Schuldiner. 1987. An alkaline shift induces the heat shock response in Escherichia coli. J. Bacteriol. 169:885–887.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*, 24: 4876.

van Gunsteren W.F. & Berendsen H.J.C (1977). Algorithms for macromolecular dynamics and constraint dynamics. Mol. Phys. 34:1311-1327.

Vriend G (1990). WHAT IF: a molecular modeling and drug design program. *J. Mol. Graph.*, 8:52.

Wu, S., Lencastre, H. & Tomasz, A. (1996). Sigma-B, a putative operon encoding alternate sigma factor of Staphylococcus aureus RNA polymerase: molecular cloning and DNA sequencing. Journal of Bacteriology 178, 6036–42.

