Fall 12-2008

The Role of MSA in the Global Regulation of Virulence in *Staphylococcus aureus*

Vijayaraj Nagarajan
University of Southern Mississippi

The University of Southern Mississippi

THE ROLE OF MSA IN THE GLOBAL REGULATION OF VIRULENCE IN STAPHYLOCOCCUS AUREUS

by

Vijayaraj Nagarajan

Abstract of a Dissertation
Submitted to the Graduate Studies Office
of The University of Southern Mississippi
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy

December 2008

The University of Southern Mississippi

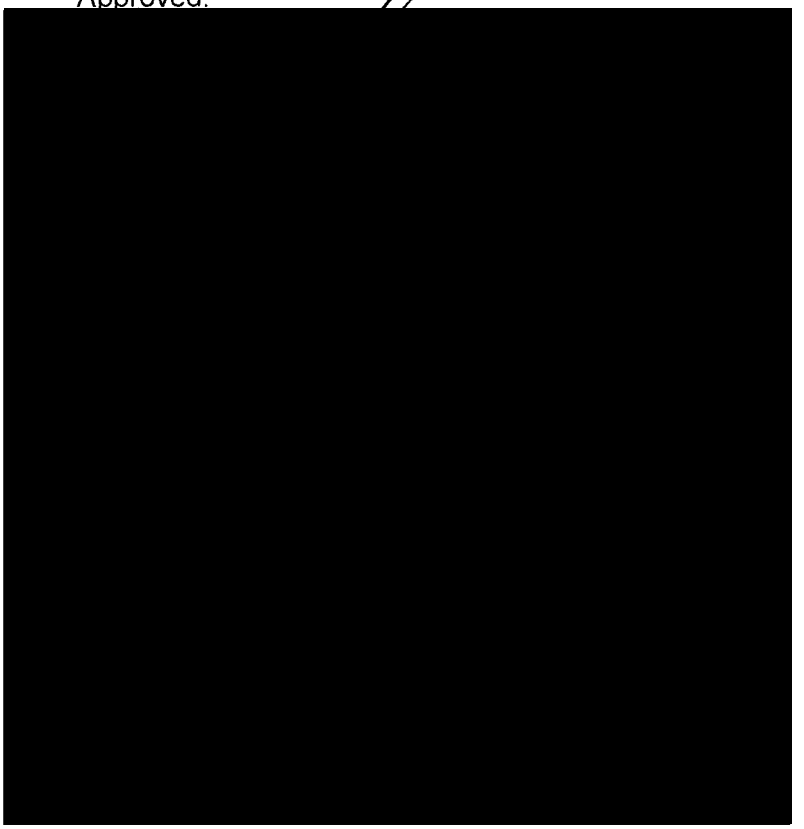
THE ROLE OF MSA IN THE GLOBAL REGULATION OF VIRULENCE IN STAPHYLOCOCCUS AUREUS

by

Vijayaraj Nagarajan

A Dissertation
Submitted to the Graduate Studies Office
of The University of Southern Mississippi
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy

Approved:



December 2008

ABSTRACT

ROLE OF MSA IN THE GLOBAL REGULATION OF VIRULENCE IN STAPHYLOCOCCUS AUREUS

by Vijayaraj Nagarajan

December 2008

Staphylococcus aureus is an important pathogen causing life threatening diseases in humans. Previously we showed that *msa* modulates the activity of *sarA* (Staphylococcal accessory regulator), which is one of a major global regulator of virulence in *S. aureus*. The objective of this study is to characterize the role of *msa* (Modulator of SarA) in the global regulation of virulence in *S. aureus*. Structure and function predictions were done using several computational tools and approaches to understand the nature of *msa*. A novel *S. aureus* microarray meta-database (SAMMD) was designed and developed to compare and contrast other transcriptomes with *msa* transcriptome. *msa* and *sar*A transcriptomes were generated using the microarray technology. Phenotypic and molecular assays were performed to support microarray results. The results show that *msa* is a putative transmembrane protein, with three transmembrane segments, a distinct N-terminal cleavable signal peptide, four phophorylation sites (two outside and two inside the membrane) and a binding site in the cytoplasmic region. Microarray results and comparative transcriptome analysis using SAMMD showed that several genes regulated by *msa* are also regulated by *sarA*. Based on these results I hypothesize that *msa* is a novel signal transducer, which modulates the activity of genes involved in virulence in a *sarA*-dependent manner, while modulating the activity of genes involved in metabolism in a *sarA*-independent manner.

ACKNOWLEDGMENTS

I sincerely thank my advisor Dr. Mohamed O. Elasri for giving me this wonderful opportunity. His thoughtful guidance and passionate encouragement led me to march successfully throughout this mission of my life time – a doctoral degree.

I thank Dr. Shearer, Dr. Wang, Dr. Murali and Dr. Guo for serving on my dissertation committee. Their insightful comments and suggestions did make me do better research. I also thank Mississippi Functional Genomics Network for the financial assistance. Last but not least, I thank my family and friends for their patience and support.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF ILLUSTRATIONS

Figure

CHAPTER I
INTRODUCTION

Objective

The Objective of my project is to decipher "The role of *msa* (Modulator of SarA) in global regulation of virulence in *Staphylococcus aureus*". *S. aureus* is a human pathogen that causes several diseases ranging from superficial skin infections to more serious manifestations like endocarditis and osteomyelitis. The arsenal of *S. aureus* consists of a wide variety of virulence factors controlled by a several global regulators like *sarA* (staphylococcal accessory regulator) and *agr (*accessory gene regulator). We earlier showed that *msa* modulates the activity the global virulence regulator *sarA* (96). Through *sarA*, *msa* could modulate several virulence factors. The proof of concept for this argument was provided by Sambanthamoorthy et al. (96), by showing that genes regulated by *sarA* are also modulated by *msa.* Since, our preliminary analysis showed that *msa* is a putative membrane protein, we hypothesized that *msa* could also affect other genes and regulators independent of *sarA* activity. Thus, my objective is to study the role of *msa* in global regulation of virulence in *S. aureus*.

Description of the Proposed Research

To study the role of *msa* in global regulation of virulence in *S. aureus*, I pursued the following goals:

1. Structure and function prediction for the *msa* protein.
2. Design and development of the *Staphylococcus aureus* Microarray Meta-Database (SAMMD).
3. Development of *msa* and *sarA* mutant transcriptomes.
4. Analysis of *msa* and *sarA* mutant transcriptomes, with SAMMD.
5. Molecular and Phenotypic assays to confirm microarray results and conclusions.

Significance of the Proposed Research

I expect my research to significantly contribute towards discovering novel drug targets and therapeutics in treating various Staphylococcal diseases. Apart from significant findings related to *msa*'s contribution to global regulation of virulence, resources (like SAMMD) that I developed during this project would also be of immense use for "Staph" researchers from through out the world. The proposed study on global regulatory network analysis using SAMMD would also throw light on some missing links in understanding the metabolism of *S. aureus*.

CHAPTER II

MSA PROTEIN STRUCTURE AND FUNCTION PREDICTIONS

Background

Introduction

Staphylococcus aureus is an important human pathogen that causes several diseases ranging from superficial skin infections to life-threatening diseases such as osteomyelitis and endocarditis. *S. aureus* is capable of infecting a wide range of tissues in humans because of the large number of virulence factors and the complex regulatory networks that control them (83). In addition, *S. aureus* is increasingly resistant to multiple antibiotics thus becoming a growing threat to public health. There is an urgent need to understand the complex regulatory networks used by *S. aureus* to cause disease. Regulatory networks are attractive therapeutic targets for future treatment of antibiotic resistant infections.

Modulator of sarA (msa)

One of the important global regulators of virulence in *S. aureus* is the Staphylococcal accessory regulator (*sarA*) (29). *sarA* regulates over 100 genes in *S. aureus* several of which are associated with virulence (40). *sarA* plays an important role in disease (9). *sarA* itself is regulated by several loci that modulate its function. We recently identified a novel gene, *msa,* that modulates the function of *sarA* (96). We showed that *msa* is essential for full expression of *sarA* and that mutation of *msa* affected the expression of several virulence factors in both *sarA*-dependent and *sarA*-independent manners (96). Microarray analyses of the *msa* mutant show that *msa* has a global effect on genes in *S. aureus* (unpublished data). These studies indicate that *msa* is an important locus in *S. aureus* and that the characterization of the Msa protein would be very useful in understanding Staphylococcal regulatory networks.

Computational Tools

Several bioinformatics tools have been developed to predict the structure and functional properties of bio molecules. These tools use a wide variety of algorithms to predict the properties of proteins at different levels (87, 93). The accuracy of these bioinformatics tools has been improving; however, each tool has its own advantages and disadvantages. A particular algorithm has its own characteristic specificity, sensitivity, robustness, computational cost, etc. These characteristics can be tested against benchmarks of known datasets (e. g., Critical Assessment of Techniques for Protein Structure Prediction – CASP). In order to make the most accurate predictions, several methods should be used to build a consensus.

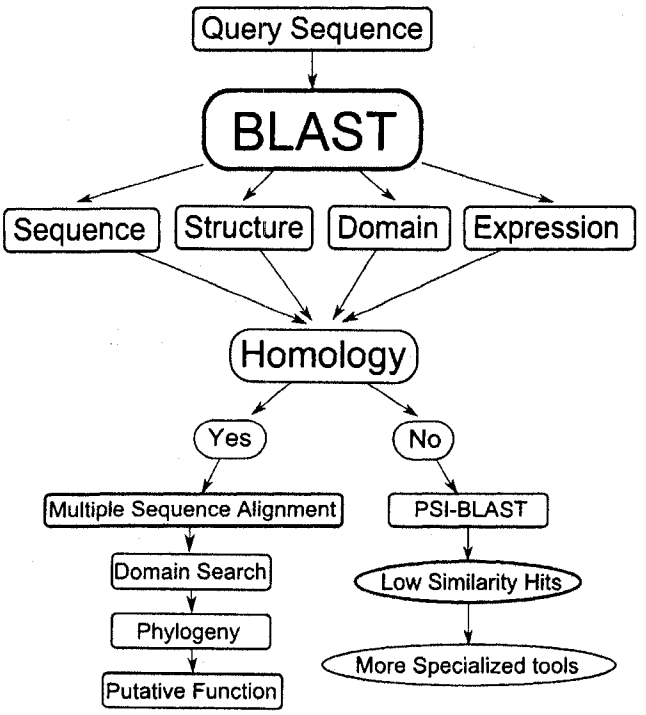


Figure 1. An overview of the approach used for predicting the structure and function of MSA protein.

The aim of this work is to predict the structure and functional properties of the Msa protein of *S. aureus* to the highest possible accuracy. Our choice of tools and the approach for making predictions and conclusions is illustrated in Figure 1.

The results show that the Msa is a putative integral membrane protein with three probable transmembrane regions. We also predict that the Msa contains phosphorylation sites in the loop regions (both inside and outside the membrane). The 3-D structure analysis of the Msa also predicts the presence of putative binding sites. Thus, based on this computational analysis, and previous experimental data (96) we hypothesise that Msa might play a role in signal transduction. The fact that Msa has no known homolog means that it would be a novel signal transducer.

Predictions and Conclusions

Primary Sequence Analysis

We used the protein sequence (Accession ID: NP_374514) obtained by conceptual translation of the *msa* open reading frame from the *S. aureus* N315 genome (NCBI database). The primary sequence analysis was performed using ProtParam, ProtScale (46) and SAPS (12). ProtScale was used to predict the Msa profile based on several amino acid scales. ProtParam computes properties like molecular weight, theoretical pI, instability index and grand average of hydropathicity (GRAVY). SAPS predict significant features of protein sequences like charge-clusters, hydrophobic regions, compositional domains etc.

The conceptually translated Msa protein is made of 133 amino acids with a predicted molecular weight of 15.6571 kDa and an isoelectric point (pI) of 6.71. The GRAVY index value 1.021 shows that Msa is probably an insoluble protein. The Codon adaptation index (CAI) value predicts the Msa as a highly expressed protein. This is consistent with experimental results described previously by our group (96).

Homology and Similarity

The Msa is highly conserved among the different strains of *S. aureus* (RF122, MRSA252, MSSA476, MW2, COL, Mu50, N315, and NCTC 8325). Even though there were several variations in the nucleotide sequences, we observed good conservation at the amino acid level. Multiple sequence alignment and phylogenetic analysis of both nucleotide sequences (SAUSA300_1294, SACOL1436, SAOUHSC_01402, SAV1401, *msa,* SAS1342, MW1289, SAR1413, and SAB1257c) and protein sequences (YP_493991, YP_186288, YP_499929, NP_371925, NP_374514, YP_043463, NP_646106) from different strains show that they are identical. The only two exceptions were strains RF122 and MRSA252 which showed slight variations in the Msa sequences. In RF122, the protein sequence (YP_416734) was 97% similar to the Msa sequence from N315 while in MRSA252, the protein sequence (YP_040815) was 98% similar to the Msa from N315. The phylogeny of the Msa protein closely resembled that of the phylogeny of these organisms as determined by Multi Locus Sequence Typing (MLST) (58). The position and effect of mutations in the Msa protein sequence of the strains MRSA252 and RF122 are discussed in the "3-D structure prediction and analysis" section.

Our similarity search results against several sequence and structure databases, using different BLAST programs, showed that there were no significant closely related homologs for the Msa protein, except for one in *S. epidermidis.* Even though there were no significant (based on E-value and score) homolog for Msa, BLAST also listed several membrane proteins with remote similarities (alignment Score of 32-35 and E values scores from 0.91 -10) only to the first few amino acids of the Msa protein (that corresponds to the predicted signal peptide region).

Sub-cellular Localization Predictions

The Sub-cellular localization and the functional categorization of Msa were predicted using ProtFun 2.2 (61), PSORT (81), ProtCompB V-3 (28), PRED-CLASS (89) and SVMProt (16). ProtFun uses *ab initio* methods to predict the cellular role category. PSORT uses a rule-based method to predict protein localization sites. ProtCompB combines several methods such as linear discriminant function-based predictions, direct comparison with homologous proteins of known localization, prediction of functional peptide sequences etc., to identify the sub-cellular localization of proteins. PRED-CLASS uses cascading neural networks to classify proteins in to different classes like membrane, globular, fibrous and mixed. SVMProt uses a support vector machine based approach to functionally classify protein sequences.

All the tools used to predict the cellular location of the protein indicated that Msa is a putative membrane protein. This prompted us to examine the sequence for presence of signal peptide and potential cleavage sites in the Msa protein sequence. Seven out of eight signal peptide prediction tools indicated the presence of a potential signal peptide in the Msa protein (Table 1). The majority of the programs also predicted an N-terminal cleavage site between the amino acid 19 and 20.

Signal Peptide and Topology Predictions

Signal peptide prediction was done using SignalP (7), PrediSi (55), sigcleave (106), PSORT (81), Phobius (66), SIG-Pred (54), SOSUIsignal (48) and iPSORT (5).

SignalP 3.0 uses artificial neural networks and hidden Markov models to predict signal peptides and their cleavage sites. PrediSi predicts signal peptide sequences and their cleavage sites based on a position weight matrix that also takes into consideration

Table 1

Signal peptide and cleavage position prediction for the Msa protein

	Signal Peptide	Cleavage Position
SignalP	Present	29
PrediSi	Present	29
sigcleave	Present	20
PSORT	Present	20
Phobius	Present	20
SIG-Pred	Present*	20
iPSORT	Present	No prediction
SOSUIsignal	Absent	No prediction
Consensus	Present	20

* Predicted as an eukaryotic signal peptide

the amino acid bias present in the proteins. Sigcleave is one of the early tools to predict the signal cleavage sites based on weight matrices. Sigcleave is distributed as part of the EMBOSS package. Phobius is a combined transmembrane protein topology and signal peptide predictor that uses a well trained hidden Markov model. SIG-Pred predicts signal peptides and their cleavage position based on weight matrices. SOSUIsignal uses a high performance system to predict signal peptides, using a three module software system that recognises the three-domain structure of signal peptides. iPSORT predicts the signal peptides based on a rule based system.

The Topology of Msa protein was predicted using TopPred (32), TMpred (57), PHDhtm (94), TMHMM (71), SPLIT (65), HMMTOP (104), MEMSAT (62), DAS (33) and TSEG (69). We also computed the charge bias of the generated models, based on the positive-inside rule (105). TopPred II predicts the topology of a protein based on its hydrophobicity profile and positive-inside rule. TMpred algorithm is based on the statistical analysis of TMbase, a database of naturally occurring transmembrane proteins, using a combination of several weight-matrices for scoring. PHDhtm uses a neural network based approach with the evolutionary information to predict the locations

of the transmembrane helices. TMHMM predicts transmembrane regions based on the hidden Markov model. SPLIT 4.0 predicts location of transmembrane helices by performing an automatic selection of optimal amino acid attribute and corresponding preference functions. HMMTOP 2.0 prediction is based on the hypothesis that the difference in the amino acid distributions in various structural parts determines the localization of the transmembrane segments. MEMSAT applies a novel dynamic programming algorithm to recognize membrane topology models by expectation maximization. DAS uses dense alignment surface method to predict transmembrane regions. TSEG uses a discriminant function to predict the transmembrane segments.

We used the NPS (Network Protein Sequence Analysis) consensus secondary structure server (39). This server runs the input sequence against several different secondary structure prediction tools and generates a consensus secondary structure out of them.

We performed topology analysis on the Msa sequence using several prediction programs that yielded widely discrepant results. Even though most programs failed to recognize the signal peptide, a consensus topology emerged (Table 2). The predicted topology of the Msa is IN-OUT with three putative transmembrane segments (from amino acid positions 27-47, 54-75, 108-125). The N-terminal is predicted as present in the cytoplasmic side of the membrane while the C-terminal is predicted as outside the membrane. Our consensus topology also passed the positive-inside rule and charge bias test (105), with a charge bias of +1 towards the inside of the membrane.

Table 2

Consensus topology for the Msa protein including the N-terminal signal peptide prediction

	N-terminal	TMS 1	TMS 2	TMS 3
TopPred	IN	27-47	-	106-126
TMHMM	IN	25-47	54-76	108-125
TMpred	IN	29-47	55-75	107-123
SPLIT	IN	27-47	54-69	107-126
HMMTOP	IN	28-47	60-77	108-125
MEMSAT	IN	30-47	54-70	108-125
PHDhtm	OUT	14-34	55-72	108-125
TSEG	No Prediction	24-47	54-75	105-128
DAS	No Prediction	27-44	54-67*	110-124
Consensus	IN	27-47	54-75	108-125

TMS, Transmembrane segment
* Probability not significant

Secondary structure prediction results indicated the presence of four distinct helical regions (Figure 2). One helical region corresponds to the cytoplasmic helix while the other three correspond to the integral membrane helices. These results are consistent with the predicted topology.

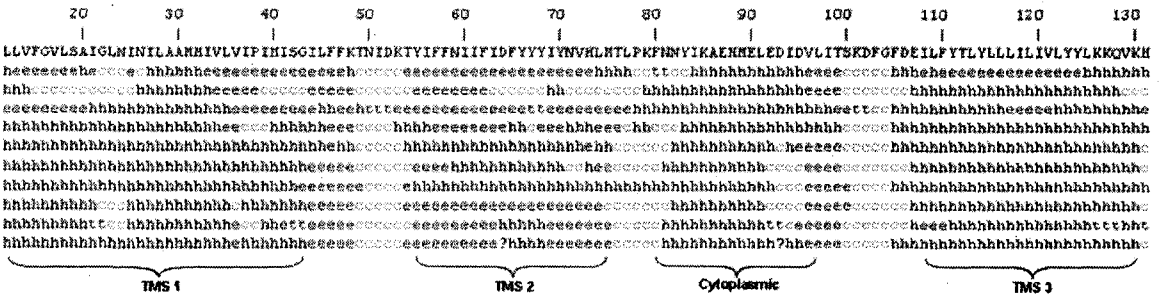


Figure 2. Consensus secondary structure predictions for the Msa protein. Three transmembrane segments (TMS) and a cytoplasmic helix are predicted.

Domains / Patterns / Motifs

SMART (Simple Modular Architecture Research Tool) (75) was used to identify the presence of any domains in the Msa protein. We used different pattern searching

applications (PPSearch (3), PSITE (100) and ScanProsite (38)) , that use PROSITE (59) database, to predict functionally relevant patterns in Msa protein.

We searched for the presence of domains, patterns and motifs in the Msa protein sequence, to gain insight into its functions and structure. The SMART results showed the presence of all the structural domains that we earlier identified using topology prediction programs and signal peptide prediction programs, viz. an N-terminal signal peptide and three transmembrane regions. In addition, SMART also predicted the presence of a PreATP-grasp domain (d1gsa_1) from the SCOP database. Even though this result had an E-value of 1.5, it was interesting because the predicted domain is a putative binding domain and falls in the predicted cytoplasmic region of Msa (residues 85-116). Our pattern search in the Msa protein sequence, using different programs against the PROSITE database, gave similar results (except for PPSearch, which did not predict the Tyrosine kinase site at position 48), showing the presence of three putative phosphorylation sites (Table 3).

Table 3

Prosite patterns predicted in the Msa protein

	Protein Kinase C	Casein Kinase II	Tyrosine Kinase
PPSearch	99	49, 99	-
PSITE	99	49, 99	48
ScanProsite	99	49, 99	48
Consensus	99	49, 99	48

Numbers denote residue position in the Msa protein sequence

All of the predicted sites were found in the exposed regions of the Msa. Analysis of the location of these putative phosphorylation sites showed that two of the putative phosphorylation sites are outside the membrane while one of them is predicted in the cytoplasmic region. We also observed that these putative phosphorylation sites are highly conserved among different strains of *S. aureus*. This suggests that Msa might be

phosphorylated by kinases in the cytoplasm as well as kinases on the outside of the membrane (e.g. from the host cells). These predictions further suggest that Msa might function as a signal transducer and provides important targets for mutagenesis experiments to test this hypothesis.

Membrane bound receptors are important components of signal transduction in all living systems. The major class of receptors in eukaryotes contains seven transmembrane segments (7 TM). Prokaryotes use 7 TM class receptors also, however, a recent study showed that prokaryotes carry novel receptor classes that have transmembrane segments ranging from one to eight (35, 44). The Msa protein sequence did not have significant homology with any of the known receptors and experimental studies are underway to evaluate its function as a signal transducer.

3-D Structure Prediction and Analysis

Initial attempts to predict the tertiary structure of Msa were done using different approaches like homology modelling, threading and *ab initio*. Automated homology modelling servers Swiss-Model (99) and ModWeb (95) were used for homology modelling. Predictions described in this study were done using fold recognition tools 123D+ (2), GenThreader (15), a new version of 3-D PSSM (Phyre) (68).

The quality of the predicted structure was examined using an online version of the WHATIF (107) program. Structure refinement was done using both WHATIF and Swiss-PDB Viewer (50). Structure visualization was done using Swiss-PDB Viewer.

The 3-D structure of the Msa protein was analysed for clefts and binding surfaces using ProFunc (73), Q-SiteFinder (74), PINUP (76) and SuMo (60).

Homology based tertiary structure prediction for the Msa protein failed, because of the lack of homologous structures. We used fold recognition based structure prediction server Phyre to model the tertiary structure of the Msa protein. Visualization and analysis of the predicted structure using Swiss-PDB Viewer (SPDBV) showed that

the predicted structure correlated with the other predicted structural features of Msa in terms of the number and positions of the transmembrane helices (Figure 3).

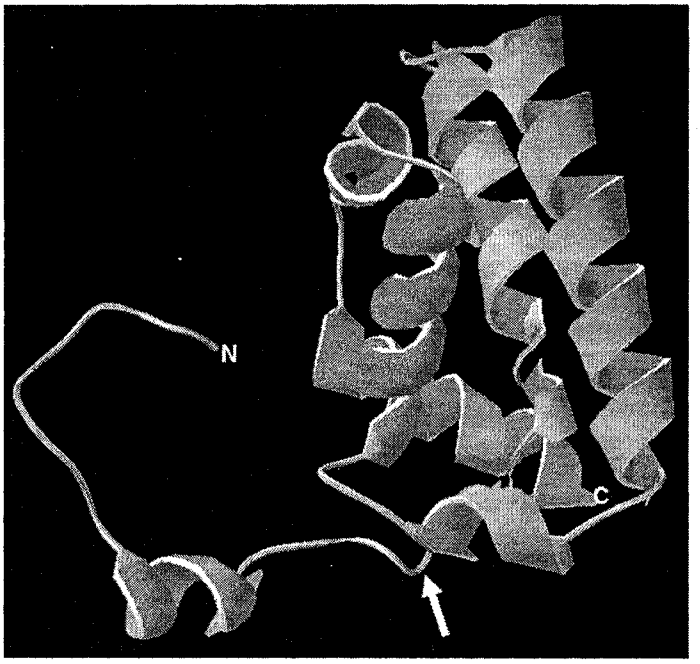


Figure 3. Predicted tertiary structure of the Msa protein showing the three transmembrane helices. Arrow indicates the predicted cleavage site for the putative signal peptide. N, N-terminus; C, C-terminus

We refined the predicted structure by fixing side chains, fixing problematic loops, removal of amino acid clashes (bumps) and energy minimization. The refinements did not yield any drastic change in the initial predicted structure. This was confirmed by visually inspecting the structure and verifying the backbone structure using Ramachandran plot (Figure 4) and computing the total energy difference between the initial model and the refined model.

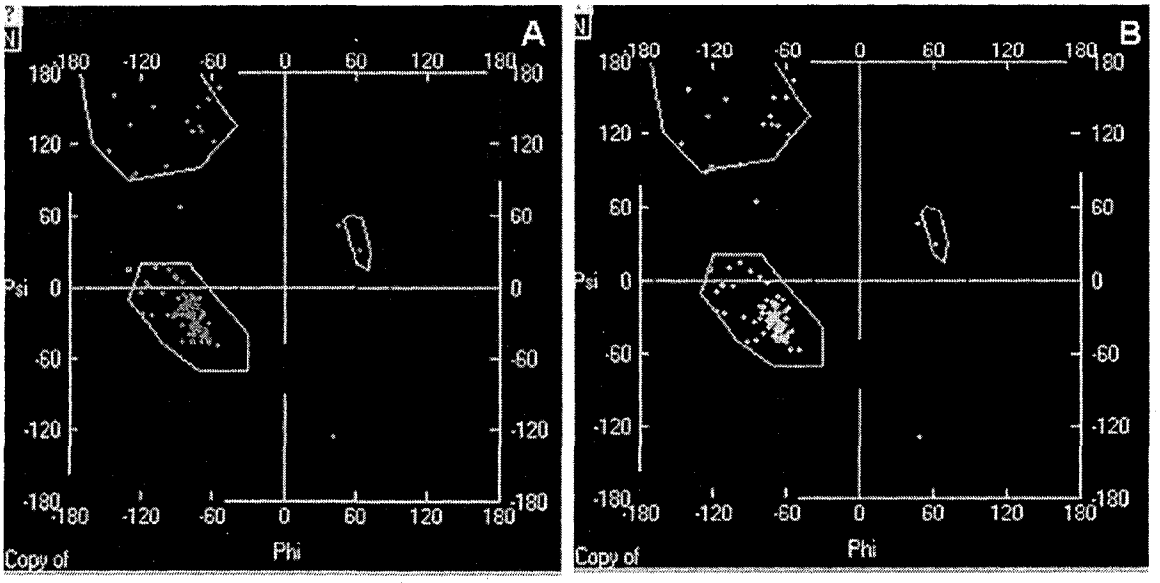


Figure 4. Ramachandran plot for the predicted tertiary structure of the Msa protein pre (A) and post (B) refinement

We analysed the predicted tertiary structure for clefts and binding sites using ProFunc server and found putative binding sites in the cytoplasmic region between the second and the third transmembrane helices (Figure 5A). We also used PINUP to predict putative interface residues in the similar region (Figure 5B). Another binding site prediction server Q-SiteFinder also predicted similar binding site and binding site residues (Figure 5C).

ProFunc also predicted a "nest" near the putative phosphorylation site (residues 47-50) which was predicted outside the membrane (108). The Msa has all the conserved residues that make up the predicted "nest". The predicted "nest" in Msa shows features of an anion-binding site. Such "nests" are characteristic functional motifs, which are found in ATP- or GTP binding proteins.

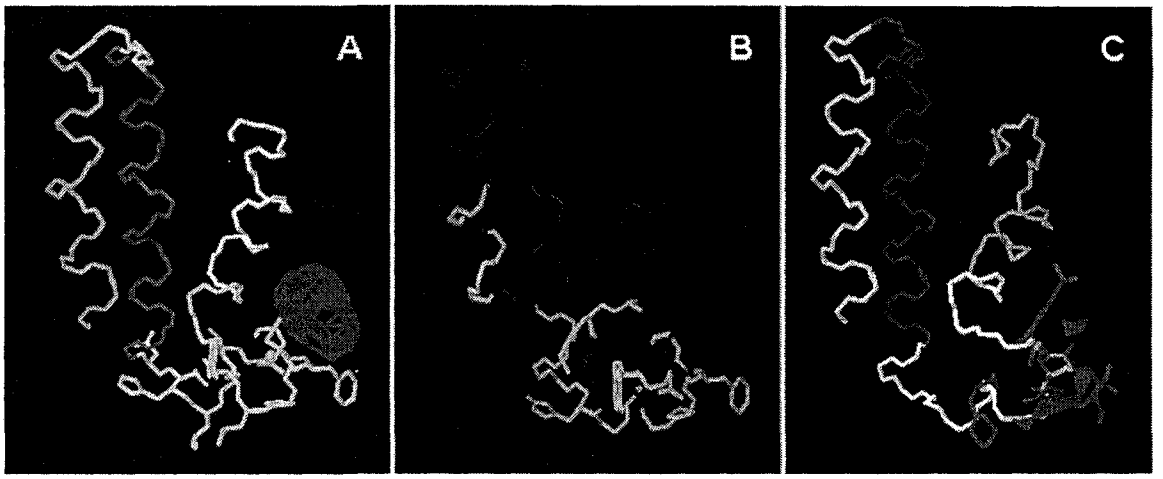


Figure 5. Binding site predictions for the Msa protein. (A) ProFunc predicted binding site (red); (B) PINUP predicted binding site (interface in green); (C) Q-SiteFinder predicted binding site and binding residues (pink)

Multiple sequence alignment of the Msa protein sequence from 11 different strains of *S. aureus* revealed 12 mutations in strain RF122 and seven mutations in strain MRSA252 relative to consensus. Mutations at amino acid positions 111, 131 and 133 were found in both MRSA252 and RF122 strains. None of these mutations were found in the predicted phosphorylation sites, predicted signal peptide sites or in the predicted anion-binding "nest". But many of the mutations were found both in the integral membrane segments as well as in the other parts of the loop regions. Only one out of the 12 mutations had the replacement (functionally different amino acid), while others were substitutions (functionally similar amino acids), in the strain RF122. In the strain MRSA152, two out of seven mutations were replacements, while others were substitutions. MRSA strain had three mutations in the predicted pre-ATP grasp domain, out of which one had an amino acid replacement. RF122 strain had only one amino acid substitution in the pre-ATP grasp domain. This indicates that the predicted functional sites are constrained from mutation.

We also used meta servers like SCRATCH (13) , ProSAL (1) and MetaPP (18) for predicting structure and functional properties of the Msa protein.

Conclusions

We predict that Msa is a membrane protein with a cleavable N-terminal signal peptide sequence, followed by three integral transmembrane regions. The Msa is also predicted to have an IN-OUT topology with at least two putative phosphorylation sites, one outside the membrane and one in the cytoplasmic region. A putative binding site is also predicted in the cytoplasmic region of Msa. Based on these predictions we put forward a model for the Msa protein (Figure 6).

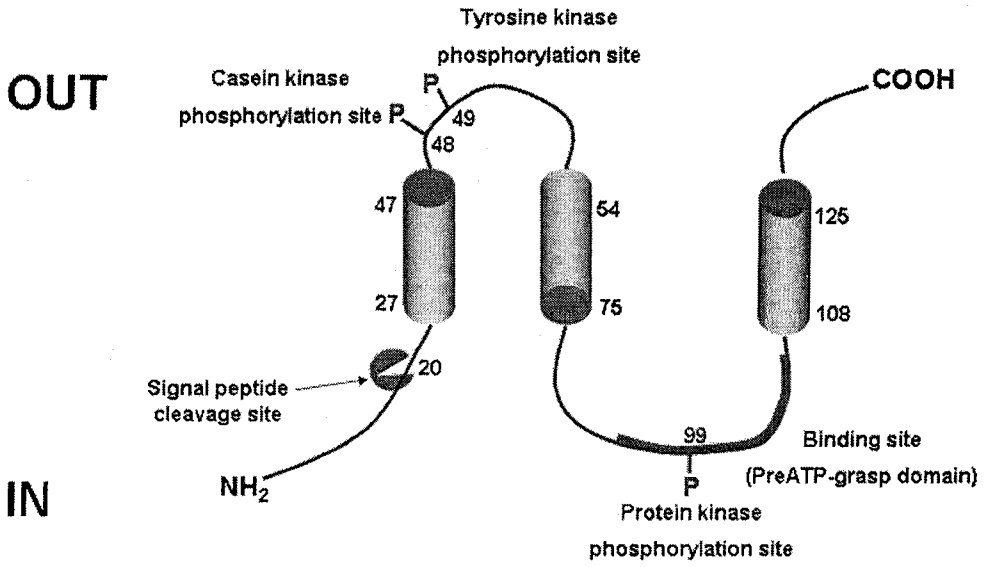


Figure 6. Predicted model for the Msa protein showing structural and functional features

This model prompted us to hypothesize that Msa might function as a novel signal transducer between the environment and the cytoplasm. If this hypothesis holds well then Msa might also play an important role in the global regulation of virulence, as it could mediate the environmental response of the regulators like *sar*A. This model will be used to design and execute experiments to confirm the functions and topology of Msa and further our understanding of its role in the pathogenesis of *S. aureus*.

CHAPTER III

DESIGN AND DEVELOPMENT OF SAMMD

Introduction

Background

Microarray studies enable the analysis of the pathogens response at a global level. There have been several studies carried out on the global expression profiles of *S. aureus* in response to different effectors like vancomycin (72), mild acid (109), stress (4) etc. There are also several transcriptional profiles of regulatory genes like *sig*B (8), *sarA* (17), *mgr*A (77) etc. To date there are about 41 published journal articles that contain about 93 microarray experiments in *S. aureus*. The use of this large amount of expression data is limited by the fact that it is not located in a centralized source. In addition, those data that have been deposited in the public databases are difficult to use for direct comparisons to data generated by researchers. We have addressed this issue by building a *Staphylococcus aureus* microarray meta-database (SAMMD) which contains all the published microarray data generated for *S. aureus*. SAMMD is a web accessible database that allows users to mine for information about a single or several genes. SAMMD can also be used to compare a whole transcriptome to published data.

Need for the Database

Databases are increasingly useful in biology as huge amount of data is generated by high throughput techniques such as Microarray technology. Computational tools are essential to analyse the vast mines of archived data and generate biological information. Scientists are encouraged to deposit published data in public databases such as the National Centre for Biotechnology Information - Gene Expression Omnibus (NCBI-GEO) (17) or European Bioinformatics Institute (EBI) Array Express (20), or Stanford Microarray Database (SMD) (26). To date, most groups however, have not complied resulting in a large amount of published DNA microarray data that is

inaccessible for further analysis by other scientists. This makes it difficult to manipulate data or make comparisons with other experiments.

Even when raw data is available online, the lack of computational tools and expertise as well as the difference in platforms used to generate the data makes it difficult to take full advantage of these resources. SAMMD addresses these issues by providing a central location for *S. aureus* microarray data. SAMMD was designed to allow users to quickly and easily mine the vast and growing collection of *S. aureus* transcriptomic data across different platforms. The search functions in SAMMD allow in depth analysis for the expression of one or collections of genes. SAMMD is a valuable tool for understanding the molecular mechanisms of pathogenesis in *S. aureus.*

Forerunners and Competitors

SAMMD is the first database that contains all the transcriptomic data for *S. aureus.* Databases devoted to other organisms have been developed. For instance, The *Saccharomyces* Genome Database (SGD) (28) is devoted to the yeast, *Saccharomyces cerevisiae.* SGD is a comprehensive genome database, which also contains information about several yeast Microarray experiments. Two other similar databases are devoted to E. coli gene expression (30), and to the human microarray data (LOLA) (37).

Potential Value

SAMMD is a highly valuable tool for staphylococcal research. SAMMD is useful to study a single gene, several genes, or a genome-wide transcriptome. SAMMD can also be used to gain insights about mutational status using transcriptomic data. Since SAMMD will be updated constantly as new transcriptomes are published, its utility and value will continue to grow.

Design and Development

Construction and Content

Database schema. SAMMD is a relational database consisting of five tables (Figure 7). The "annotation" table includes information about ORF IDs from six different strains of *S. aureus* (N315, COL, Mu50, MW2, MRSA252 and MSSA476), obtained from the primary source (National Centre for Biotechnology Information).

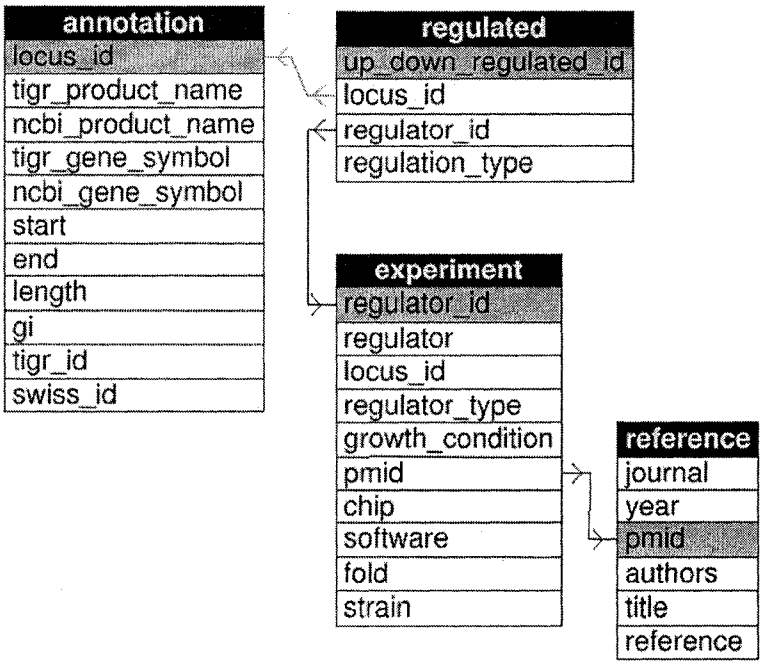


Figure 7. SAMMD relational database schema

The "experiment" table includes information about the Microarray experiments including the regulator or growth condition, number of replicates, strains, array platform, data analysis software, fold change cut-off value, and Pubmed ID (PMID). In addition, the experiment table will indicate whether RNA stabilizing agents were used and if the raw datasets are available. The "reference" table contains references to published journal articles from which the data was extracted. The "regulated" table contains the list

of differentially expressed genes (represented from the all the above mentioned strains of *S. aureus*), the effect of the mutation or growth condition on each gene (up or down). A table labeled "others" consists of information about non-transcriptomic DNA Microarray experiments such as genome comparisons. Corresponding primary and foreign keys are used to link the tables used in SAMMD.

Implementation. SAMMD was implemented with MySQL 4.1.10 as the back end. PHP 4.3.9, JavaScript and DHTML were used to develop the front end user interface (Figure 8). phpMyAdmin 2.9.2 was used as the database administration tool as well as for data entry. The entire project is hosted as a bioinformatics.org project. The user entered query is sent to a PHP script through an interactive form.

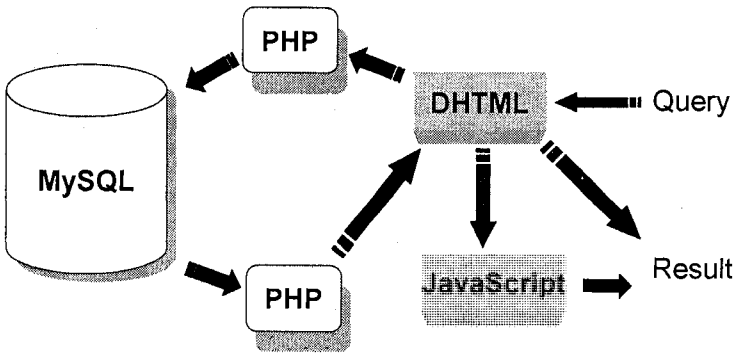


Figure 8. Implementation scheme of SAMMD

The PHP script sends the queries to MySQL database. The PHP script then retrieves and displays results, with hyperlinks to additional information. For every hit that is returned, additional information is presented through a JavaScript pop up window that gets the data from another appropriate PHP script.

Data sources and quality control. Data for SAMMD is obtained by manually curating the published journal articles related to DNA Microarray studies in *S. aureus* (Figure 9). The following keywords were used to find appropriate articles in the PubMed database: "stimulon", "transcriptome", "transcriptomics", "transcription profile", "transcription profiling", "microarray". Each keyword was paired with the term "*Staphylococcus aureus*".

We are using the same set of keywords to receive automatic email alerts from PubMed about the new *S. aureus* Microarray research articles, for the purpose of updating our database.

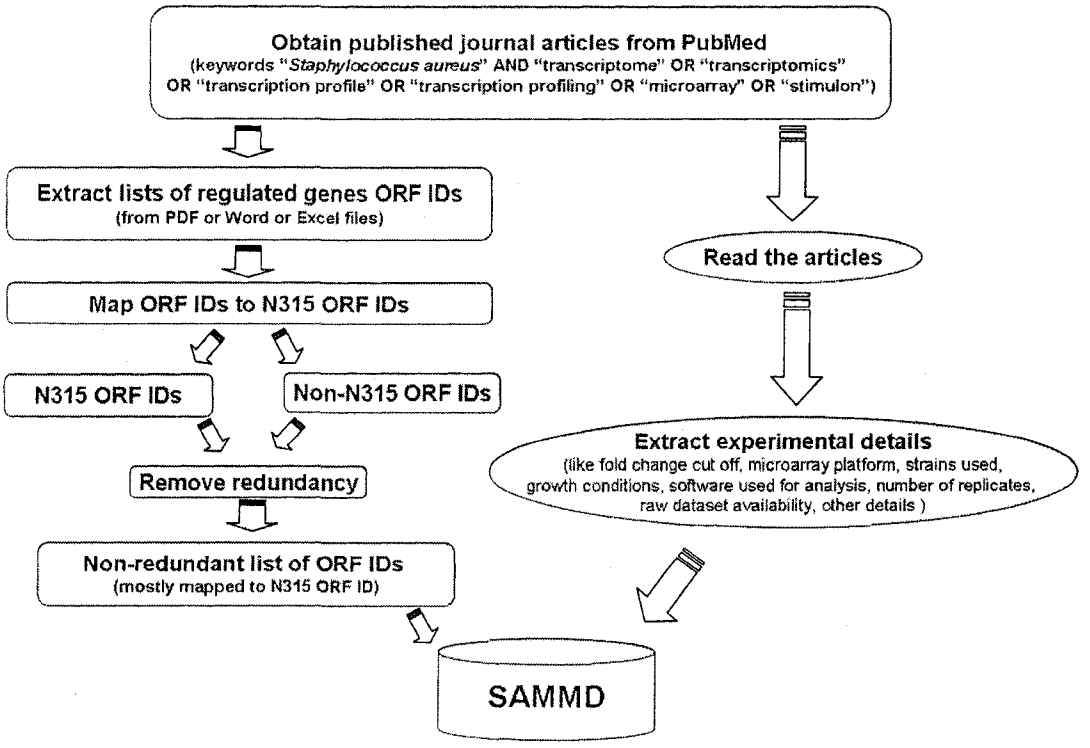


Figure 9. Data extraction and quality control

Data extraction. Details about the experiments such as fold change cut off, Microarray platform, strains, etc. were obtained by carefully studying the journal articles.

The related lists of regulated genes were extracted from either the journal article (from PDF files) or their respective supplemental files (Word or Excel files). The extracted lists of ORF IDs were mapped to N315 ORD IDs, using perl scripts, to enable comparison studies. The mapping scripts were written based on TIGR annotation files (*S. aureus* Version 6). Currently the mapping to N315 ORF ID is done for ORF IDs from 5 strains (MW2, COL, Mu50, MRSA252, and MSSA476). Genes that are not found in strain N315 are included in the database using their original name. The extracted data was entered in to the database using phpMyAdmin interface.

Quality of the data was checked at various points either by perl scripts or manual inspections. For instance, most of the *S. aureus* transcriptional experiments have few redundant genes from different strains represented in the same slide. These redundant genes are also reported as different entries in the list of regulated genes. In SAMMD, such duplicate entries were removed from the extracted and mapped gene lists. SAMMD has a feed back form, which the users could use to email the authors about any errors or other problems that they encounter.

Utility and Discussion

User interface. The user interface was developed using PHP, JavaScript and DHTML. Users can browse through the transcriptomic content using the following categories: "transcriptomes of regulatory genes", "transcriptomes under different growth conditions", "transcriptomes of different strains", "transcriptomes published during a specific year", "transcriptomes published in a specific journal" and "transcriptomes in a specific reference".

A powerful search tool provided in the home page of the database provides users the ability to search for the transcriptomic status of a particular gene (Figure 10). Users can search using an ORF ID (e.g. SA1233) or a Gene name (e.g. *sarA*) or a gene product name (e.g. urease). The search by ORF ID option accepts ORF IDs from any of

the following six strains of *S. aureus* (N315, MW2, Mu50, COL, MRSA252, MSSA476) as input.

After execution of the search, the user is provided with an output that lists all of the experiments or transcriptomes where the query gene is found. The output also shows how the query gene was affected by the regulatory gene or growth conditions i.e. up regulated, down regulated or not affected.

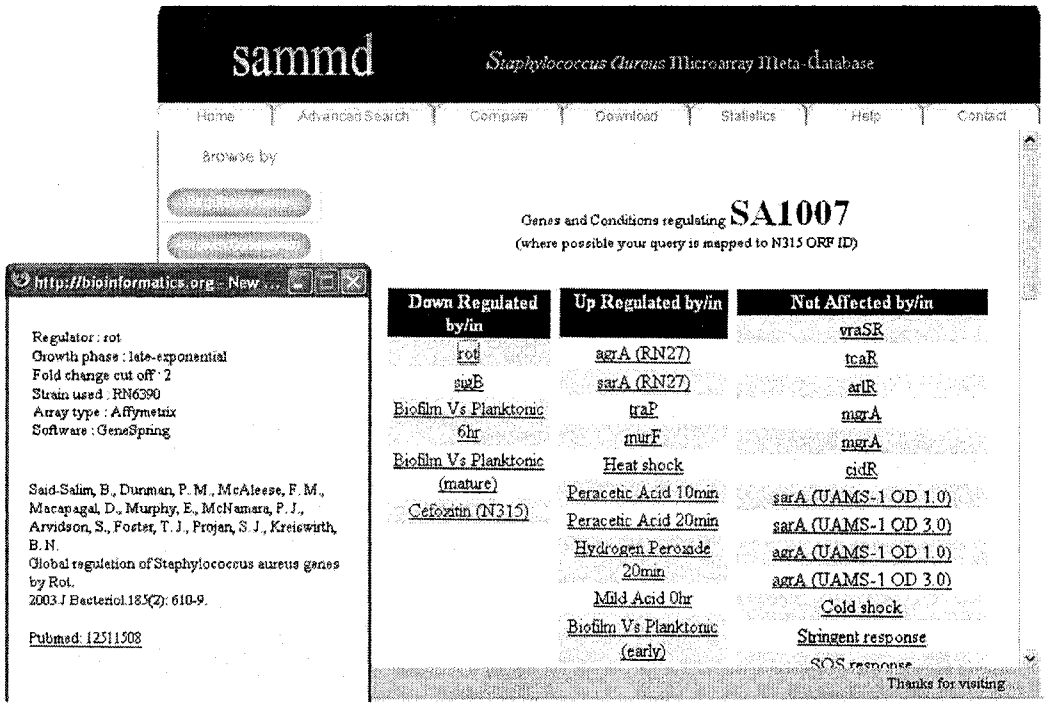


Figure 10. Output of a simple search in SAMMD using an ORF ID. ORF ID "SA1007" has been used as a query term in this search. All terms in the output are hyperlinked to further information.

For a search using a Gene name, SAMMD accepts any standard *S. aureus* gene name and first returns a list of ORF IDs that have similar gene name (Figure 11) from the above mentioned strains of *S. aureus*. Clicking on a particular ORF ID shows the transcriptomic status of that particular ORF ID (usually mapped to N315 ID) in all different transcriptomes available in SAMMD.

We have also implemented a full text search against the "NCBI Product Name" column of the "annotation" table, under the search by "ORF Function" option. This lets users execute searches in SAMMD using key words. This full text search is Boolean operator enabled allowing users to add operators such as "AND", "OR", "NOT", "" and "*" to limit their search. The SAMMD help page contains more detailed help with examples about the usage of these Boolean operators.

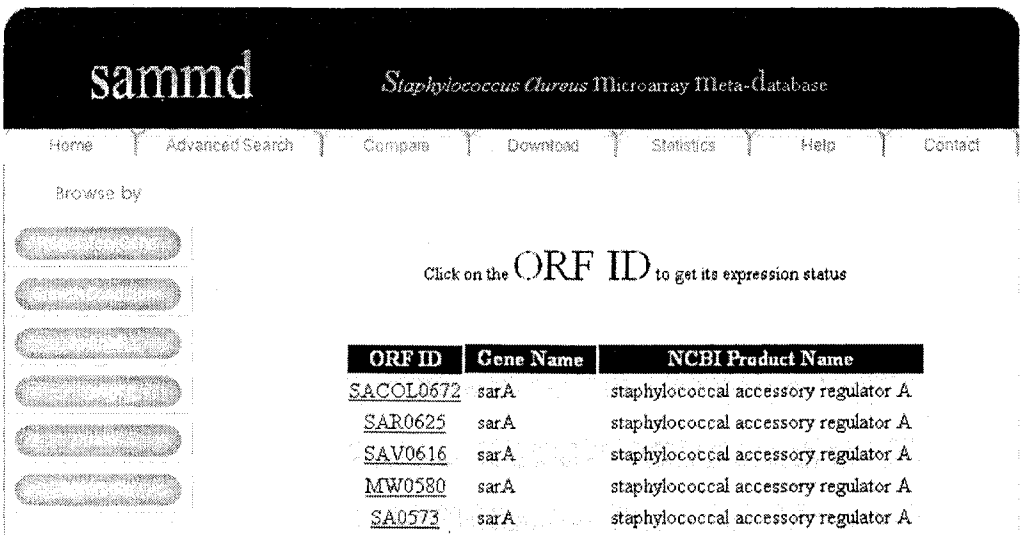


Figure 11. Output of a simple search in SAMMD using a Gene name. "sarA" has been used as a query term in this search. Clicking on the ORF ID would pull out its transcriptomic status in SAMMD, after mapping it to N315 ID.

The advanced search lets the users to search SAMMD using a list of genes. Users can input a list of genes (any number of genes), as corresponding N315 ORF IDs. The output is the list of transcriptomes that has overlapping genes with that of the user entered gene list (Figure 12).

Comparing datasets. SAMMD users can compare two or three datasets using the "Compare" menu option (Figure 13). Comparison between datasets is performed by set operations using PHP. The results are displayed in a Venn diagram, with the numbers in

the diagram linked to the list of overlapping genes (Figure 14). The SAMMD help page shows an illustration of the dataset comparison option.

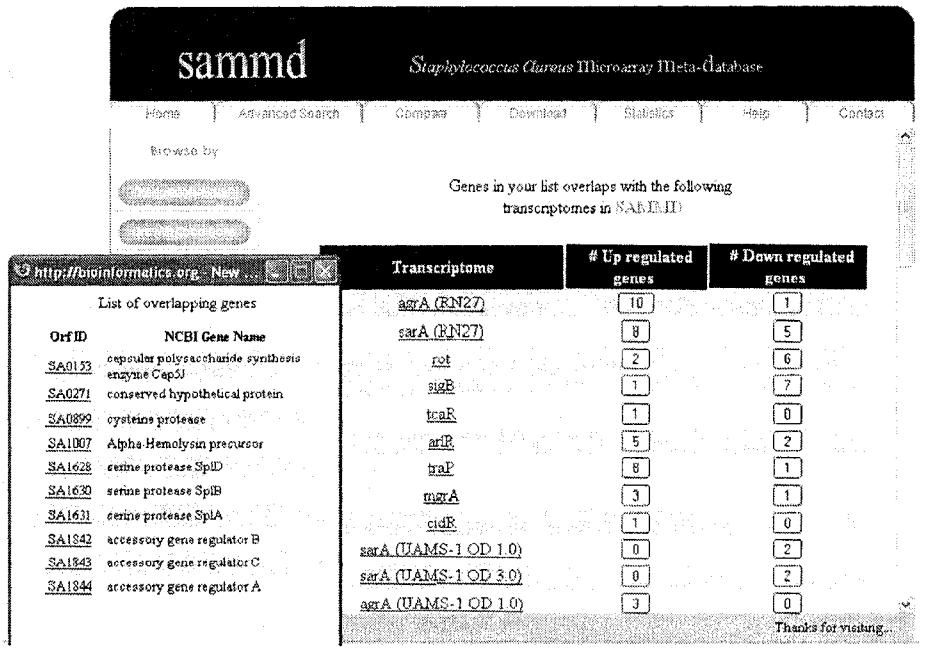


Figure 12. Output of an advanced search in SAMMD using a list of ORF IDs. Output shows the overlapping transcriptomes, with the corresponding list of genes. All terms in the output are hyperlinked to further information.

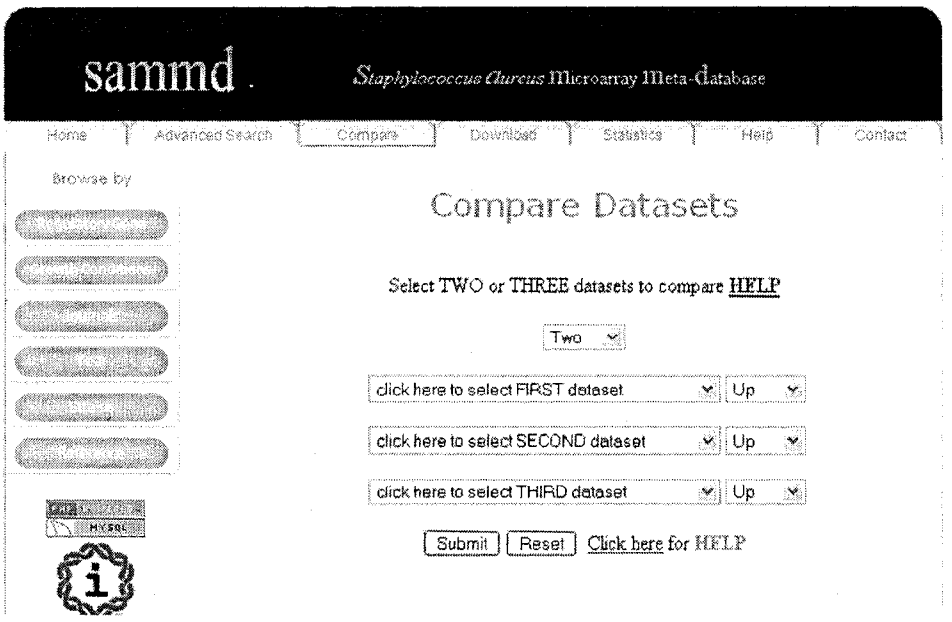


Figure 13. Compare Datasets option. Users can compare among different datasets using this option.

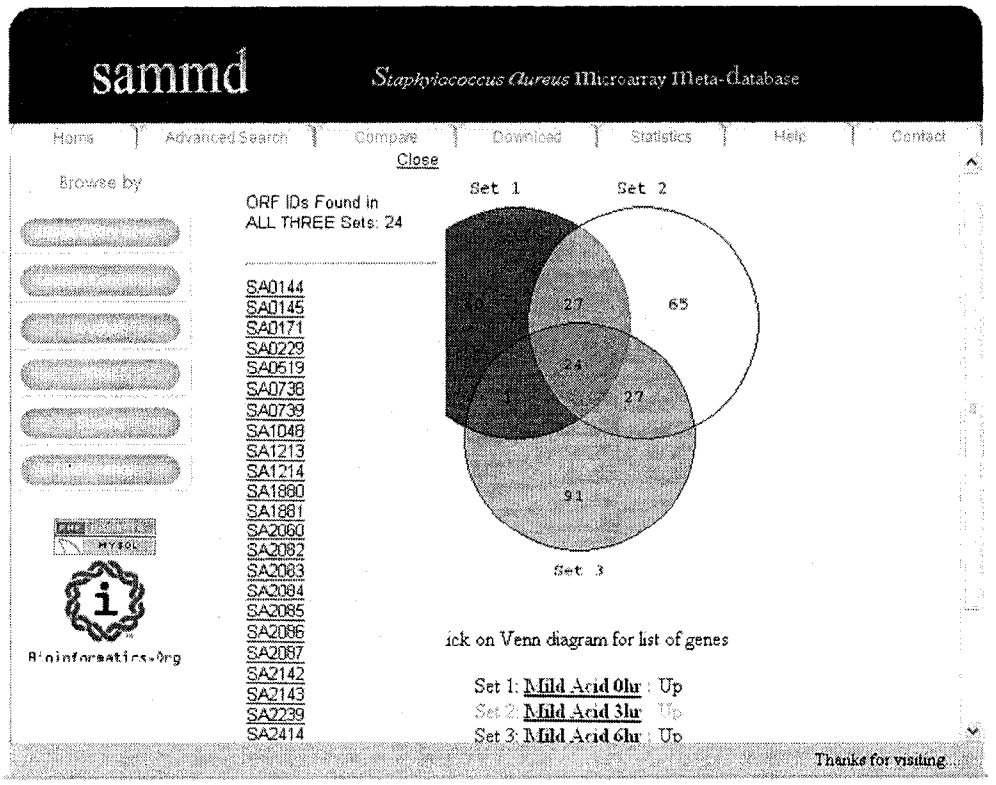


Figure 14. Venn diagram showing the results of comparing three datasets. Clicking on the numbers in the Venn diagram shows the corresponding list of genes in a pop up window.

Options are also available for users to download relevant data from the database. Current statistics about the number of records in the database and a detailed help page with example gene lists and usage illustrations are also available. Contact information is also included for additional help with SAMMD and to receive comments and suggestions from users in order to improve the database.

Intended use and benefits. SAMMD would be of immense use to scientists who are working on *S. aureus*. Knowing the transcriptional status of a particular gene from the literature might be a cumbersome task, because of the different semantics that are used to denote the genes from different strains. Manual searching for the transcriptional status of a particular gene becomes a laborious task, given the number of experiments

and the huge number of regulated gene lists. SAMMD helps molecular microbiologists to over come these problems.

Comparative transcriptome analysis becomes conveniently possible with the help of SAMMD. Scientists who perform Microarray based experiments, could now easily compare their list of regulated genes (transcriptome) to that of the other transcriptomes in the database. They could also compare among the datasets that are already in SAMMD. This could help them to find out regulatory patterns and connections between their transcriptome and other transcriptomes, using the list of overlapping genes.

Future developments. We are planning to incorporate a graphic module to represent the list of overlapping genes that is generated as a result of advanced search using a user entered list of genes.

Conclusion

Importance and Relevance of the Database

Microarray gene expression databases like NCBI-GEO and EBI-ArrayExpress hold the raw data only for a very few of the 41 published Microarray papers that are listed in SAMMD. The raw data that are published in these databases are not easily accessible for the use by biologists.

By developing SAMMD, we have addressed these issues. SAMMD is a searchable database of Microarray gene expression data of *S. aureus*. Such a database is valuable in staphylococcal research in light of increasing multiple-antibiotic resistance in *S. aureus*. SAMMD will allow scientists to study the role of individual genes in the context of global transcriptomes as well as enable comparison of new transcriptomes to published ones. SAMMD will facilitate understanding of the complex regulatory networks of *S. aureus*.

Availability and Requirements

The database is entirely based on open source concept and hence its usage is licensed under GNU General Public License (GPL). The database is available at the URL: http://bioinformatics.org/sammd/

SAMMD works best using the open source browsers like Firefox. JavaScript based pop-up windows are used to display some of the data in SAMMD, so the users should have JavaScript enabled their browsers for the full functionality of the database. Much of the data in the database could be directly downloaded from the database website. Any other file used in the development of the database would be provided up on request.

I will use SAMMD to perform the comparative transcriptomic analysis on *msa* transcriptome. This would help us to put the *msa* regulatory network in the context of other global regulators.

CHAPTER IV

msa AND *sarA* REGULON

Introduction

Staphylococcus aureus is an important pathogen that causes several diseases in humans ranging from superficial skin infections to severe disorders like osteomyelitis and endocarditis. *S. aureus* is able to cause such diverse diseases, by virtue of its host of virulence factors and complex regulatory networks. The virulence factors generally grouped as surface proteins and secreted proteins are controlled by a variety of regulators, of which *sarA* and *agr* play the central role (27). The *agr* system is a quorum sensing mechanism which positively controls several exoproteins and toxins while negatively controlling surface proteins. *sarA* positively regulates *agr*, thereby producing both *agr*-dependent and *agr*-independent effect on several virulence factors. *sarA* activity is also modulated by several environmental factors.

Earlier we showed that *msa* modulates the activity of *sarA* (96). We also predicted that *msa* is a putative membrane protein, which could mediate the transduction of environmental signals to the regulatory elements like *sarA* (80). In this paper, we explore the role of *msa* in the global regulation of virulence in *S. aureus*.

Materials and Methods

Bacterial Strains and Growth Conditions

The wild type COL and its isogenic mutant *msa*- (96), were grown at 37^0C with constant aeration (200 rpm at a medium flask volume ratio of 0.5), in TSB (Trptic Soy Broth) or on TSA (Tryptic Soy Agar) supplemented with antibiotics when appropriate. *Escherichia coli* was also grown in TSB or TSA. Antibiotics erythromycin 10µg/ml, tetracycline 5µg/ml, kanamycin 50µg/ml and ampicillin 100µg/ml were used in the study.

RNA Isolation

The wild type *S. aureus* COL strain and its isogenic *msa* mutant were grown in the absence of selective antibiotics, in a 500 ml conical flask containing 200 ml of TSB. Two ml of the culture was collected at OD 0.3, 1.5 and 4.0 (560nm), and mixed with 4ml of Bacterial RNA Protect solution (Qiagen, Valencia, CA). The cells were harvested by centrifuging the mixture for 10 min at 7300 rpm in a table top centrifuge. The pellets were then resuspended in 1 ml of Trizol (Invitrogen, Grand Island, NY). Resuspended cells were broken using the FastPrep sytem (Qbiogene, Irvine, CA) at a speed of 6.0 for 40 seconds and RNA from the lysate was extracted as per the manufacturer's instructions. RNeasy mini kit (Qiagen) was used to purify the extracted RNA. An on-column DNase (Qiagen) treatment was done. The quality and the concentration of the RNA were then analyzed using the RNA 6000 Nano LabChip system (Agilent Technologies, Santa Clara, CA).

RNA Labeling and Hybridization

RNA labeling and hybridization were done as per TIGR protocol (Standard operating procedure no. M0007: http://pfgrc.tigr.org/protocols/protocols.shtml). Briefly, random hexamers (Invitrogen) as primers were annealed to 2.5 µg of the total RNA (70°C for 10 min, followed by snap-freezing in ice for 1 min), along with RNaseOUT recombinant ribonuclease inhibitor (Invitrogen). The primers were then extended using SuperScript II reverse transcriptase (Invitrogen) with 0.1 M DTT and 12.5 M of aa-dNTP/aa-dUTP mix (Ambion, Austin, TX) at 42°C overnight. Residual RNA was removed by alkaline treatment with 1 M NaOH followed by neutralization with 1 M Tris (pH 7.4). The resulting cDNA was purified using Qiagen MinElute PCR purification kit (Qiagen) and labeled with Cy3 and Cy5 mono-Reactive dyes (Amersham Biosciences, Pittsburg, PA) as per manufacturer's instruction.

Labeled cDNA was purified using MinElute PCR purification kit and hybridized with *S. aureus* COL genome microarrays version 6.0, provided by Pathogen Functional Genomics Resource Center (PFGRC), as per the TIGR protocol (Standard operating procedure no. M0008: http://pfgrc.tigr.org/protocols/protocols.shtml). Briefly, the slides were treated with a preheated pre-hybridization buffer (5 X SSC [Ambion], 0.1 % SDS and 1 % BSA [Sigma, St. Louis, MO]) at 42°C for 1 hr. Slides were then washed with distilled water followed by washing with isopropanol and dried by centrifuging. The dried Cy3 and Cy5 labeled cDNA was resuspended using hybridization buffer (50 % formamide, 5 X SSC, 0.1 % SDS and 300 µg sheared salmon sperm DNA [Ambion]) and denatured at 95°C. The sample was then applied to the pre-hybridized slides under the LifterSlip (Fisher, Pittsburg, PA). The slides were incubated at 42°C overnight. The hybridized slides were serially washed with low, medium and high stringency wash buffers, dried by centrifuging and scanned using Axon 4200A scanner (Molecular Devices, Sunnyvale, CA). The microarray scans were saved as TIFF image files.

Microarray Data Analysis

GenePix Pro 6 (Molecular Devices) was used to extract the spot intensity values and generate GPR files from the microarray image files. Further data analysis was done using the Acuity 4.0 software (Molecular Devices). We did four hybridizations for this experiment, including a biological replicate and a dye-swap experiment for each replicate to account for dye-bias. Spots flagged as empty or bad were excluded and the raw data from each slide was normalized using LOWESS method, with background correction. The data from the replicates were combined (using the median value) and a one sample t-test was performed. The volcano plot was used with a fold change cut-off of >= 3 and a p-value of <0.05, to filter the genes that were differentially expressed.

The ORF IDs in the list of differentially expressed genes were then mapped to *S. aureus* N315 IDs (using TIGR-PFGRC annotation file:

http://pfgrc.tigr.org/annotationPages.shtml). The transcriptome of *msa* (late-exponential phase) was then compared with the other transcriptomes of *S. aureus* using SAMMD.

Quantitative Real-Time PCR Analysis

qRT-PCRs were done following the method described by Sambanthamoorthy et al., 2006 (96), for *cap*A gene, to confirm the microarray results. Briefly, *cap*A gene was cloned in to plasmid pCR2.1 TOPO to determine plasmid efficiency. The primers with good efficiency were used (LP 5'- CAGCAGTTAAAGTCGCACCA-3', RP 5'-TAAGCGCGACAACTAATCCT-3') to perform qRT-PCRs. The RNA from the mutant and the wild type was reverse transcribed using iScript cDNA synthesis kit (Bio-Rad) and the qRT-PCR reactions were done using IQ SYBR Green Supermix (Bio-Rad) as per the manufacturer's quidelines, using an iCycler (Bio-Rad). Analysis of expression of each gene was done based on two independent experiments.

Western Blot Analysis

Post-exponential phase culture supernatant was collected for wild type, *msa* and *sarA* mutants. The culture supernatants were concentrated using Amicon Ultra 3000 daltons centrifugal filters (Millipore). Protein concentration was measured using Bradford assay. Proteins (2µg) for Western blot analysis were electrophoresed under denaturing conditions on precast gels (Bio-Rad). Polyclonal anti-SEB from sheep (Abcam Inc, Cambridge, MA) was used as per the manufacturer's instruction. Blotting and analysis was done as previously described by Blevins et al. (10).

Hemolytic Activity

Tryptic soy agar supplemented with 5% v/v defibrinated blood was used to test the activity of different hemolysins (*hla, hlb, hld, hlg*). Defibrinated or heparinized blood from sheep, rabbit, horse (Hemostat, Dixon, CA) and human (Interstate Blood Bank, Inc. Memphis, TN) were used. Blood agar plates streaked with wild type, *msa* and *sarA* mutants were incubated for 24 hours at 37^0C and then refrigerated overnight. The

hemolytic activity before and after refrigeration were recorded. The interpretation of the hemolytic reactions obtained on blood-agar plates were made as previously described by Haque and Baldwin (51).

Lipolytic Assay

The lipolytic activity was qualitatively assayed using Tributyrin agar plates. Tributyrin agar plates were prepared using Tributyrin agar base (EMD, San Diego, CA) and 1% w/v Tributyrin (Sigma, St. Louis, MO) as per manufacturer's instruction. *msa* mutant, *sarA* mutant and the wild-type grown for 24 hrs in the tributyrin agar plates at 37^0C, were kept in the refrigerator overnight for clearance of the zone of lysis.

Results and Discussion

Transcriptional Profile of msa Mutant

Our goal in this study was to evaluate the global effect of *msa* on gene expression. We had three objectives. First, given the role of *msa* in modulating the expression of the global regulator *sarA* in *S. aureus*, we hypothesized that the *msa* mutant will have a global effect on gene expression. In order to study to global effect of *msa* in *S. aureus*, we examined the transcriptional profile of the *msa* mutant relative to wild type in strain COL. Second, we examined the transcriptional profile of the *msa* mutant in three growth phases: mid-exponential, post-exponential and late exponential to determine any growth phase dependence in *msa* regulation. Third, we compared the transcriptional profile of the *msa* mutant to all known transcriptomes using SAMMD analysis to define any overlaps. Finally, since we previously had evidence that some genes are regulated by *msa* in a *sarA* independent fashion (96), we compared the *msa* and *sarA* transcriptomes in order to survey the genes that are regulated in a *sarA*-dependent versus *sarA*-independent manner.

The transcriptome of the *msa* mutant showed that a total of 325 genes were differentially transcribed between the *msa* mutant and wild type in at least one of the

three growth phases tested. These data are based on at least threefold difference in the level of transcription and statistical significance ($P<0.05$) in two independent experiments. The transciptome analysis indicated that 287 genes were up-regulated and 38 genes were down-regulated by *msa.*

We performed a gene ontology analysis on the *msa* profile using the TIGR primary role category. We found that most of the up-regulated genes belong to five categories of genes with known functions: Transport and binding proteins; Cell envelope; Cellular Processes; regulatory functions; energy metabolism; and signal transduction (Figure 15).

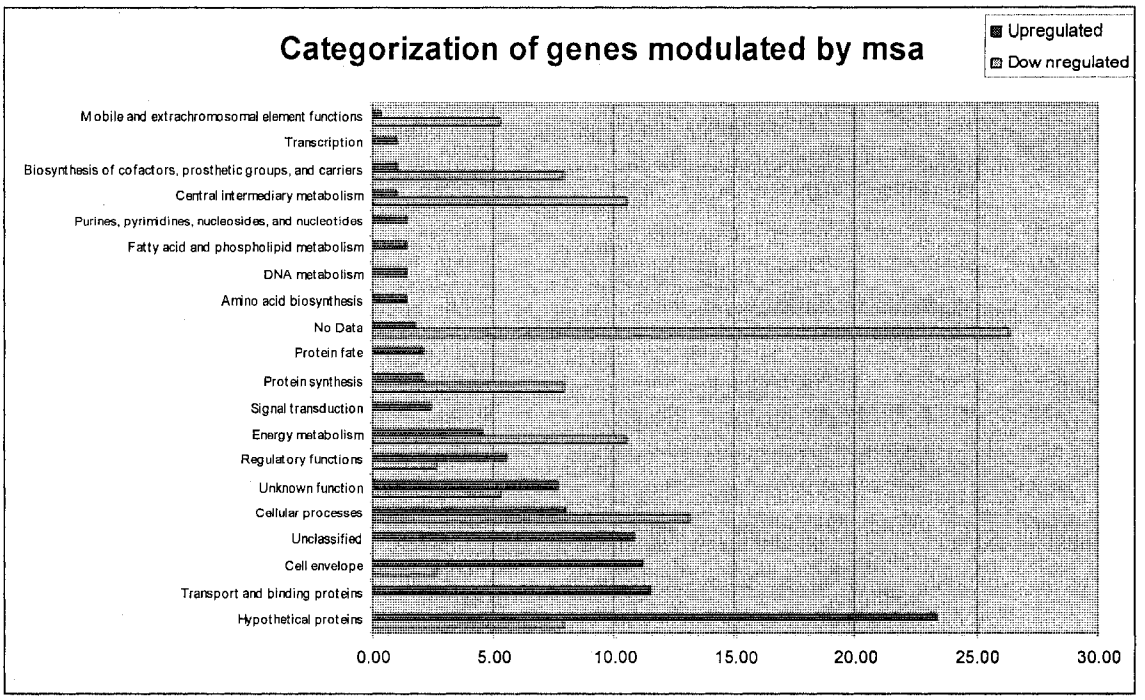


Figure 15. Categorization of genes modulated by *msa* in at least one growth phase

Most of the genes that are down-regulated by *msa* belong to five categories of known function: Cellular processes; Central intermediary metabolism; Energy metabolism; protein synthesis; and biosynthesis of cofactors, prosthetic groups, and carriers.

The gene ontology analysis shows that genes up-regulated by *msa* are involved in a wide variety of functions; mostly metabolism and transport, similar to *sarA*. It is important to note that the cellular processes category contains most of the virulence factors which suggests an important role of *msa* in virulence. It is also important to note that a significant number of genes that are differentially expressed in the *msa* mutant are hypothetical proteins or have no data associated with them. Therefore, we anticipate that the *msa* mutant profile will change as more gene functions are determined.

Growth Phase Dependency of msa Regulation

Analysis of the *msa* transcriptome at different growth phases showed that 77 genes are up-regulated by *msa* at the mid-exponential phase, while 244 genes are up-regulated at the late-exponential (1.5 OD) and 52 genes up-regulated at the post-exponential phase (4.0 OD). *msa* downregulates 17 genes at the mid-exponential phase, 1 gene at the late-exponential and 20 genes at the post-exponential phase. Several genes were differentially expressed in more than one growth phase. Indeed, comparison of the genes affected at different growth phases shows that only eight genes are up-regulated in all three growth phases (Figure 16A). These genes include (*agr*B, D, C2, *epi*E (epidermin immunity protein), SACOL1578 (FtsK/SpoIIIE family protein), SA2262 (AhpD family protein), SA2367 (hydrolase family protein) and SAS1940a (delta-hemolysin precursor)). There are 52 genes that are up-regulated only in the mid and late-exponential phase and 16 genes that are up-regulated in late and post-exponential phase. Just two genes are up-regulated only in mid and post-exponential phase.

The large number of differentially expressed genes during the late exponential growth phase suggests that *msa* plays a major regulatory role during the transition from mid-exponential to late exponential growth phase. Indeed, this is supported by our previous finding that *msa* transcript peaked at the late exponential growth phase relative to the mid and post exponential phases (96).

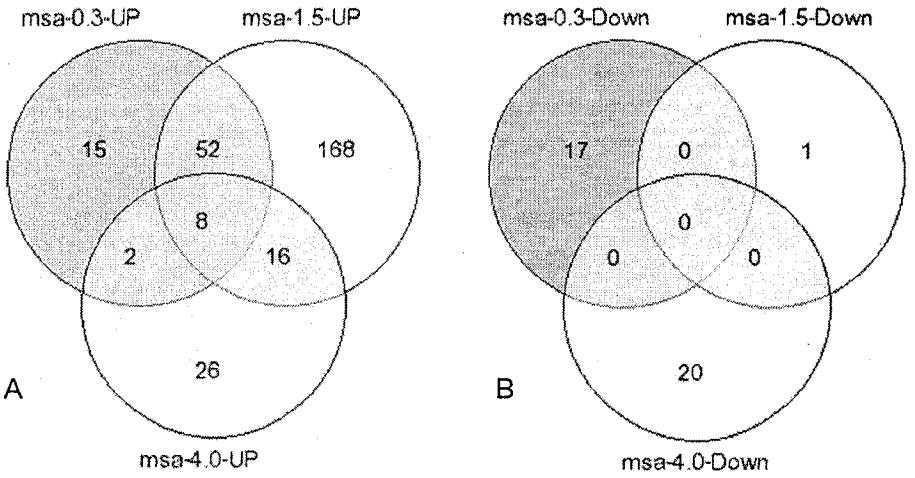
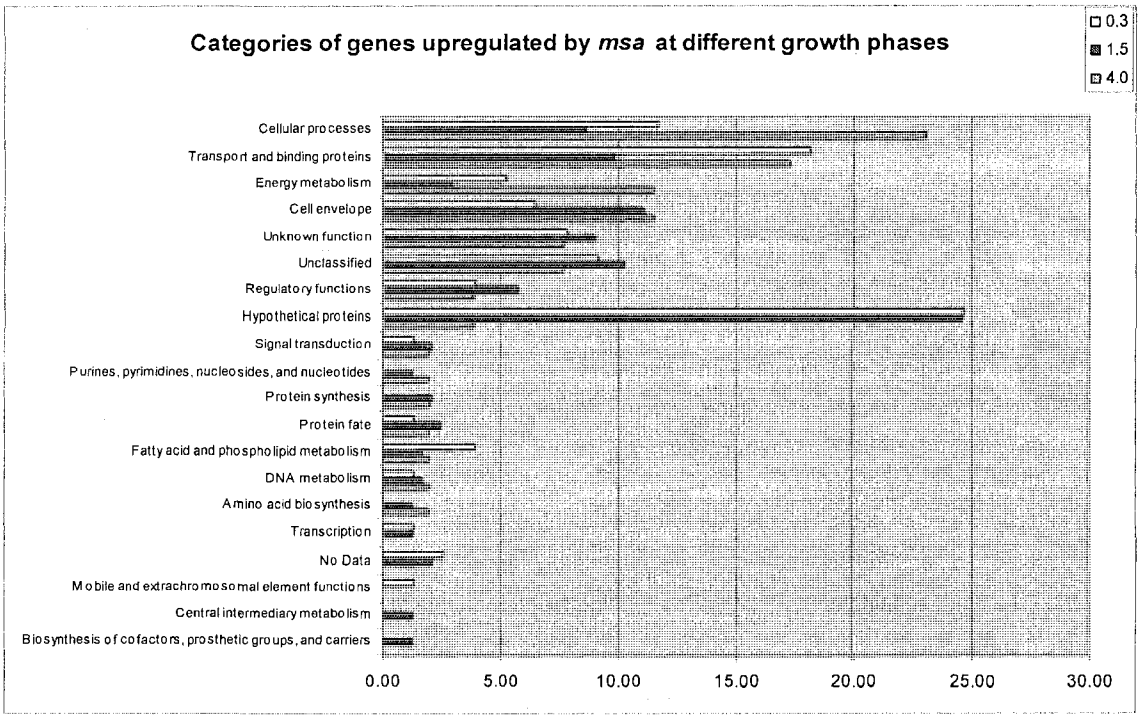


Figure 16. A) Comparison of genes up-regulated by *msa* at different growth phases. B) Comparison of genes down-regulated by *msa* at different growth phases.

Analysis of the role categories of genes regulated by *msa* at different growth phases revealed that genes that are up-regulated by *msa* at the mid-expo growth phase genes belong to Transport and binding proteins (Figure 17). At the post-exponential growth phase genes that are up-regulated are mostly related to pathogenesis (Cellular processes). At late and post-exponential growth phases genes up-regulated by *msa* mostly belong to Cell envelope related.

Genes involved in energy metabolism, central intermediary metabolism, protein synthesis and cofactors biosynthesis were down-regulated by *msa* at post-exponential phase. We also found that 168 genes are up-regulated only at late-exponential phase. The genes that are up-regulated only at late-expo phase had 21 genes that are related to cell envelope. These 21 genes were mostly related to biosynthesis and degradation of surface polysaccharides, several membrane proteins, cell wall anchored lipoproteins etc. Interestingly only one such membrane protein was up-regulated in mid-expo and post-expo phase.



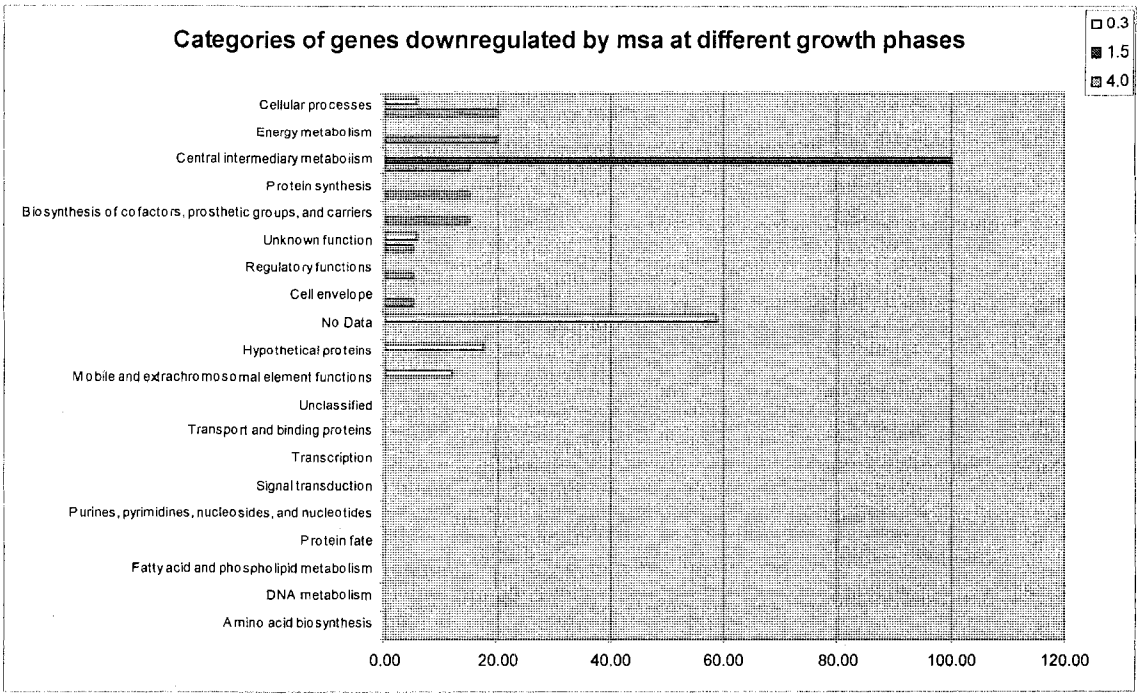


Figure 17. Categorization of genes modulated by *msa* at different growth phases

Our microarray data shows that the difference in mid and late-expo phase in *msa* is that, *msa* mutant failed to produce several cell envelope related products, which could explain the defective biofilm forming phenotype. There was no overlap between down-regulated genes expressed in any of the three time points. One striking feature observed was that, while 168 genes are up-regulated only at late-expo, only one was down-regulated at late-expo, but there were about equal number of genes up and down-regulated at only post-expo phase.

Transcriptional Profile of COL sarA Mutant

We earlier showed that *msa* modulates the activity of *sarA*. *msa* mutant shows reduced expression of *sarA* (96). We compared the *msa* transcriptome to other *sarA* transcriptome (18, 40) using SAMMD (79). Since, *msa* modulates *sarA* activity, we expected some kind of overlap between *msa* and *sarA* transcriptomes. But surprisingly there were only 10 genes of *msa* transcriptome that overlapped with RN27 *sarA* transcriptome (40) and 2 genes overlapping with UAMS-1 *sarA* transcriptome (17). We also know that COL is closely related to RN27 that with UAMS-1 (18). Comparison of RN27 and UAMS-1 *sarA* transcriptomes in SAMMD showed that there was only 5 genes that overlapped between the two transcriptomes. This emphasized the point of strain differences in *S. aureus.* So, we generated the *sarA* transcriptome in COL to compare with *msa* transcriptome.

Analysis of *sarA* transcriptome in COL showed that *sarA* upregulates 93 genes and downregulates 79 genes. Most of the genes that are up-regulated by *sarA* belong to Cell envelope and cellular processes category (Pathogenesis related). Most of the genes that are down-regulated by *sarA* belong to Transport and metabolism related.

Comparison of sarA and msa Transcriptomes at Post-Expo Phase

Comparison of *sarA* and *msa* transcriptomes at post-expo phase showed that about 50% of the genes up-regulated by *msa* are also up-regulated by *sarA* (Figure 18A). But, only few genes were down-regulated by both *sarA* and *msa* (Figure 18B).

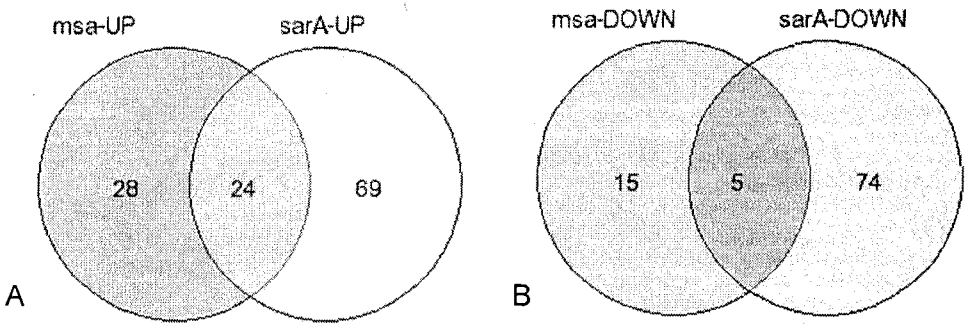


Figure 18. Comparison of genes regulated by *msa* and *sarA* in COL. A) Up-regulated. B) Down-regulated.

Most of the genes that are up-regulated both by *msa* and *sarA* belong to the cellular processes and cell envelope category (Figure 19) (that is the major category of genes up-regulated by *sarA* as well). This could mean that effect of *msa* on genes involved in cellular processes and cell envelope are mediated through *sarA*. Most of the genes that were up-regulated only in *msa,* belong to the category of transport and binding proteins (that is the major category of genes up-regulated by *msa* as well). This could mean that *msa* upregulates genes involved in transport independent of *sarA*. Interestingly most of the genes down-regulated by *sarA* belongs to the transport and binding category.

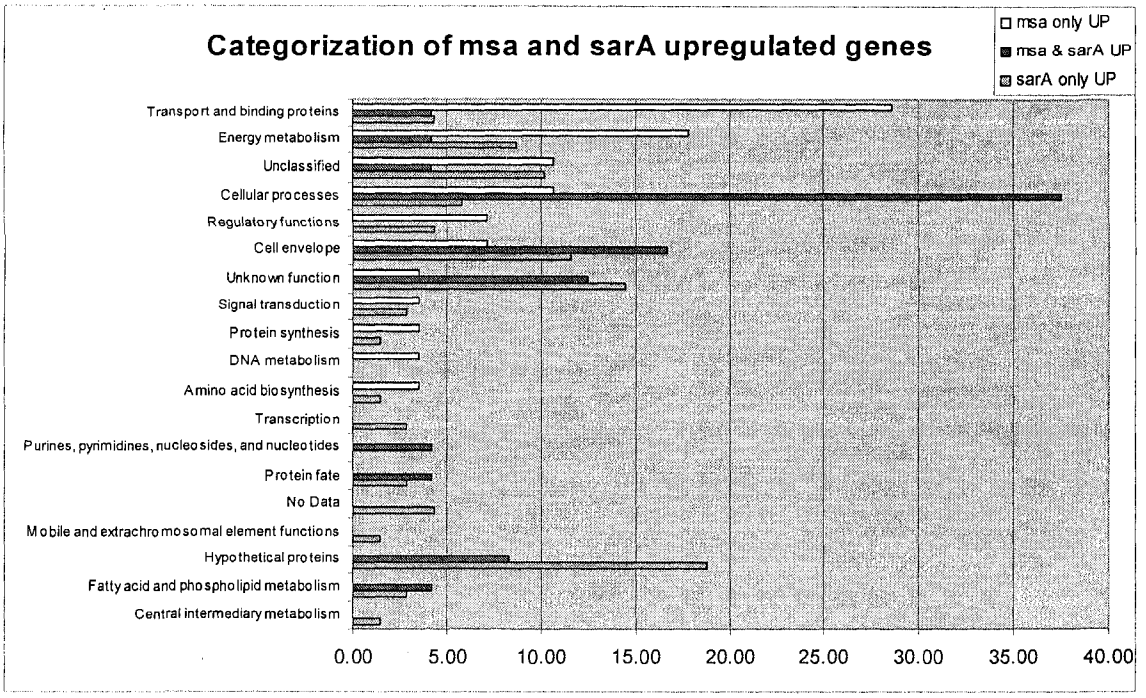


Figure 19. Categorization of *msa* and *sarA* up-regulated genes

Seven genes were oppositely regulated in *msa* and *sarA*. When we examined these genes we found that five of them belonged to the Arginine Deiminase operon (SA2424, SA2425, SA2426, SA2427, SA2428), 2 other genes were SA1269, SA1270. Regulation of genes involved in energy metabolism by *msa* and *sarA* looks like mutually exclusive. Even though several energy metabolism related genes are up-regulated by both *msa* and *sarA*, just one gene is found as up-regulated by both. A similar effect was observed in down-regulated genes – ie, even though energy metabolism related genes are the major category of genes that are down-regulated by *sarA* and *msa,* not even a single energy metabolism related gene is down-regulated by both *sarA* and *msa. msa* downregulates genes involved in cellular processes and energy metabolism. There was no single gene involved in transport that is down-regulated by *msa,* while *sarA* downregulates several of the genes involved in transport and binding.

Expression Patterns of Selected Genes

Influence of msa on known regulatory elements. msa affects a large number of regulators (20) belonging to regulator families *bgl*G, *rpi*R, GntR, LysR, MarR, LuxR, *mer*R, Crp/Fnr and Cro/CI. A few of these are also regulated by *sarA* (SA1956, SA2108, SA2424). Transcriptional regulators TcaR, IcaR and NirR are also differentially affected by *msa.*

Of these, NirR is the only regulator that is down-regulated by *msa* (post-expo significant, others non-significant, but differentially expressed), all the other regulators are up-regulated by *msa.*

All the regulators that are up-regulated at mid-expo (SA1961, SA1697, and SA2108) are also up-regulated at late-expo, but not at post-expo phase. Most of the regulators are up-regulated only at mid-expo. This corresponds to the large number of genes that are up-regulated during the mid-expo phase.

Since, *msa* is a predicted transmembrane protein (80), we hypothesize that *msa* could mediate transfer of environmental signals to these regulators, thereby bringing about the cellular response.

arcR and argR. ADI pathway is an important arginine catabolic pathway in prokaryotes, where arginine is broken down in to ornithine, ammonia and CO_2 with release of energy (ATP). Energy depletion and oxygen condition (like anaerobiasis) are the major inducers of ADI pathway (34). *arc*R is an activator of the ADI pathway (111). *arc*R expression is up-regulated in biofilms at 48hr (92), but down-regulated in different stress conditions like Peracetic acid, Stringent response, SOS response and Nitrite response (4, 23, 98). *arc*R is up-regulated by *agr*A and *tra*P (RN27 and UAMS-1) (17, 40, 70). *arc*R is also up-regulated by *msa* in the post-exponential phase. But, *arc*R is down-regulated by *sarA* in RN27 (40) and COL (This paper). Interestingly the entire *arc* operon genes (SA2424, SA2425, SA2426, SA2427, SA2428) also show expression

pattern similar to *arc*R (Figure 20). This shows that *msa* mutant produces less *arc*R, the expression resembling like those under stress condition and like *agr*A mutant. Expression level of *arc*R in *msa* is in opposite to the *sarA* mutant and biofilm phenotypes, where *arc*R is produced more (Figure 20).

*arg*R acts both as the repressor for the anabolic ornithine carbamoyltransferase and as an activator of the ADI pathway (78). *arg*R is up-regulated in biofilm (6) but *arg*R level is reduced in *msa* mutant at least in one growth phase.

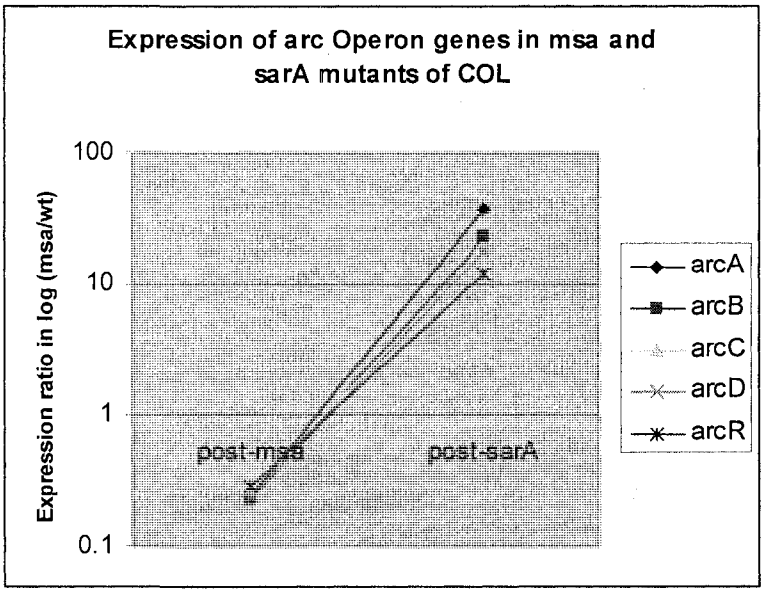


Figure 20. Expression of arc Operon genes in *msa* and *sarA* mutants

*arg*R and *arc*R helps in the adaptive response during biofilm formation, but they do not have any direct contribution for initial attachment in biofilm formation. *arg*R and *arc*R products of the *arc* operon genes are produced less in the *msa* mutant in post-exponential phase compared to the mid and early expo-phase.

Since the arc operon genes are oppositely regulated in biofilm and *msa* mutant, that could explain the reason for the defective biofilm formation by *msa* mutant that we reported earlier (96).

Effect of msa on sarA. msa transcriptome showed that *sarA* is affected by *msa. sarA* has been well defined as a global regulator in *S. aureus* (17). *msa* mutant was shown to produce less *sarA* in all three time points (96), microarray data also showed that, *sarA* expression is reduced in *msa* mutant in all three time points (only at mid-expo the p-value was significant, not in late and post-expo).

sarA expression was about 6 fold reduced in late-expo, while it was only 2 fold reduced in mid and post-expo phase. Interestingly similar *sarA* expression pattern has been reported by Cassat et al., (17), where it has been suggested that *sarA* might repress its own expression in at least some strains.

Effect of msa on agr. msa transcriptome showed that *agr* system is affected by *msa.* The *agr* system has been characterized as involved in the density dependent regulation of virulence in *S. aureus*. This system positively regulates several of the toxins and enzymes involved in virulence, while negatively regulating the some of the cell surface proteins involved in virulence. *sarA* positively regulates *agr* in RN27 (40). Thus, it has been shown that some of the down stream effects of *sarA* could be mediated through *agr* (25, 40, 53).

The entire *agr* operon (agrABCD) genes are up-regulated by *msa,* in all three growth phases. They are also up-regulated by *sarA* in COL, in the post-exponential phase (our data).

Influence of msa on expression of virulence determinants. Extracellular proteins (Exoenzymes) are affected by *msa.* There are three major proteases in *S. aureus* (67). Serine protease, Cysteine protease and metallo protease.

Several serine proteases and cysteine proteases are (*spl* and *ssp* operon genes) up-regulated by *msa. msa* mutant expressed several folds less of proteases *splA, splB, splC, splD* and *splE* in the late-expo phase, where as the expression of these genes were not affected in the mid-expo and post-expo phase. *spl* operon genes are maximally

expressed during the transition period to stationary phase and that *agr* acted as a positive regulator of these genes. *agr* genes are expressed several fold less in *msa* mutant. *msa* mutant expressed several folds less of *sspA, sspB* and *sspC* only in lat-expo phase. There was no statistically significant change in expression of these genes at mid-expo and post-expo phase.

agr upregulates *ssp*C in RN27 (40). *ssp* proteases are also up-regulated in UAMS-1 biofilm, compared to mid-expo phase planktonic cultures (6).

msa mutant also expresses less amount of *clpC, clpA* and Pfpl proteases in the late-expo phase. Other *clp* proteases *clp*P, Q, X, Y are not affected in *msa* mutant. *clp*C was shown as induced in biofilm. *clp*C is also shown as required for the stress resistance and growth recovery (24). *clp*C is expressed less in *msa* mutant, which might contribute to the rapid disintegration of biofilm formed by *msa* mutant.

Phospholipase C is up-regulated by *msa* significantly at late-expo phase. Other lipases *lip* (SA2463), *geh* (SA0309) are also affected in *msa* mutant at late-expo phase. We performed a qualitative lipolytic assay on Tributyrin agar to analyze the lipolytic phenotype of *msa* and *sarA* mutant along with the wild-type COL strain.

Our results clearly show (Figure 21) reduced lipolytic activity by the *msa* mutant (very thin zone of lysis) compared to the wild-type (large zone of lysis), while the *sarA* mutant showed very high lipolytic activity (larger zone of lysis), higher than the wild-type. This results well correlates with our microarray, where we observed that *msa* produces less lipases, while *sarA* produces a lot more lipases in comparison to the wild-type (Table 4).

S.aureus lipases have been associated with survival of the pathogen inside the host. *geh* has also been shown to be able to hydrolyze antibacterial molecule GML (glycerol monolaurate), thus playing a crucial role in pathogenesis. Lipases are also up-

regulated by *agr*, while down-regulated by *sarA*. Lipases are also down-regulated in biofilm.

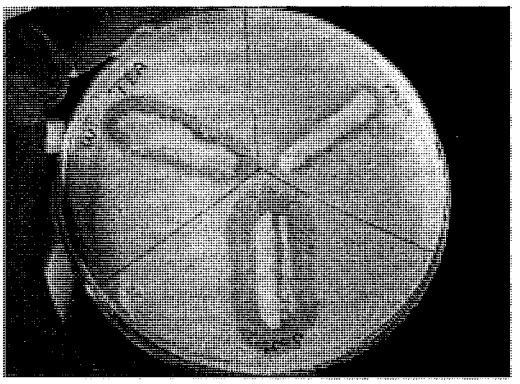


Figure 21. Tributyrin agar showing clear zones of lipolytic activity. *msa* shows very thin zone of lysis, while *sarA* shows a very large zone of lysis, since *sarA* mutant produces more of lipases, while *msa* mutant produces less of the lipases

Table 4

Expression values of lipases plc and geh obtained through sarA and msa microarray experiments. While sarA mutant shows increased expression, msa mutant shows decreased level of expression

Expression Profile of lipases in *S. aureus* COL mutants *sar*A- and *msa*- at the late-exponential phase					
COLMutant	Gene Name	N315 ID	p-value	Ratio	Fold Change
*sar*A*-	*plc*	SA0091	0.00322	12.951	12.951
	geh	SA0309	0.03070	2.098	2.098
msa-	*plc*	SA0091	0.07506	0.078	12.871
	geh	SA0309	0.23728	0.418	2.395

Toxins. Hemolysins (alpha, beta and gamma) have been shown as important virulence factors involved in *S.aureus* pathogenesis (36, 63, 82, 88). *hla* is up-regulated by *sarA* and *agr* in RN27 (40). *hla* is also several folds up-regulated by *msa* in the late and post-expo phase.

Beta and Delta hemolysins (*hlb* & *hld*) are also up-regulated by *msa* at least in late-expo phase. Beta hemolysin is up-regulated by *sarA* (31), but it did not show up in

the *sarA* transcriptome of RN27 and UAMS-1. Gamma hemolysins *hlg*A, B, C and leukocidin were also found up-regulated by *msa* at least in late-expo phase. *hlg*B and C are also up-regulated by *sarA* and *agr*A in RN27, but these hemolysins (*hlg*B and C) did not show up in the UAMS-1 *sarA* or *agr* transcriptomes. Leukotoxins *luk*D, E, M and F were all up-regulated by *msa* in late-expo phase. *luk*M and *luk*F were up-regulated by *sarA* COL transcriptome as well.

Blood agar hemolytic assays, using rabbit, sheep and human blood showed zone of hemolysis in all three blood by the wild type COL strain (Figure 22A, B, C), essentially demonstrating the production of different hemolysins (alpha, beta and delta) .

msa mutant and *sarA* mutant did not show any clear zone of hemolysis in all three blood samples after 24hrs of incubation at 37^0C. But, *sarA* mutant showed a hazy, dull zone of hemolysis in sheep blood agar (Figure 22C1). Overnight refrigeration of the sheep agar plate with 24 hr grown cultures at 37^0C, cleared the haziness in the *sarA* zone of hemolysis (Figure 22C2). This showed that *sarA* mutant produced beta hemolysin, while *msa* mutant did not.

Our microarray array results also show that *msa* failed to produce any of the hemolysins, while *sarA* mutant transcriptome showed only the hla reduction in the mutant (Table 5).

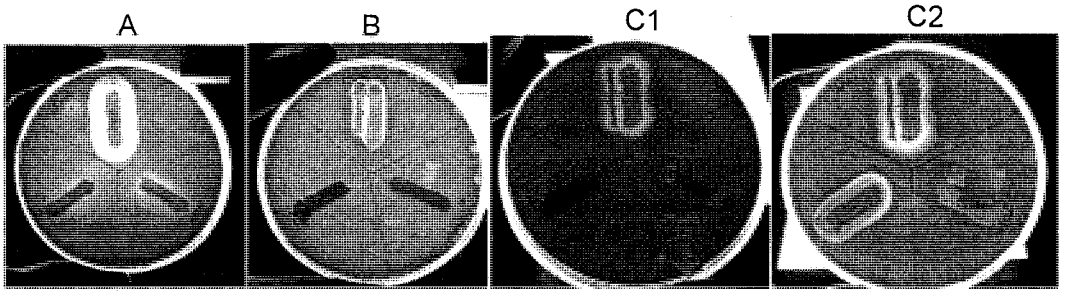


Figure 22. A) Rabbit blood agar. B) Human blood agar. C1) Sheep blood agar after 24 hr incubation at 37C showing hazy zone of hemolysis in *sarA* mutant. C2) C1 plate overnight refrigerated, showing cleared zone of hemolysis in *sarA* mutant.

Table 5

Expression values of hemolysins hla, hld and hlb obtained through sarA and msa microarray experiments. Both sarA mutant and msa mutant show decreased level of expression

Expression Profile of hemolysins in *S. aureus* COL mutants *sar*A- and *msa*- at post-exponential phase					
COLMutant	Gene Name	N315 ID	p-value	Ratio	Fold Change
sarA-	*hla*	SA1007	0.00778	0.374	2.676
	hld	SAS1940a	0.00007	0.004	232.486
	*hlb**				
msa-	*hla*	SA1007	0.04486	0.096	10.375
	hld	SAS1940a	0.00878	0.019	51.733
	hlb	SA1811	0.51296	0.413	2.422

Enterotoxin *seb* (*sec*3) is up-regulated by *msa* in late and post-expo phase. *seb* is also up-regulated by *sarA* in COL (this paper) (Table 6). *seb* is regulated by *agr* system in a *rot* dependent manner (102), but *rot* expression is not affected in *msa* mutant. *seb* is positively regulated by *sarA* (19).

Our western blot analysis of the COL wild type, *msa* and *sarA* mutants using polyclonal antibody against *seb*, clearly showed that *sarA* mutants failed to produce *seb*, while seb expression was relatively less in *msa* mutant compared to the wild type (Figure 23. Westernblot of COL wild type, *msa* mutant and *sarA* mutant, using anti-seb).

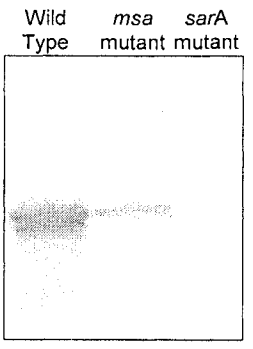


Figure 23. Westernblot of COL wild type, *msa* mutant and *sarA* mutant, using anti-seb

Table 6

Expression values of seb obtained through sarA and msa microarray experiments. Both sarA mutant and msa mutant show decreased level of expression

Expression Profile of SEB in *S. aureus* COL mutants *sar*A- and *msa*- at the post-exponential phase					
COLMutant	N315 ID	Other ID	p-value	Ratio	Fold Change
*sar*A-	SA1817	SAV2009	0.00001	0.003	390.452
	SA1817	SACOL0907	0.00027	0.003	309.758
msa-	SA1817	SACOL0907	0.02141	0.156	6.426
	SA1817	SAV2009	0.01192	0.108	9.241

spa is a major staphylococcal surface protein, involved in virulence. *spa* expression is down-regulated by *agr* and *sarA* in RN27 (40, 45). Our transcriptomic data shows that *spa* is down-regulated by *msa* several folds. Expression of *spa* in *msa* mutant gradually increases from mid-expo to post-expo. *spa* expression is also down-regulated by *sarA* in COL.

Adhesion of *S. aureus* to foreign materials during biofilm formation depends on the expression of *clf*A and *clf*B (103). *clf*B is up-regulated by *agr* in RN27 (40). *clf*A and *clf*B are up-regulated in biofilm (92). *clf*B also promotes adherence (84). *clf*A level is significantly reduced in *msa* planktonic and biofilm cultures (96). *msa* transcriptome shows that both *clf*A and B are up-regulated by *msa* more than 3 fold at the post-expo phase. This correlates well with our previous results (96) and moreover, *clf*A and B expression are not affected by *sarA* in COL (this paper). This shows that both *clf*A and *clf*B could be modulated by *msa* independent of *sarA*.

*sdr*C and *sdr*E are up-regulated by *msa* at the post-expo phase. *sdr*C and *sdr*E are members of the serine-aspartate repeat (Sdr) family of proteins that have been identified as mediating interactions of *S. aureus* with the extracellular matrix (47). *sdr*C is a Ser-Asp-rich fibrinogen-binding protein (92). Sdr family proteins have both

organizational and sequence similarity to ClfA and ClfB proteins (64). *sdr*C is up-regulated in biofilm (92). Reduced levels of *sdr*C and *sdr*E in *msa* mutant could also contribute to the defective biofilm phenotype in *msa*. *sdr*C and *sdr*E are not affected by *sarA* in COL.

Stress response proteins. Cold shock protein *csp*A, alkaline shock protein *asp*23 and a general stress protein 39 are all up-regulated by *msa* in the mid and late-expo phase. *csp*A is down-regulated in Peracetic acid (20min) (23) and under Chlorination (20min) (21).

*csp*A is down-regulated by *sarA* in COL. *csp*C is down-regulated by *msa* at mid-expo phase and also down-regulated by *sarA* in COL. *csp*C is down-regulated in Nitrite stress (98).

*asp*23 is up-regulated by *msa* as well as by *sarA* in COL. *asp*23 is also up-regulated in biofilm and in acid shock (11, 92). Even though Asp23 is an alkaline pH tolerance protein, it is interesting to note that its expression is induced in biofilm and in acid shock conditions. *asp*23 is also induced by hydrogen peroxide (22). This shows that *asp*23 could contribute to the adaptive response of *S. aureus* to change in pH. *asp*23 is down-regulated in Nitrite stress, which could be contributing for defective biofilm formation under Nitrite stress.

The general stress protein 39 (SA2119) up-regulated by *msa* in the mid and late-expo phase is down-regulated by *sarA* in COL. This protein is also up-regulated under acid-shock condition (11) and Anaerobic condition (43).

While *csp*A and the general stress protein SA2119 seems to be oppositely regulated by *msa* and *sarA*, *csp*C and *asp*23 seems to be regulated by *msa* through *sarA*.

Capsular polysaccharide synthesis genes. Our microarray results showed that *cap* genes (*cap* A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P) are significantly up-regulated by *msa* (less expression) at the late-expo phase, while not greatly affected in other time points (Figure 24). qRT-PCR on *cap*A gene showed its expression profile similar to that of the Microarray results.

Expression of *cap in vitro* is highly sensitive to various environmental signals. *cap*A,B,C,D,E,F,G,H,I,J,L,M,N,O operon genes are up-regulated by *sig*B and mild acid (8, 109). *cap*A,B,C,D,E,F,G genes are up-regulated in biofilm (6), while *cap*A,C,D,E,F,G,I,K,L,M are down-regulated by nitrite stress (98).

cap operon genes are positively controlled by *agr* (less expression in early exponential, while maximal production occurs during post exponential phase). Capsule production is under the regulation of a variety of regulators (*agr, mgrA, sbc*DC, *clp*C,*sae* and *sig*B) (8, 77). SAMMD showed that *cap*J is up-regulated by *sarA* and *agr* (RN27).

*sar*A in COL also showed *cap*B and *cap*G significantly up-regulated. Other *cap* genes were also up-regulated several folds, but the p-value was not significant. Cassat et al. (17) showed that *agr* upregulates "*cap*" genes at the post-exponential phase. Our microarray results showed that *msa* upregulates *agr*. Microarray data also showed that the *cap* genes are up-regulated by *msa.* This upregulation of *cap* genes by *msa* could be either dependent or independent of *agr*. But considering the fact that *msa* is a putative integral membrane protein and that capsule production is highly sensitive to various environmental signals (85), it is likely that *msa* upregulates *cap* genes through *agr*.

Inverse correlation has been observed between capsule production and infectivity, which might be due to capsule masking the adhesins that have been shown to be important determinants of virulence (85, 90).

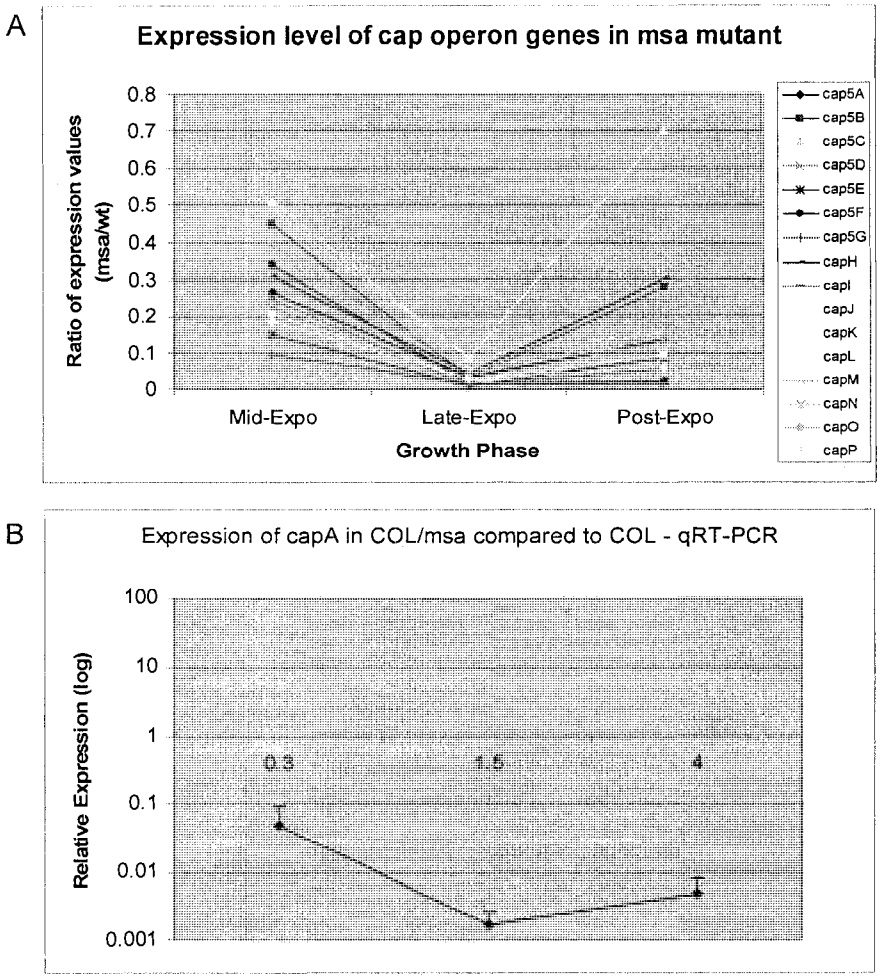


Figure 24. A) Expression profile of cap operon genes from Microarray data. All the late-expo values had significant p-values, some of the mid-expo and post-expo values did not have significant p-values, and two missing values were imputed with the average value. B) Expression profile of capA gene from qRT-PCR.

Influence of msa on Known 2 Component Systems

msa didn't show any effect on most of the two-component systems (*pho*RP, *srr*AB, *vic*K, *vra*R), but it did affect *sae*RS like *sarA*. *kdp*DE two-component system was affected by *msa,* as was seen in biofilms (*kdp*DE expression was high in *msa* mutant and in biofilm).

Influence of msa on biofilm forming genes. Bacteria need to overcome acidic environment during biofilm formation, usually this is achieved by producing alkaline compounds (ammonia). Microarray data shows that genes involved in ammonia production pathway (arginine deiminase) are induced by *msa.*

The arc operon. ADI or arginine deiminase pathway plays an important role in biofilm formation, by utilizing arginine to generate ATP, in anaerobic conditions, as well as in maintaining alkaline pH, by releasing ammonia. Energy depletion and oxygen condition (like anaerobiasis) are the major inducers of ADI pathway (34). *arc*R is an activator of the ADI pathway (111).

*arc*R is up-regulated in biofilms (6). The entire *arc* operon is also up-regulated in biofilms at 48hr (92).

*arc*R helps in the adaptive response during biofilm formation, but does not have any direct contribution for initial attachment in biofilm formation. *arc* operon genes are down-regulated under nitrite stress (98). Nitrite stress adversely affects biofilm formation in *S. aureus* (98). The entire *arc* operon is up-regulated by *agr*A both in UAMS-1 and RN27 strains (17, 40).

*arc*R level is significantly reduced in *msa* mutant at the post-expo phase. Other arginine deiminase cluster (*arc*A,B,C,D) genes are also expressed at significantly low level during the post-expo phase in *msa* mutant, compared to the mid-expo phase (expression values even though were not much affected in mid and late-expo phase, they also did not have a significant p-value).

*arc*R and *arc*C are down-regulated by *sarA* (RN27), while the entire operon (*arc*ABCDR) is significantly down-regulated by *sarA* in COL (our data).

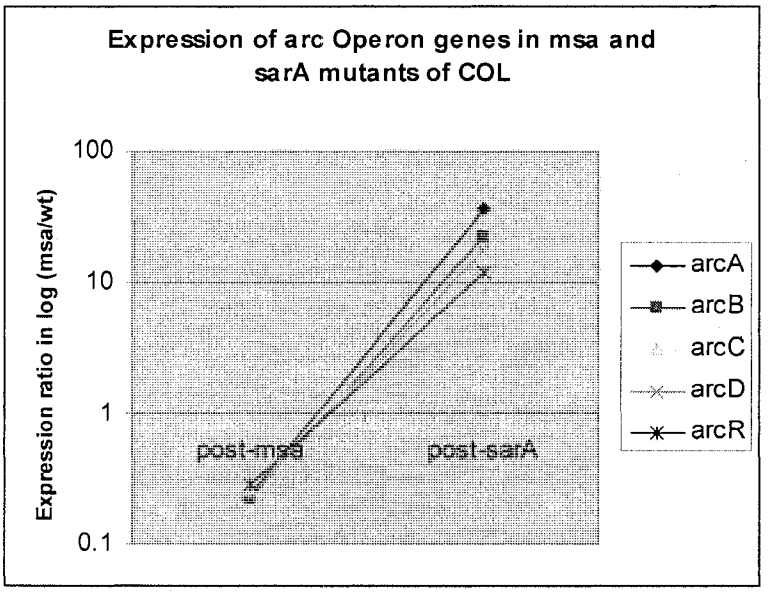


Figure 25. Expression profile of arc operon genes in the post-expo phase *msa* mutant and post-expo phase *sarA* mutant. Profile shows that arc operon genes are produced less in *msa* mutant, while it is produced more in *sarA* mutant.

arc expression is oppositely regulated in biofilm and *msa* mutant (Figure 25). This could be one of the reasons for the observed defective biofilm phenotype in the *msa* mutant. Nitrite stress affects biofilm formation and *arc* operon genes are down-regulated under nitrite stress. *arc* operon genes are also expressed less in *msa* mutant and that could attribute to the defect in *msa* mutants biofilm.

Other biofilm genes. Sdr CDE has been show to affect biofilm formation, by involving in initial attachment (97). Sdr C and E are significantly up-regulated by *msa.* More than 3 fold reduction in *sdr*C and E was observed in the *msa* mutant. No significant change was seen in *sarA* mutant.

*clf*B promotes adherence (84). *clf*B expression is up-regulated in biofilm (92). *clf*B expression has also been reported as up-regulated by *agr*A (40). *clf*B expression is significantly down-regulated under nitrite stress condition, where biofilm defective

phenotype has been observed. *clf*B expression is significantly reduced in the *msa* mutant, but not in the *sarA* mutant. This could also contribute to our previously observed defective biofilm phenotype in *msa* mutant.

Nitrate is reduced to Nitrite by nitrate reductase. Nitrite could then be reduced to nitric oxide or ammonia. But, under hypoxic conditions, like under biofilm conditions or when pH is around 5 or when the acidification of the growth medium occurs, like under biofilm conditions, nitrite is reduced to nitric oxide, which is a reactive oxygen species. Nitric oxide generated under this condition, has been shown to represses biofilm formation (98). The *nir* operon genes *nirD,nirE,nirR* and *nas*F are all expressed significantly higher in *msa* mutant at the post-expo phase. *nir* genes product (nitrite reductase) convert nitrite in to ammonia or nitric oxide. The *nir* operon genes are also expressed at least 2 fold higher in the *sarA* mutants. The *nar* operon gene (nitrate reductase) *nar*G and K are expressed significantly higher in *msa* mutant at the post-expo phase. Other *nar* operon genes *nar* I, J and K are expressed higher but p-value was not significant.

The entire *pur* operon genes (*pur* A,D,E,F,H,K,L,M,N,Q,S) are down-regulated in the acid-shock condition. *pur* operon genes (*pur* K, C, S, D) are also down-regulated in the biofilm . The entire *pur* operon genes are highly expressed in the *sarA* mutant of COL (statistically significant). *pur*M has been reported as highly expressed in *sarA* mutant of RN27 (40). In the *msa* mutant, all the *pur* operon genes are expressed less during the late-expo phase. But, they are not affected much during the mid and post-exponential phase.

*als*SD operon encodes for alpha-acetolactate synthase and alpha-acetolactate decarboxylase, which are involved in acetoin production (49, 91, 110). Cassat showed that mutation of *als*S,D results in reduced biofilm formation (17) and that reduced level was comparable to the *sarA* mutants biofilm forming capacity. They also argue that

acetoin production by *als*SD is important for survival in the acidic environment, like in biofilm. Yang et al., 2006 showed that disruption of *als*SD operon results in increased stationary-phase killing in UAMS-1. *als*S,D is found at increased levels in biofilm. *als*SD expression is also induced under peracetic acid, mild acid, biofilm and acid-shock conditions. *als*SD is found at reduced levels in *msa* mutant at the late-exponential phase, but was found in higher levels in the post-exponential phase (also found at reduced levels in *sarA* mutant (17)). Relatively less expression of *als*SD in the *msa* mutant during the late-expo phase could contribute to its defective biofilm phenotype.

*arl*RS has been shown to act as repressor of biofilm formation, with a major role in the initial attachment or primary adherence (42, 101). *arl*R also positively regulates capsule production (77). *arl*RS production was significantly low in *msa* mutant during the mid and late-exponential phase. But there was no significant change in *arl*RS expression during the post-expo phase.

Autolysins are peptidoglycan hydolases. They digest the cell-wall peptidoglycan of the gram positive bacteria. Autolysins are involved in cell division and cell separation. The gene *atl* encodes for a major autolysin (41, 86). *atl* mutants showed rough cell surface, formed clumps of cells. *atl* mutant is also deficient in primary attachment. *atl* expression is more than 3 fold significantly reduced in *msa* mutant in the post-expo phase.

aaa (SA0423) is a multifunctional autolysin/adhesin. It has been shown to possess adhesion properties and as involved in colonization (52). There was no significant change in the aaa expression in *msa* mutant.

*lyt*RS controls the rate of autolysis. *lyt*S mutants showed increased autolysis (14). *lyt*RS expression in *msa* mutant is more than 2 fold reduced in the late-expo phase (but p-value not significant). *lyt*RS positively regulates *lrg*AB genes (14). *lrg*AB are

involved in murein hydrolase activity. *lrg*AB expression is more than 3 fold reduced in *msa* mutant, in the late-exponential phase (but p-value not significant).

Influence of msa on transporters and surface proteins. msa upregulates 33 genes involved in transport and binding. Relatively higher percentage of genes involved in transport and binding were up-regulated in the mid and post-expo phase, than the late-exponential phase. No single gene involved in transport and binding is significantly down-regulated by *msa.* In contrast, *sarA* downregulates 12 genes involved in transport and binding, while upregulating only 4 of them.

Several known and unknown transport systems are regulated by *msa.* Several of the genes up-regulated by *msa* are involved in "Amino acids, peptides and amines" and "Cation and iron carrying" category. It was interesting to note that almost all of the genes that are involved in "Cations and iron carrying" are up-regulated only in the mid and late-expo phase. Most of the genes under t his category are "Na+/H+ antiporters". These type of transporters are involved in pH balance, aid symport of other molecules including drugs efflux (56).

Transporters involved in lactate, formate, and citrate transport are up-regulated by *msa* in the late-exponential phase. 10 ABC type transporters were among those up-regulated by *msa.* ABC transporters play crucial role in the transport of amino acids, carbohydrates, oligopeptides and drug molecules.

msa upregulates 31 cell envelope related genes, while downregulating just one (SA0129). Similar expression pattern is seen in *sarA* of COL (more 12 up-regulated and 3 down-regulated). Out of 31 up-regulated genes, almost all of them except 6 (*atl, sdrC, sdrE*, SA2006, SA2158, SA2320) are up-regulated only during the mid or late-exponential phase. A majority of the genes up-regulated during the late-exponential phase are capsular polysaccharide synthesis related.

Antibiotic resistance. Several genes involved in antibiotic resistance are up-regulated by *msa*. SA0151, which is an antibiotic acetyl transferase, is several folds reduced in the *msa* mutant, significantly at the late-expo phase. Antibiotic acetyl transferases confer resistance to antibiotics by attaching an acetyl group to the antibiotic and thereby making it unable to bind to its target. SA1269, a drug transporter (which belongs to the Major Facilitator Superfamily –MFS) is several folds reduced in the *msa* mutant, significantly during the post-expo phase. MFS proteins are secondary carriers that transport solutes in response to stress. SA1269 has been characterized as a quinolone efflux transporter gene and has been named as *norB* . *norB* confers resistance to quinolones like norfloxacin, ciprofloxacin, sparfloxacin and moxifloxacin. SA2203, a drug resistance transporter belonging to the EmrB/QacA subfamily, is significantly up-regulated by *msa* in the late and post-expo phase. This is the family of drug efflux proteins, a part of the major facilitator family. Another hypothetical major facilitator family drug transporter SA0099 is also significantly up-regulated more than 2 fold by *msa* in the mid-expo phase (it is also up-regulated more than 2 fold in the late and post-expo, but p-value was not significant).

Differences and Similarities with sar regulon

In part, *msa* transcriptome looks like *sarA* transcriptome (comparison of the *msa* post-expo and *sarA* post-expo transcriptomes). About 50% of *msa* up-regulated genes (24 out of 52) are also up-regulated by *sarA* (Figure 26, 27). While 28 genes are up-regulated only in *msa,* 69 are up-regulated only in *sarA*.

While several of the *msa* up-regulated genes are also up-regulated by *sarA*, only a very few (5) genes that are down-regulated by *msa* are also down-regulated by *sarA* (Figure 26). There were 15 genes down-regulated only by *msa,* while 74 genes are down-regulated only by *sarA*.

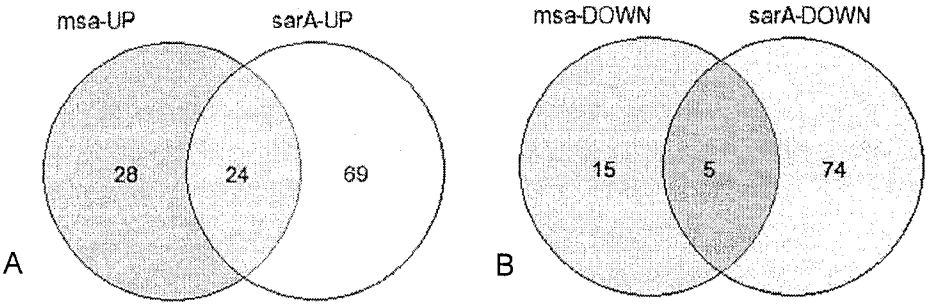


Figure 26. Venn diagram of genes up-regulated (A) and down-regulated (B) by *msa* and *sarA* at post-expo phase

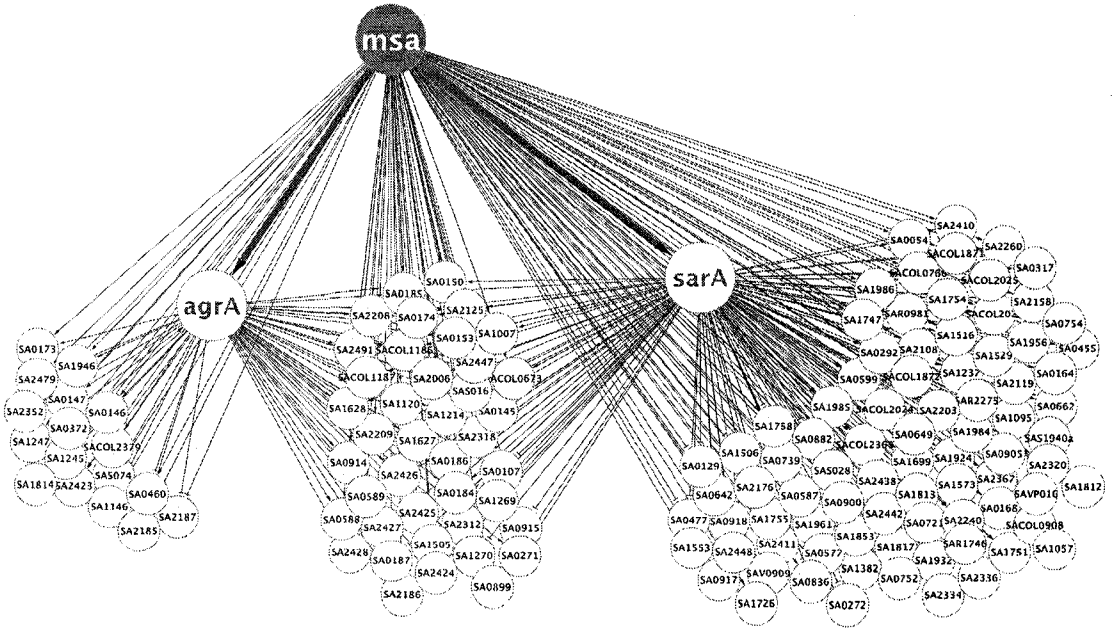


Figure 27. A network representation of genes regulated by *msa, sarA* and agr. *S. aureus* N315 IDs are shown in the circles.

Categorization of the genes regulated by msa and sarA. Gene Ontology categorization shows that even though several genes involved in energy metabolism and Transport are differentially regulated by *msa* and *sarA*, only a very few of them overlap. I.e., only a few of the genes involved in energy metabolism and transport are regulated by both *sarA* and *msa* (Figure 28).

This could mean that genes involved in these two categories are regulated independent of each other for most of the time. Similarly, genes involved in Transport and binding seems to be regulated by *msa* independent of *sarA*, even though this category of genes is oppositely regulated by *msa* and *sarA* (*sarA* mostly downregulates transport genes, while *msa* mostly upregulates transport genes).

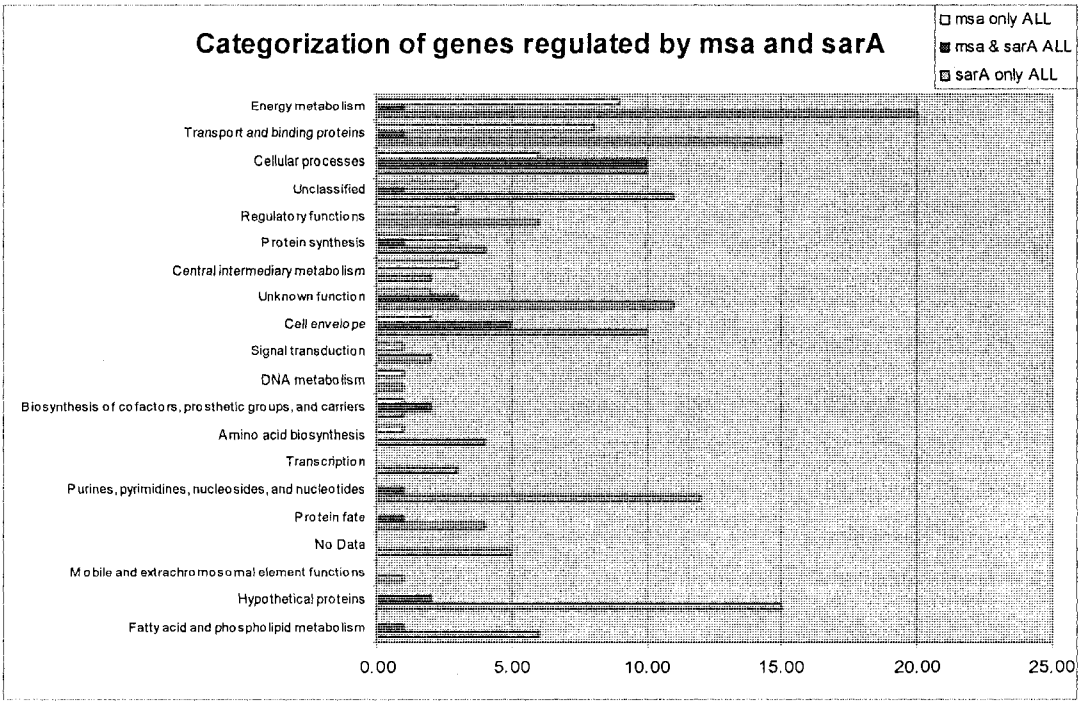


Figure 28. Gene Ontology categorization of genes regulated by *sarA* only, *msa* only and by both *sarA* and *msa* (both up and down regulated).

Interestingly, several of the genes involved in the cellular processes are regulated both by *sarA* and *msa,* infact a majority of the cellular processes genes regulated by *msa* are also regulated by *sarA*. This could mean that genes involved in cellular processes (Pathogenesis, Toxin production etc.,) are regulated by *msa* in a *sarA* dependent manner.

Opposite regulation. There were 7 genes oppositely regulated by *msa* and *sarA*. 5 of those 7 genes belong to the *arc* operon (*arc*A,B,C,D,R) and the other two are SA1269 and SA1270 (both are putative transport permeases). ADI or arginine deiminase pathway plays an important role in biofilm formation, by utilizing arginine to generate ATP, in anaerobic conditions, as well as in maintaining alkaline pH, by releasing ammonia. Energy depletion and oxygen condition (like anaerobiasis) are the major inducers of ADI pathway (34). *arc*R is an activator of the ADI pathway (111). *arc*R is up-regulated in biofilms (6). The entire *arc* operon is up-regulated in biofilms at 48hr (92). *arc*R helps in the adaptive response during biofilm formation, but does not have any direct contribution for initial attachment in biofilm formation. *arc* operon genes are down-regulated under nitrite stress (98). The entire *arc* operon is up-regulated by *agr*A both in UAMS-1 and RN27 strains (17, 40).

*arc*R level is significantly reduced in *msa* mutant at the post-expo phase. Other arginine deiminase cluster (*arc*A,B,C,D) genes are also expressed at significantly low level during the post-expo phase in *msa* mutant, compared to the mid-expo phase (expression values even though were not much affected in mid and late-expo phase, they also did not have a significant p-value). *arc*R and *arc*C are down-regulated by *sarA* (RN27), while the entire operon (*arc*ABCDR) is significantly down-regulated by *sarA* in COL (our data).

arc expression is oppositely regulated in biofilm and *msa* mutant. This could be one of the reasons for the observed defective biofilm phenotype in the *msa* mutant. Nitrite stress affects biofilm formation and *arc* operon genes are down-regulated under nitrite stress. *arc* operon genes are also expressed less in *msa* mutant and that could attribute to the defect in *msa* mutants biofilm. Interestingly *msa* is also up-regulated in biofilm (6) and peracetic acid (23) conditions in the early stages (OD 1.0 in biofilm and

10 min in peracetic acid). These data show that *msa* does have an important role in the biofilm formation in *S. aureus*.

CHAPTER V

SUMMARY

We earlier showed that *msa* modulates the expression of *sarA.* Also, sarA regulated genes are also regulated by *msa.* To further characterize the role of *msa*, I first attempted to predict its structure and functional properties using computational tools. The structure and function prediction studies showed that *msa* is a putative transmembrane protein, with a distinct N-terminal signal peptide, IN-OUT topology, 4 phosphorylation sites (two inside and two outside membrane) and a putative cytoplasmic binding site. Based on these predictions I hypothesize that *msa* is a potential signal transducer, which could play an important role in the global regulation of virulence in *S. aureus*, by mediating the interaction between the environmental signals and the regulatory components.

To compare and contrast the *Staphylococcus aureus* transcriptomes I designed and developed the *Staphylococcus aureus* Microarray Meta-Database (SAMMD). SAMMD was developed as a web-based open-source resource and is hosted as a bioinformatics.org project. SAMMD consists of manually curated microarray data from more than 90 different *S. aureus* microarray experiments.

I generated the *msa* and *sarA* transcriptome using microarray technology. Analsysis of the *msa* transcriptome showed that *msa* modulates global regulation of virulence in *S. aureus.* I also observed that *msa* modulates global regulation of virulence genes in a growth-phase dependent manner. Microarray results showed that *msa* modulates activity of genes involved in transport, metabolism and virulence globally. I also confirmed some of the microarray results by performing molecular and phenotypic assays. Comparative transcriptomic studies on the *msa* transcriptome using SAMMD showed that *msa* regulates genes in both *sarA*-dependent and *sarA*-independent

manner. Virulence genes are regulated by *msa* in a *sarA*-dependent manner, while genes involved in metabolism are regulated by *msa* in a *sarA*-independent manner.

The structure and function prediction results would help us to design rationale experiments to test the hypothesis that *msa* is a novel signal transducer. We will also direct our future studies towards understanding how *msa* transduces the environmental signals to the global regulators. We have planned on developing microarray meta-databases for other important pathogens and including novel data-mining tools to help query and visualize data from these meta-database.

APPENDIX

PUBLICATIONS

Parts of this dissertation work have been published in peer-reviewed journals.

Nagarajan, V. and M. O. Elasri (2007). "Structure and function predictions of the Msa protein in *Staphylococcus aureus*." *BMC Bioinformatics* 8 Suppl 7: S5.

Nagarajan, V. and M. O. Elasri (2007). "SAMMD: *Staphylococcus aureus* microarray meta-database." *BMC Genomics* 8: 351.

Nagarajan, V. and O. M. Elasri (2008). Case Study: Structure and Function Prediction of a Protein with No Functionally Characterized Homolog. *Computational Intelligence in Biomedicine and Bioinformatics : Current Trends and Applications*. G. T. Smolinski, G. M. Milanova and A.-E. Hassanien. Berlin, Heidelberg, Springer-Verlag: 379-395.

REFERENCES

1. ProSAL. http://xray.bmc.uu.se/sbnet/prosal.html
2. **Alexandrov, N. N., R. Nussinov, and R. M. Zimmer.** 1996. Fast protein fold recognition via sequence to structure alignment and contact capacity potentials. Pac Symp Biocomput:53-72.
3. **Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res **25:**3389-402.
4. **Anderson, K. L., C. Roberts, T. Disz, V. Vonstein, K. Hwang, R. Overbeek, P. D. Olson, S. J. Projan, and P. M. Dunman.** 2006. Characterization of the *Staphylococcus aureus* heat shock, cold shock, stringent, and SOS responses and their effects on log-phase mRNA turnover. J Bacteriol **188:**6739-56.
5. **Bannai, H., Y. Tamada, O. Maruyama, K. Nakai, and S. Miyano.** 2002. Extensive feature detection of N-terminal protein sorting signals. Bioinformatics **18:**298-305.
6. **Beenken, K. E., P. M. Dunman, F. McAleese, D. Macapagal, E. Murphy, S. J. Projan, J. S. Blevins, and M. S. Smeltzer.** 2004. Global gene expression in *Staphylococcus aureus* biofilms. J Bacteriol **186:**4665-84.
7. **Bendtsen, J. D., H. Nielsen, G. von Heijne, and S. Brunak.** 2004. Improved prediction of signal peptides: SignalP 3.0. J Mol Biol **340:**783-95.
8. **Bischoff, M., P. Dunman, J. Kormanec, D. Macapagal, E. Murphy, W. Mounts, B. Berger-Bachi, and S. Projan.** 2004. Microarray-based analysis of the *Staphylococcus aureus* sigmaB regulon. J Bacteriol **186:**4085-99.
9. **Blevins, J. S., M. O. Elasri, S. D. Allmendinger, K. E. Beenken, R. A. Skinner, J. R. Thomas, and M. S. Smeltzer.** 2003. Role of *sarA* in the pathogenesis of *Staphylococcus aureus* musculoskeletal infection. Infect Immun **71:**516-23.
10. **Blevins, J. S., A. F. Gillaspy, T. M. Rechtin, B. K. Hurlburt, and M. S. Smeltzer.** 1999. The Staphylococcal accessory regulator (sar) represses transcription of the *Staphylococcus aureus* collagen adhesin gene (cna) in an agr-independent manner. Mol Microbiol **33:**317-26.
11. **Bore, E., S. Langsrud, O. Langsrud, T. M. Rode, and A. Holck.** 2007. Acid-shock responses in *Staphylococcus aureus* investigated by global gene expression analysis. Microbiology **153:**2289-303.
12. **Brendel, V., P. Bucher, I. R. Nourbakhsh, B. E. Blaisdell, and S. Karlin.** 1992. Methods and algorithms for statistical analysis of protein sequences. Proc Natl Acad Sci U S A **89:**2002-6.
13. **Bronner, S., H. Monteil, and G. Prevost.** 2004. Regulation of virulence determinants in *Staphylococcus aureus*: complexity and applications. FEMS Microbiol Rev **28:**183-200.

14. **Brunskill, E. W., and K. W. Bayles.** 1996. Identification of LytSR-regulated genes from *Staphylococcus aureus*. J Bacteriol **178:**5810-2.
15. **Bryson, K., L. J. McGuffin, R. L. Marsden, J. J. Ward, J. S. Sodhi, and D. T. Jones.** 2005. Protein structure prediction servers at University College London. Nucleic Acids Res **33:**W36-8.
16. **Cai, C. Z., L. Y. Han, Z. L. Ji, X. Chen, and Y. Z. Chen.** 2003. SVM-Prot: Web-based support vector machine software for functional classification of a protein from its primary sequence. Nucleic Acids Res **31:**3692-7.
17. **Cassat, J., P. M. Dunman, E. Murphy, S. J. Projan, K. E. Beenken, K. J. Palm, S. J. Yang, K. C. Rice, K. W. Bayles, and M. S. Smeltzer.** 2006. Transcriptional profiling of a *Staphylococcus aureus* clinical isolate and its isogenic *agr* and *sarA* mutants reveals global differences in comparison to the laboratory strain RN6390. Microbiology **152:**3075-90.
18. **Cassat, J. E., P. M. Dunman, F. McAleese, E. Murphy, S. J. Projan, and M. S. Smeltzer.** 2005. Comparative genomics of *Staphylococcus aureus* musculoskeletal isolates. J Bacteriol **187:**576-92.
19. **Chan, P. F., and S. J. Foster.** 1998. The role of environmental factors in the regulation of virulence-determinant expression in *Staphylococcus aureus* 8325-4. Microbiology **144 (Pt 9):**2469-79.
20. **Chan, P. F., and S. J. Foster.** 1998. Role of SarA in virulence determinant production and environmental signal transduction in *Staphylococcus aureus*. J Bacteriol **180:**6232-41.
21. **Chang, M. W., F. Toghrol, and W. E. Bentley.** 2007. Toxicogenomic response to chlorination includes induction of major virulence genes in *Staphylococcus aureus*. Environ Sci Technol **41:**7570-5.
22. **Chang, W., D. A. Small, F. Toghrol, and W. E. Bentley.** 2006. Global transcriptome analysis of *Staphylococcus aureus* response to hydrogen peroxide. J Bacteriol **188:**1648-59.
23. **Chang, W., F. Toghrol, and W. E. Bentley.** 2006. Toxicogenomic response of *Staphylococcus aureus* to peracetic acid. Environ Sci Technol **40:**5124-31.
24. **Chatterjee, I., P. Becker, M. Grundmeier, M. Bischoff, G. A. Somerville, G. Peters, B. Sinha, N. Harraghy, R. A. Proctor, and M. Herrmann.** 2005. *Staphylococcus aureus* ClpC is required for stress resistance, aconitase activity, growth recovery, and death. J Bacteriol **187:**4488-96.
25. **Cheung, A. L., M. G. Bayer, and J. H. Heinrichs.** 1997. sar Genetic determinants necessary for transcription of RNAII and RNAIII in the *agr* locus of *Staphylococcus aureus*. J Bacteriol **179:**3963-71.
26. **Cheung, A. L., K. J. Eberhardt, E. Chung, M. R. Yeaman, P. M. Sullam, M. Ramos, and A. S. Bayer.** 1994. Diminished virulence of a sar-/agr- mutant of *Staphylococcus aureus* in the rabbit model of endocarditis. J Clin Invest **94:**1815-22.
27. **Cheung, A. L., J. M. Koomey, C. A. Butler, S. J. Projan, and V. A. Fischetti.** 1992. Regulation of exoprotein expression in *Staphylococcus*

aureus by a locus (sar) distinct from agr. Proc Natl Acad Sci U S A **89:**6462-6.

28. **Cheung, A. L., C. C. Nast, and A. S. Bayer.** 1998. Selective activation of sar promoters with the use of green fluorescent protein transcriptional fusions as the detection system in the rabbit endocarditis model. Infect Immun **66:**5988-93.
29. **Cheung, A. L., and S. J. Projan.** 1994. Cloning and sequencing of *sarA* of *Staphylococcus aureus*, a gene required for the expression of agr. J Bacteriol **176:**4168-72.
30. **Cheung, A. L., M. R. Yeaman, P. M. Sullam, M. D. Witt, and A. S. Bayer.** 1994. Role of the sar locus of *Staphylococcus aureus* in induction of endocarditis in rabbits. Infect Immun **62:**1719-25.
31. **Cheung, A. L., and P. Ying.** 1994. Regulation of alpha- and beta-hemolysins by the sar locus of *Staphylococcus aureus*. J Bacteriol **176:**580-5.
32. **Claros, M. G., and G. von Heijne.** 1994. TopPred II: an improved software for membrane protein structure predictions. Comput Appl Biosci **10:**685-6.
33. **Cserzo, M., E. Wallin, I. Simon, G. von Heijne, and A. Elofsson.** 1997. Prediction of transmembrane alpha-helices in prokaryotic membrane proteins: the dense alignment surface method. Protein Eng **10:**673-6.
34. **Cunin, R., N. Glansdorff, A. Pierard, and V. Stalon.** 1986. Biosynthesis and metabolism of arginine in bacteria. Microbiol Rev **50:**314-52.
35. **D'Souza, M., E. M. Glass, M. H. Syed, Y. Zhang, A. Rodriguez, N. Maltsev, and M. Y. Galperin.** 2007. Sentra: a database of signal transduction proteins for comparative genome analysis. Nucleic Acids Res **35:**D271-3.
36. **Dajcs, J. J., B. A. Thibodeaux, D. O. Girgis, and R. J. O'Callaghan.** 2002. Corneal virulence of *Staphylococcus aureus* in an experimental model of keratitis. DNA Cell Biol **21:**375-82.
37. **David, T. E., G. Gavra, C. M. Feindel, T. Regesta, S. Armstrong, and M. D. Maganti.** 2007. Surgical treatment of active infective endocarditis: a continued challenge. J Thorac Cardiovasc Surg **133:**144-9.
38. **de Castro, E., C. J. Sigrist, A. Gattiker, V. Bulliard, P. S. Langendijk-Genevaux, E. Gasteiger, A. Bairoch, and N. Hulo.** 2006. ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. Nucleic Acids Res **34:**W362-5.
39. **Deleage, G., C. Blanchet, and C. Geourjon.** 1997. Protein structure prediction. Implications for the biologist. Biochimie **79:**681-6.
40. **Dunman, P. M., E. Murphy, S. Haney, D. Palacios, G. Tucker-Kellogg, S. Wu, E. L. Brown, R. J. Zagursky, D. Shlaes, and S. J. Projan.** 2001. Transcription profiling-based identification of *Staphylococcus aureus* genes regulated by the *agr* and/or *sarA* loci. J Bacteriol **183:**7341-53.

41. **Foster, S. J.** 1995. Molecular characterization and functional analysis of the major autolysin of *Staphylococcus aureus* 8325/4. J Bacteriol **177:**5723-5.
42. **Fournier, B., and D. C. Hooper.** 2000. A new two-component regulatory system involved in adhesion, autolysis, and extracellular proteolytic activity of *Staphylococcus aureus*. J Bacteriol **182:**3955-64.
43. **Fuchs, S., J. Pane-Farre, C. Kohler, M. Hecker, and S. Engelmann.** 2007. Anaerobic gene expression in *Staphylococcus aureus*. J Bacteriol **189:**4275-89.
44. **Galperin, M. Y.** 2005. A census of membrane-bound and intracellular signal transduction proteins in bacteria: bacterial IQ, extroverts and introverts. BMC Microbiol **5:**35.
45. **Gao, J., and G. C. Stewart.** 2004. Regulatory elements of the *Staphylococcus aureus* protein A (Spa) promoter. J Bacteriol **186:**3738-48.
46. **Gasteiger, E., C. Hoogland, A. Gattiker, S. Duvaud, M. R. Wilkins, R. D. Appel, and A. Bairoch.** 2005. Protein Identification and Analysis Tools on the ExPASy Server, p. 571-607. *In* J. M. Walker (ed.), The Proteomics Protocols Handbook. Humana Press.
47. **George, N. P., Q. Wei, P. K. Shin, K. Konstantopoulos, and J. M. Ross.** 2006. *Staphylococcus aureus* adhesion via Spa, ClfA, and SdrCDE to immobilized platelets demonstrates shear-dependent behavior. Arterioscler Thromb Vasc Biol **26:**2394-400.
48. **Gomi, M., M. Sonoyama, and S. Mitaku.** 2004. High performance system for signal peptide prediction: SOSUIsignal. Chem-Bio Informatics Journal **4:**142-147.
49. **Grundy, F. J., D. A. Waters, T. Y. Takova, and T. M. Henkin.** 1993. Identification of genes involved in utilization of acetate and acetoin in Bacillus subtilis. Mol Microbiol **10:**259-71.
50. **Guex, N., and M. C. Peitsch.** 1997. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis **18:**2714-23.
51. **Haque, R. U., and J. N. Baldwin.** 1964. Types of Hemolysins Produced by Staphylococcus Aureus, as Determined by the Replica Plating Technique. J Bacteriol **88:**1442-7.
52. **Heilmann, C., J. Hartleib, M. S. Hussain, and G. Peters.** 2005. The multifunctional *Staphylococcus aureus* autolysin aaa mediates adherence to immobilized fibrinogen and fibronectin. Infect Immun **73:**4793-802.
53. **Heinrichs, J. H., M. G. Bayer, and A. L. Cheung.** 1996. Characterization of the sar locus and its interaction with *agr* in *Staphylococcus aureus*. J Bacteriol **178:**418-23.
54. **Hill, E. E., P. Herijgers, P. Claus, S. Vanderschueren, M. C. Herregods, and W. E. Peetermans.** 2006. Infective endocarditis: changing epidemiology and predictors of 6-month mortality: a prospective cohort study. Eur Heart J.

55. **Hiller, K., A. Grote, M. Scheer, R. Munch, and D. Jahn.** 2004. PrediSi: prediction of signal peptides and their cleavage positions. Nucleic Acids Res **32:**W375-9.
56. **Hiramatsu, T., K. Kodama, T. Kuroda, T. Mizushima, and T. Tsuchiya.** 1998. A putative multisubunit Na+/H+ antiporter from *Staphylococcus aureus*. J Bacteriol **180:**6642-8.
57. **Hofmann, K., and W. Stoffel.** 1993. TMBASE - A database of membrane spanning protein segments. Biol. Chem. Hoppe-Seyler **374:**166.
58. **Holden, M. T., E. J. Feil, J. A. Lindsay, S. J. Peacock, N. P. Day, M. C. Enright, T. J. Foster, C. E. Moore, L. Hurst, R. Atkin, A. Barron, N. Bason, S. D. Bentley, C. Chillingworth, T. Chillingworth, C. Churcher, L. Clark, C. Corton, A. Cronin, J. Doggett, L. Dowd, T. Feltwell, Z. Hance, B. Harris, H. Hauser, S. Holroyd, K. Jagels, K. D. James, N. Lennard, A. Line, R. Mayes, S. Moule, K. Mungall, D. Ormond, M. A. Quail, E. Rabbinowitsch, K. Rutherford, M. Sanders, S. Sharp, M. Simmonds, K. Stevens, S. Whitehead, B. G. Barrell, B. G. Spratt, and J. Parkhill.** 2004. Complete genomes of two clinical *Staphylococcus aureus* strains: evidence for the rapid evolution of virulence and drug resistance. Proc Natl Acad Sci U S A **101:**9786-91.
59. **Hulo, N., A. Bairoch, V. Bulliard, L. Cerutti, E. De Castro, P. S. Langendijk-Genevaux, M. Pagni, and C. J. Sigrist.** 2006. The PROSITE database. Nucleic Acids Res **34:**D227-30.
60. **Jambon, M., A. Imberty, G. Deleage, and C. Geourjon.** 2003. A new bioinformatic approach to detect common 3D sites in protein structures. Proteins **52:**137-45.
61. **Jensen, L. J., R. Gupta, H. H. Staerfeldt, and S. Brunak.** 2003. Prediction of human protein function according to Gene Ontology categories. Bioinformatics **19:**635-42.
62. **Jones, D. T., W. R. Taylor, and J. M. Thornton.** 1994. A model recognition approach to the prediction of all-helical membrane protein structure and topology. Biochemistry **33:**3038-49.
63. **Jonsson, P., M. Lindberg, I. Haraldsson, and T. Wadstrom.** 1985. Virulence of *Staphylococcus aureus* in a mouse mastitis model: studies of alpha hemolysin, coagulase, and protein A as possible virulence determinants with protoplast fusion and gene cloning. Infect Immun **49:**765-9.
64. **Josefsson, E., K. W. McCrea, D. Ni Eidhin, D. O'Connell, J. Cox, M. Hook, and T. J. Foster.** 1998. Three new members of the serine-aspartate repeat protein multigene family of *Staphylococcus aureus*. Microbiology **144 (Pt 12):**3387-95.
65. **Juretic, D., L. Zoranic, and D. Zucic.** 2002. Basic charge clusters and predictions of membrane protein topology. J Chem Inf Comput Sci **42:**620-32.
66. **Kall, L., A. Krogh, and E. L. Sonnhammer.** 2004. A combined transmembrane topology and signal peptide prediction method. J Mol Biol **338:**1027-36.

67. **Karlsson, A., and S. Arvidson.** 2002. Variation in extracellular protease production among clinical isolates of *Staphylococcus aureus* due to different levels of expression of the protease repressor *sarA*. Infect Immun **70:**4239-46.
68. **Kelley, L. A., R. M. MacCallum, and M. J. Sternberg.** 2000. Enhanced genome annotation using structural profiles in the program 3D-PSSM. J Mol Biol **299:**499-520.
69. **Kihara, D., T. Shimizu, and M. Kanehisa.** 1998. Prediction of membrane proteins based on classification of transmembrane segments. Protein Eng **11:**961-70.
70. **Korem, M., Y. Gov, M. D. Kiran, and N. Balaban.** 2005. Transcriptional profiling of target of RNAIII-activating protein, a master regulator of staphylococcal virulence. Infect Immun **73:**6220-8.
71. **Krogh, A., B. Larsson, G. von Heijne, and E. L. Sonnhammer.** 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol **305:**567-80.
72. **Kuroda, M., H. Kuroda, T. Oshima, F. Takeuchi, H. Mori, and K. Hiramatsu.** 2003. Two-component system VraSR positively modulates the regulation of cell-wall biosynthesis pathway in *Staphylococcus aureus*. Mol Microbiol **49:**807-21.
73. **Laskowski, R. A., J. D. Watson, and J. M. Thornton.** 2005. ProFunc: a server for predicting protein function from 3D structure. Nucleic Acids Res **33:**W89-93.
74. **Laurie, A. T., and R. M. Jackson.** 2005. Q-SiteFinder: an energy-based method for the prediction of protein-ligand binding sites. Bioinformatics **21:**1908-16.
75. **Letunic, I., R. R. Copley, S. Schmidt, F. D. Ciccarelli, T. Doerks, J. Schultz, C. P. Ponting, and P. Bork.** 2004. SMART 4.0: towards genomic data integration. Nucleic Acids Res **32:**D142-4.
76. **Liang, S., C. Zhang, S. Liu, and Y. Zhou.** 2006. Protein binding site prediction using an empirical scoring function. Nucleic Acids Res **34:**3698-707.
77. **Luong, T. T., P. M. Dunman, E. Murphy, S. J. Projan, and C. Y. Lee.** 2006. Transcription Profiling of the mgrA Regulon in *Staphylococcus aureus*. J Bacteriol **188:**1899-910.
78. **Maghnouj, A., T. F. de Sousa Cabral, V. Stalon, and C. Vander Wauven.** 1998. The arcABDC gene cluster, encoding the arginine deiminase pathway of Bacillus licheniformis, and its activation by the arginine repressor argR. J Bacteriol **180:**6468-75.
79. **Nagarajan, V., and M. O. Elasri.** 2007. SAMMD: *Staphylococcus aureus* microarray meta-database. BMC Genomics **8:**351.
80. **Nagarajan, V., and M. O. Elasri.** 2007. Structure and function predictions of the Msa protein in *Staphylococcus aureus*. BMC Bioinformatics **8 Suppl 7:**S5.

81. **Nakai, K., and P. Horton.** 1999. PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. Trends Biochem Sci **24:**34-6.
82. **Nilsson, I. M., O. Hartford, T. Foster, and A. Tarkowski.** 1999. Alpha-toxin and gamma-toxin jointly promote *Staphylococcus aureus* virulence in murine septic arthritis. Infect Immun **67:**1045-9.
83. **Novick, R. P.** 2003. Autoinduction and signal transduction in the regulation of staphylococcal virulence. Mol Microbiol **48:**1429-49.
84. **O'Brien, L. M., E. J. Walsh, R. C. Massey, S. J. Peacock, and T. J. Foster.** 2002. *Staphylococcus aureus* clumping factor B (ClfB) promotes adherence to human type I cytokeratin 10: implications for nasal colonization. Cell Microbiol **4:**759-70.
85. **O'Riordan, K., and J. C. Lee.** 2004. *Staphylococcus aureus* capsular polysaccharides. Clin Microbiol Rev **17:**218-34.
86. **Oshida, T., M. Sugai, H. Komatsuzawa, Y. M. Hong, H. Suginaka, and A. Tomasz.** 1995. A *Staphylococcus aureus* autolysin that has an N-acetylmuramoyl-L-alanine amidase domain and an endo-beta-N-acetylglucosaminidase domain: cloning, sequence analysis, and characterization. Proc Natl Acad Sci U S A **92:**285-9.
87. **Pandey., G., V. Kumar., and M. Steinbach.** 2006. Computational Approaches for Protein Function Prediction: A Survey TR 06-028. Department of Computer Science and Engineering, University of Minnesota.
88. **Park, P. W., T. J. Foster, E. Nishi, S. J. Duncan, M. Klagsbrun, and Y. Chen.** 2004. Activation of syndecan-1 ectodomain shedding by *Staphylococcus aureus* alpha-toxin and beta-toxin. J Biol Chem **279:**251-8.
89. **Pasquier, C., V. J. Promponas, and S. J. Hamodrakas.** 2001. PRED-CLASS: cascading neural networks for generalized protein classification and genome-wide applications. Proteins **44:**361-9.
90. **Pohlmann-Dietze, P., M. Ulrich, K. B. Kiser, G. Doring, J. C. Lee, J. M. Fournier, K. Botzenhart, and C. Wolz.** 2000. Adherence of *Staphylococcus aureus* to endothelial cells: influence of capsular polysaccharide, global regulator agr, and bacterial growth phase. Infect Immun **68:**4865-71.
91. **Renna, M. C., N. Najimudin, L. R. Winik, and S. A. Zahler.** 1993. Regulation of the Bacillus subtilis alsS, alsD, and alsR genes involved in post-exponential-phase production of acetoin. J Bacteriol **175:**3863-75.
92. **Resch, A., R. Rosenstein, C. Nerz, and F. Gotz.** 2005. Differential gene expression profiling of *Staphylococcus aureus* cultivated under biofilm and planktonic conditions. Appl Environ Microbiol **71:**2663-76.
93. **Rost, B.** 2001. Review: protein secondary structure prediction continues to rise. J Struct Biol **134:**204-18.
94. **Rost, B., R. Casadio, P. Fariselli, and C. Sander.** 1995. Transmembrane helices predicted at 95% accuracy. Protein Sci **4:**521-33.

95. **Sali, A., and T. L. Blundell.** 1993. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol **234:**779-815.
96. **Sambanthamoorthy, K., M. S. Smeltzer, and M. O. Elasri.** 2006. Identification and characterization of *msa* (SA1233), a gene involved in expression of SarA and several virulence factors in *Staphylococcus aureus*. Microbiology **152:**2559-72.
97. **Schaffer, A. C., R. M. Solinga, J. Cocchiaro, M. Portoles, K. B. Kiser, A. Risley, S. M. Randall, V. Valtulina, P. Speziale, E. Walsh, T. Foster, and J. C. Lee.** 2006. Immunization with *Staphylococcus aureus* clumping factor B, a major determinant in nasal carriage, reduces nasal colonization in a murine model. Infect Immun **74:**2145-53.
98. **Schlag, S., C. Nerz, T. A. Birkenstock, F. Altenberend, and F. Gotz.** 2007. Inhibition of staphylococcal biofilm formation by nitrite. J Bacteriol **189:**7911-9.
99. **Schwede, T., J. Kopp, N. Guex, and M. C. Peitsch.** 2003. SWISS-MODEL: An automated protein homology-modeling server. Nucleic Acids Res **31:**3381-5.
100. **Solovyev, V. V., and N. A. Kolchanov.** 1994. Search for functional sites using consensus, p. 16-21. *In* N. A. Kolchanov and H. A. Lim (ed.), Computer analysis of Genetic macromolecules. World Scientific.
101. **Toledo-Arana, A., N. Merino, M. Vergara-Irigaray, M. Debarbouille, J. R. Penades, and I. Lasa.** 2005. *Staphylococcus aureus* develops an alternative, ica-independent biofilm in the absence of the arlRS two-component system. J Bacteriol **187:**5318-29.
102. **Tseng, C. W., and G. C. Stewart.** 2005. Rot repression of enterotoxin B expression in *Staphylococcus aureus*. J Bacteriol **187:**5301-9.
103. **Tung, H., B. Guss, U. Hellman, L. Persson, K. Rubin, and C. Ryden.** 2000. A bone sialoprotein-binding protein from *Staphylococcus aureus*: a member of the staphylococcal Sdr family. Biochem J **345 Pt 3:**611-9.
104. **Tusnady, G. E., and I. Simon.** 2001. The HMMTOP transmembrane topology prediction server. Bioinformatics **17:**849-50.
105. **von Heijne, G.** 1992. Membrane protein structure prediction. Hydrophobicity analysis and the positive-inside rule. J Mol Biol **225:**487-94.
106. **von Heijne, G.** 1986. A new method for predicting signal sequence cleavage sites. Nucleic Acids Res **14:**4683-90.
107. **Vriend, G.** 1990. WHAT IF: a molecular modeling and drug design program. J Mol Graph **8:**52-6, 29.
108. **Watson, J. D., and E. J. Milner-White.** 2002. A novel main-chain anion-binding site in proteins: the nest. A particular combination of phi,psi values in successive residues gives rise to anion-binding sites that occur commonly and are found often at functionally important regions. J Mol Biol **315:**171-82.
109. **Weinrick, B., P. M. Dunman, F. McAleese, E. Murphy, S. J. Projan, Y. Fang, and R. P. Novick.** 2004. Effect of mild acid on gene expression in *Staphylococcus aureus*. J Bacteriol **186:**8407-23.

110. **Yang, S. J., P. M. Dunman, S. J. Projan, and K. W. Bayles.** 2006. Characterization of the *Staphylococcus aureus* CidR regulon: elucidation of a novel role for acetoin metabolism in cell death and lysis. Mol Microbiol **60:**458-68.

111. **Zuniga, M., C. Miralles Md Mdel, and G. Perez-Martinez.** 2002. The Product of arcR, the sixth gene of the arc operon of Lactobacillus sakei, is essential for expression of the arginine deiminase pathway. Appl Environ Microbiol **68:**6051-8.