Posted May 22, 2007. The text of this research plan is copyrighted. See more samples here: http://www.niaid.nih.gov/ncn/grant/app/default.htm

This sample was created using the PHS 398. It does not conform to all the requirements for electronic application, including page limits and other items. We will be posting samples of electronic applications soon.

RESEARCH DESIGN AND METHODS

This section is organized by aims as largest level of organization – then subheads: rationale, experimental design, methods for each research question. Aims are designed to answer the question posed by hypothesis.

Specific Aim #1: Defining the relationship between sar transcription and the production of SarA.

Rationale. Northern blot analysis of the *sar* locus reveals the presence of three transcripts (sarA, sarB and sarC), all of which share a common termination site and include the entire *sarA* structural gene (2). Each transcript is produced from its own promoter in a growth-phase dependent manner. Specifically, expression of the sarA and sarB transcripts is highest during exponential growth while expression of sarC is highest during the post-exponential growth phase (2). In our experience with clinical isolates of *S. aureus*, the growth-phase dependency of sarB expression is much tighter than the growth-phase dependency of sarA expression (i.e. sarB is not made during post-exponential growth while sarA is made is reduced amounts by comparison to exponential phase levels). **The overall objective of these experiments is to correlate the temporal production of the sarA, sarB and sarC transcripts with the production and activity of SarA.** The experiments include assays for the direct detection of SarA (Part A) as well as assays capable of assessing SarA function (Parts B and C). The direct detection of SarA will be done by Western blot using the affinity-purified anti-SarA antibody preparation discussed in Preliminary Results. We will correlate the results of these experiments with functional assays assessing the activity of SarA as a DNA-binding protein (Part B) and the activity of SarA as a transcriptional activator (Part C). Establishing the correlation between the accumulation of SarA and the functional activity of SarA will allow us to address the possibility that the small ORFs contained within the sarB and sarC transcripts serve a functional role with respect to SarA activity. The activity of SarA as a DNA-binding protein will be assessed by EMSA experiments performed with whole cell lysates and DNA fragments that contain the heptad repeats upstream of the *agr* P_2 and P_3 promoters. The function of SarA as a transcriptional activator will be assessed using fusions between each of the *agr* promoters and a *xylE* reporter gene. Because these experiments will employ fusions with the *agr* promoters, the experiments will specifically address the function of SarA as a transcriptional activator. Similar experiments utilizing fusions between the *xylE* reporter and the *cna* promoter region (P_{cna}) will allow us to correlate the results with the functional activity of SarA as a transcriptional repressor (Specific Aim #2).

Key points are bold for emphasis. This lets non-reviewers scan and retrieve information quickly during the review meeting.

Gives detailed experimental methods for primary reviewers. Organization allows other reviewers to bypass these sections.

Experimental design. The Western blot and EMSA experiments (Parts A and B respectively) will be done using cell lysates from *S. aureus* strains RN6390, DB, Sar R, 11D2 and RN6911. RN6390 and DB are wild-type strains that encode and express all three *sar* transcripts (7,11). The inclusion of both strains is based on the observation that, by comparison to DB, the *sarA* gene in RN6390 has a nonsense mutation that results in the deletion of 11 amino acids from the C-terminus of the SarA protein (2). Although Bayer et al. (2) suggest that this truncation has no effect on the production or activity of SarA, the phenotype of RN6390 *sar* mutants (e.g. Sar R) is somewhat distinct by comparison to a DB *sar* mutant (e.g. 11D2) (9,11). Based on that, we believe the inclusion of both strains will allow us to make a more a comprehensive assessment of the functional activity of SarA in *S. aureus*. It is anticipated that SarA will be detected at some point in the growth cycle of RN6390 and DB, with one of the experimental questions being whether the temporal accumulation and/or activity of SarA in RN6390 differs from the pattern observed in DB. Sar R and 11D2 are *sar* mutants generated in RN6390 and DB respectively. These strains carry transposon insertions in the *sarA* structural gene and are included as negative controls for their respective parent strains. The inclusion of RN6911 in these experiments requires comment. RN6911 is an RN6390-derived *agr*-null mutant in which the entire *agr* locus (including the region encoding RNAIII) has been replaced with a tetracycline-resistance determinant (
in *tetM*) (38). Its inclusion is based on the hypothesis that RNAIII may contribute to the production of SarA. That hypothesis requires explanation since it has been demonstrated both in our laboratory and by other investigators that mutation of *agr* does not affect the production of any *sar* transcript (7). However, we believe the inclusion of RN6911 is warranted because 1) the impact of *agr* on *sar* has not been examined at any level other than Northern blot analysis of *sar* transcription in *agr* mutants (6,23) and 2) the *agr*-encoded RNAIII molecule is known to affect the production of at least one *S. aureus* protein (alpha-toxin) at the post-transcriptional level (26). Importantly, the post-transcriptional effect of RNAIII on alpha-toxin production involves an "anti-attenuation" in which translation of the *hla* mRNA is enhanced in the presence of RNAIII. **In the absence of RNAIII**, the *hla* mRNA transcript forms a stem-loop structure that sequesters the ribosome-binding site within the duplexed stem region (26). The unavailability of the ribosome-binding site results in the translational attenuation of alpha-toxin production. **In the presence of RNAIII**, this attenuation is relieved due to the formation of an RNAIII:*hla* mRNA duplex molecule in which the ribosome-binding site on the *hla* mRNA is exposed and available for translation (26). The observation that this "anti-attenuation" involves a short (~80 bp) region of the RNAIII molecule that exhibits only partial complementarity (~75%) with *hla* mRNA suggests that a similar effect may occur with the transcripts encoding other proteins including SarA. It should be re-emphasized that the Northern blot analysis of *sar* transcription in *agr* mutants does not address the possibility that RNAIII has an effect on the production of SarA from an existing *sar* transcript. Based on that, the second experimental question to be addressed by these studies is whether RN6911 produces SarA and, if so, whether the temporal pattern of SarA production is affected by the absence of RNAIII.

Primary reviewers expect citations to the latest relevant publications.

Since they developed and had access to reagents in preliminary studies, they don't need to state how they will obtain them here.

A. Determination of the temporal accumulation of SarA by Western blot.

> Experiments are grouped by experimental question. Application includes extensive details on how they will proceed.

Methods. To quantitate the accumulation of SarA, whole cell lysates will be prepared from each strain using cells taken at various stages of *in vitro* growth. Specifically, each strain will be grown in tryptic soy broth at 37°C with constant aeration. Two-liter cultures will be used to allow for the removal of relatively large volumes (~500 ml) at each time point. It is anticipated that relatively large volumes may be required to compensate for the relatively low cell density during the exponential growth phase. Cells will be harvested after 2, 4, 6 and 8 hours of growth. Under these growth conditions, the transition between exponential and post-exponential growth can be anticipated to occur between the 4 and 6 hr time points. Additionally, based on our Northern blot analysis (data not shown), these time points encompass each of the growth phases associated with maximal production of the sarA, sarB and sarC transcripts. In this as well as all other time course experiments, our assumptions about growth phase will be verified by monitoring the optical density of each culture at 560 nm. Because the mutations in Sar R, 11D2 and RN6911 involve chromosomal insertions, all cultures will be grown without antibiotic selection. However, to ensure the stability of each insertion, duplicate plate counts utilizing medium with and without antibiotics will be done at the completion of each experiment.

Lysates will be prepared according to the method of Mahood and Kahn (25). Briefly, cells will be harvested by centrifugation, washed and resuspended in 4.0 ml TEG buffer (25 mM Tris-Cl (pH 8.0), 5 mM EGTA). The cell suspension will be subjected to two freeze-thaw cycles prior to adding KCl to a final concentration of 0.15 M and lysostaphin to a final concentration of 0.2 mg/ml. After incubating on ice for 45 min., the cell suspension will be subjected to two additional freeze-thaw cycles and then centrifuged at 30,000 rpm in a Beckman SW40 rotor. After harvesting the supernatant, glycerol will be added to a final concentration of 20% (v/v). The supernatant will then be dialyzed against a buffer consisting of 10 mM Tris-Cl (pH 7.5), 1 mM EDTA, 1 mM dithiothreitol, 50 mM $NaCl_2$ and 20% (v/v) glycerol. After determining the protein concentration using the Bradford protein assay (Bio-Rad Protein Assay, Bio-Rad Laboratories, Hercules, CA), aliquots of each lysate will be resolved by SDS-PAGE (10-20% gradient gels) and the resolved proteins transferred to nitrocellulose membranes. Western blot analysis will be performed using the Phototope-HRP Western Blot Detection Kit (New England Biolabs, Beverly, MA). Briefly, the nitrocellulose membrane will be sequentially incubated in 1) blocking buffer consisting of Tris-buffered saline containing 0.1% Tween-20 and 5% gelatin, 2) an appropriate dilution of the affinity-purified anti-SarA antibody preparation (see below) and 3) goat anti-rabbit IgG conjugated to horseradish peroxidase (HRP). Detection of the HRP-conjugated antibody will be done using the LumiGlo chemiluminescent substrate and exposure to X-ray film. The use of a radiographic images will allow us to quantitate signal strengths by densitometric analysis.

> Alternatives are provided to accommodate different results.

To determine the appropriate dilution of the anti-SarA antibody preparation, Western blot analysis will be performed as described above using purified SarA. In the first set of experiments, a constant amount of SarA (~500 ng) will be resolved in multiple lanes of the same gel and then blotted using varying dilutions of the antibody. In the second set of experiments, the dilution of antibody that results in a reproducible, strong signal in the first experiment will be tested against decreasing amounts of purified SarA. If the sensitivity of the assay is sufficient (estimated at 10 ng of purified SarA), subsequent Western blots will be done using the same dilution of antibody. If the detection limit is >10 ng, then we will decrease the antibody dilution to bring the concentration of SarA antibody up to

level capable of detecting lesser amounts of protein. Alternatively, we can scale up the procedure for production of cell lysates to bring the final yield of SarA into the detectable range. In either case, the sensitivity of the Western blot protocol will be confirmed by repetitive blots using IgG purified from pre-immune serum obtained from the same rabbit used for the SarA immunization. The IgG fraction will be purified by protein G-sepharose affinity chromatography. The dilution of the SarA antibody to be used in subsequent experiments will be the concentration of antibody that yields a reproducible signal with the least amount of purified SarA without a background signal with the same dilution of pre-immune serum. Once the appropriate dilution is determined, a standard curve will be constructed based on the densitometric analysis of Western blot radiographs. The amount of SarA in each cell lysate will be determined by comparison to this standard curve.

The amount of SarA in lysates prepared from each strain at different time points will be correlated with the production of individual *sar* transcripts in the same strains at the same time points. The *sar* transcripts will be detected by Northern blot using a *sarA*-specific gene probe (because all three transcripts include the entire *sarA* gene, all three can be detected with equal efficiency using the same gene probe). Protocols for the isolation of RNA, resolution of RNA by agarose gel electrophoresis, and the detection of mRNA's using digoxygenin-labeled DNA probes and the Lumigen chemiluminescent substrate (Boehringer Mannheim, Indianapolis, IN) are all well-established in the PI's laboratory (20, 22).

Well-established methods are referenced, not described.

B. Determination of the temporal accumulation of SarA by EMSA.

Methods. The DNA-binding activity of SarA will be assessed using the DB, 11D2, RN6390, Sar R and RN6911 lysates produced for our Western blot experiments. The EMSA experiments will be done using the 45-bp DNA fragment corresponding to the *agr* P_3 promoter region (see Preliminary Results) and a similar fragment corresponding to the *agr* P_2 promoter region. The inclusion of both DNA targets is based on the results of Heinrichs et al. (23), who suggested that SarA exhibits differential binding to the *agr* P_2 and P_3 promoters. The concentration of DNA to be used in each set of experiments will be determined based on equilibrium dissociation constants (K_D) determined in EMSA experiments using varying concentrations of purified SarA and each of the DNA targets (13, 14). Specifically, the amount of DNA used in EMSA experiments with whole cell lysates will be 100-fold higher than the K_D determined using purified SarA. Under those conditions, stoichiometric binding occurs such that the concentration of protein required to bind 50% of the available DNA ($P_{1/2}$) is equal to 50% of the input DNA concentration [DNA_o] (5). For example if the K_D is 200 pM, then we will use 20 nM of input DNA. Based on the conversion formula $P_{1/2} = 1/2\ [DNA_o]$, the point at which 50% of the DNA (10 nM) is bound will indicate that the extract contains 10 nM of SarA capable of binding an appropriate DNA target.

Once the appropriate concentration of each DNA target is determined, EMSA experiments will be done using a constant amount of ^{32}P-labeled DNA and various dilutions of the whole cell lysate prepared from each strain at each time point. Bound and unbound DNAs will be resolved by native gel electrophoresis and the unbound fraction quantitated by laser densitometry. The bound fraction, determined by subtracting the unbound fraction from the total input DNA, will be plotted against the volume of crude extract used in the mixture. The specificity of the protein-DNA complex observed by EMSA will be confirmed by supershift with

Indicates sequence of events. Shows how results will be interpreted and built upon.

anti-SarA antibody and by competition experiments employing unlabeled DNAs. The amount of SarA capable of binding DNA will be expressed in moles of SarA per volume of extract and will be compared with the accumulation of each of the *sar* transcripts and the accumulation of SarA as defined by the Western blot protocol described in Specific Aim #1, Part A. To determine whether the DNA-binding activity of SarA is a direct function of the concentration of SarA in the cell, the results will be reported as the ratio of the total amount of SarA as defined by Western blot (Specific Aim #1, Part A) versus the amount of SarA bound to each DNA target (e.g. [total SarA/bound SarA]). Because an increase in the DNA-binding activity of SarA will be reflected by a decrease in this ratio, any decrease will be taken as an indication that the DNA-binding activity of SarA can be affected by post-translational modifications. Correlating such changes with the production of specific *sar* transcripts will allow us to assess the possibility that the short ORFs encoded within the sarB and sarC transcripts are important for SarA function. Finally, by correlating the results obtained with a DNA fragment corresponding to the *agr* P_2 promoter region with the results obtained with a DNA fragment corresponding to the *agr* P_3 promoter region, any functional difference in the ability to bind the two *agr* promoters can be assessed.

C. Determination of the temporal accumulation of SarA by transcriptional activation.

Describes additional experiments to confirm results.

Methods. To confirm the results of our experiments correlating the production of each of the *sar* transcripts with the production of SarA and the activity of SarA as a DNA-binding protein, we will also assess the growth-phase dependent activity of SarA as a transcriptional activator. These experiments will also allow us to determine whether transcriptional activation from the *agr* P_2 and P_3 promoters occurs at the same time and under the same conditions. These experiments will be done using deriviatives of RN6390, Sar R, DB, 11D2 and RN6911 carrying either a P_2-*xylE* reporter fusion or a P_3-*xylE* reporter fusion. The reporter vector was constructed by inserting a promoter-less *xylE* gene downstream from the multiple cloning site in the pLI50 *E. coli-S. aureus* shuttle vector (see Preliminary Results). To generate the appropriate fusions, we used PCR to amplify the entire 186-bp region between the RNAII and RNAIII transcription start sites. This amplification was done with primers that contain engineered restriction sites incorporated for use in cloning into the pLI50-*xylE* vector. Specifically, *Hind*III restriction sites were added at both the 5' and 3' ends of the amplified fragment. The presence of these restriction sites will allow us to clone the 186-bp fragment upstream of the *xylE* reporter gene in both orientations. Because the heptad repeats upstream of the P_2 and P_3 promoters are directional, this approach should allow us to assess activation from each promoter while at the same time allow for any influence of the distal binding sites on activity from the proximal promoter. To select for the *xylE* reporter plasmid, 5 µg/ml chloramphenicol will be added to the growth medium. Variants of each strain that contain the promoter-less *xylE* vector will be included as negative controls.

Catechol 2,3-dioxygenase (CATase) activity will be assessed using cellular lysates prepared from samples taken at the same time points described for Specific Aim #1, Parts A and B. Our ability to correlate CATase activity with the SarA-mediated activation of transcription from the P_2 and P_3 promoters is based on the observations that (i) mutation of *sar* in RN6390 results in reduced production of both RNAII and RNAIII and (ii) the production of RNAII and RNAIII is restored when a clone carrying the *sarA* structural gene is introduced into the RN6390 *sar* mutant (6). These results clearly indicate that transcription from the P_2 and P_3 promoters in RN6390 is at least partially dependent on SarA. CATase assays will be done essentially as described by Ray et al. (34). Briefly, 10 ml samples will be harvested from each culture at 2 hr intervals and the cells harvested by centrifugation. Cell pellets will be washed with 2 ml of 20

mM potassium phosphate buffer (pH 7.2) and then frozen overnight at -70°C. Pellets will then be thawed, resuspended in 2 ml of APEL buffer (100 mM potassium phosphate (pH 7.5), 20 mM EDTA, 10% acetone, 200 µg lysostaphin per ml), and incubated for 30 min. at 37°C. After adding 20 µl Triton X-100 and incubating on ice for an additional 30 min., extracts will be centrifuged to pellet the cellular debris. CATase assays will be done by mixing 250 µl of each cell extract with 2.75 ml of 100 mM potassium phosphate buffer (pH 7.5) containing 0.2 mM catechol. All reactions will be carried out at 30°C (34, 44). Results will be monitored spectrophotometrically by measuring the absorbance at 375 nm (17). Under these conditions, 1 milliunit of catechol 2,3'-dioxygenase activity corresponds to the formation of 1 nmol of 2-hydroxymuconic semialdehyde per minute (44).

To ensure that the results reflect the functional activity of SarA, the CATase activity observed when the *xylE* fusions to the *agr* P_2 and P_3 promoters are introduced into the Sar R and 11D2 mutants will be taken as the background CATase activity associated with each of the parent strains. The background activity observed with each mutant will be subtracted from the results obtained with the corresponding parent strain. The results will then be compared with the values obtained for experiments assessing the accumulation of SarA as defined by Western blot analysis (Specific Aim #1, Part A) and the DNA-binding activity of the accumulated SarA as defined by EMSA (Specific Aim #1, Part B). If the activity of SarA as a transcriptional activator is dependent only on the production of SarA, then changes in CATase activity as a function of growth phase should be consistent with changes in the amount of SarA. An increase in CATase activity without a corresponding increase in the concentration of SarA will be taken as an indication that SarA exists in alternative forms that differ in their ability to activate transcription. Comparison of the results obtained with the *agr* P_2 and P_3 promoters will allow us to assess whether such differences are promoter-dependent. Correlation of the results with the production of different *sar* transcripts will provide an indication of whether the small *sar* ORFs are required for SarA activity.

Specific Aim #2: Characterization of the mechanism by which sar regulates expression of the S. aureus collagen adhesin gene (cna).

Rationale. Expression of the *S. aureus* collagen adhesin gene (*cna*) is highest during the exponential growth phase and falls to almost undetectable levels during post-exponential growth (20).To assess the regulatory impact of *sar* and *agr* on *cna* transcription, we introduced *cna* into RN6390, Sar R and RN6911. Comparison of *cna*-positive derivatives of each strain indicates that *sar* is the primary regulatory element controlling expression of *cna* and that the regulatory impact of *sar* is independent of the interaction between SarA and *agr* (20,21). Subsequent experiments indicate that the *sar*-mediated regulation of *cna* transcription involves a direct interaction between SarA and *cis* elements upstream of *cna* (see Preliminary Results, Fig. 12). **Importantly, these results represent the first indication that SarA binds any DNA target other than the heptad repeats upstream of the *agr* P_2 and P_3 promoters.** Moreover, the SarA-mediated regulation of *cna* transcription differs from the SarA-mediated regulation of *agr* transcription in two respects. First, mutation of *sar* results in the increased expression of *cna* (21) and the decreased expression of RNAII and RNAIII (23). Second, the DNA upstream of *cna* does not contain heptad repeats like those associated with the *agr* promoters (19, 30). Based on these differences, we believe it is important that the mechanism of the *sar*-mediated regulation of *cna* transcription be addressed in detail. That is the overall objective of Specific Aim #2.

Experimental design. The experiments in this section are divided into three areas. The first is directed toward identification of the specific *sar* transcripts required to restore the *sar*-mediated

regulation of *cna* transcription. These experiments will be done by introducing plasmids encoding each *sar* transcript into a *cna*-positive *sar* mutant and assessing the effect on *cna* transcription. The results will be correlated with the production of each *sar* transcript from each clone (it should be emphasized that the *sar* transcripts overlap such that the sarB transcript is all inclusive while the sarC transcript includes the promoter for the sarA transcript; see Preliminary Results, Fig. 13). Additionally, once the SarA-binding site upstream of *cna* has been defined (see below), EMSA experiments will be done using cell lysates from *sar* mutants complemented with each *sar* transcript and the appropriate DNA target. These experiments will allow us to correlate the results of our complementation analysis with the production and activity of SarA. The second set of experiments can be divided into three phases, all of which are directed toward identification and characterization of the SarA-binding site upstream of *cna*. Specifically, we will 1) localize the SarA binding site using EMSA experiments done with purified SarA and DNA fragments derived from the region upstream of *cna*, 2) identify and partially characterize the binding site within the target fragment by DNA footprinting and methylation interference assays and 3) characterize the sequence-specificity of the binding site using EMSA experiments done with mutagenized DNA targets and purified SarA. Importantly, all of the protocols required to complete the Specific Aim #2, Part B experiments are well-established in the Co-I's laboratory (14,15). The final set of experiments in this section is directed toward assessing the *in vivo* significance of SarA binding. These experiments will be done using transcriptional fusions between sequence variants of the *cis* elements upstream of *cna* and a promoter-less *xylE* reporter gene.

A. Complementation of the defect in *cna* transcription observed in *sar* mutants.

Methods. The experimental approach taken in these experiments is based on studies demonstrating that complementation of different defects in *sar* mutants requires the introduction of DNA fragments encoding different *sar* transcripts (23). As a first step toward defining the mechanism by which *sar* regulates expression of *cna*, we will attempt to complement the defect in *cna* transcription using plasmid-borne DNA fragments encoding each *sar* transcript (see Preliminary Results, Fig. 13). Each of these fragments has already been cloned into the pLI50 *E. coli*-*S. aureus* shuttle vector. Each clone will be introduced into the *cna*-positive, *sar* mutant UAMS-175. Transformation will be done by electroporation as previously described (20). Transformants will be selected by plating on tryptic soy agar containing 5 µg per ml chloramphenicol. After verifying the presence of the appropriate plasmid, transformants will be grown in tryptic soy broth and sampled at 2 hr intervals. Transformants will be characterized with respect to (i) restoration of the expected *sar* transcripts, (ii) the temporal pattern of *sar* transcription, (iii) the capacity to bind collagen and (iv) the temporal pattern of *cna* transcription. Northern blot analysis will be done as previously described (22) using probes corresponding to *sarA* and *cna*. As discussed above, the fact that all three *sar* transcripts include the *sarA* gene will allow us to detect all three transcripts using the same *sarA* probe. The *cna* gene probe was generated by PCR using genomic DNA from *S. aureus* strain UAMS-1 as template (20). Collagen-binding assays will be done using ^{125}I-labeled type I collagen as previously described (20). Restoration of the wild-type phenotype with respect to each of the parameters noted above will be made by comparison to the *cna*-positive, *sar*-positive strain UAMS-174 (see Preliminary Results, Table 1). As the experiments aimed at defining the SarA-binding site upstream of *cna* progress (see below), we will correlate the results of these experiments with the accumulation of SarA as determined by Western blot (see Specific Aim #1, Part A) and by EMSA experiments performed with the appropriate DNA fragment and purified SarA.

Gives rationale for approach.

B. Characterization of the SarA DNA-binding site upstream of *cna*.

Methods #1: Localization of the SarA-binding site upstream of cna. EMSA experiments employing a series of short, overlapping fragments derived from the region upstream of *cna* and purified SarA indicate that at least one SarA-binding site exists within approximately 200 bp upstream of the *cna* start codon (see Preliminary Results, Fig. 12). To further localize the SarA-binding site(s) upstream of *cna*, we will synthesize DNA fragments corresponding to progressively smaller regions of the DNA upstream of *cna* and use these fragments in EMSA experiments utilizing purified SarA. Appropriate fragments will be generated either by synthesizing and annealing complementary pairs of oligonucleotides or by PCR. The fact that we have already sequenced the region extending 930 bp upstream of the *cna* transcriptional start site (19) will greatly facilitate the synthesis of the appropriate DNA targets. EMSA experiments will be done using standard procedures (5). Briefly, ^{32}P-labeled DNA fragments will be mixed with purified SarA diluted to concentrations ranging from 1 to 100 nM (this concentration range was chosen based on preliminary experiments demonstrating a band shift with 55 nM SarA). After an appropriate incubation period, SarA-DNA complexes will be resolved by native gel electrophoresis as illustrated in Preliminary Results. It is anticipated that these experiments will allow us to localize the SarA-binding site to a region spanning 50-100 bp. However, we recognize that SarA-binding sites may exist across a relatively long stretch of DNA with intervening and perhaps irrelevant nucleotides in between each site. To address that issue, all fragments that are bound by SarA will be used in “mix and match” competition experiments aimed at defining the relative affinity of SarA for different binding sites. The experiments described below will be done starting with the DNA fragments that exhibit the highest affinity for SarA (i.e. that exhibit a gel shift with the lowest concentration of SarA and are not competitively inhibited in the presence of other, unlabeled DNA fragments).

Methods #2: Mapping of the SarA-binding site upstream of cna. Once the smallest DNA fragment(s) bound by SarA have been identified, the same fragment(s) will be characterized by DNase I footprinting and methylation interference assays. Both sets of experiments will be done with purified SarA and ^{32}P-labeled DNA fragments. **For DNaseI footprinting**, SarA will be diluted to an appropriate concentration (based on the results of the EMSA experiments described above) in an assay buffer consisting of 10 mM Tris-Cl (pH 8.0), 5 mM MgCl2, 1 mM CaCl2, 100 mM KCl, 2 mM dithiothreitol (DTT), 50 µg/ml bovine serum albumin (BSA) and 2 µg/ml calf thymus DNA. After allowing the mixture to equilibrate, DNase I diluted in assay buffer without BSA or calf thymus DNA will be added and the exposure allowed to proceed for 2 minutes. The appropriate concentration of DNase I will be determined empirically (4). A control tube without DNase I will also be included in all experiments. Reaction products will be examined by polyacrylamide gel electrophoresis and exposure to X-ray film according to standard protocols (4). **For methylation protection experiments**, labeled DNA fragments will be incubated for 5 minutes at room temperature with dimethyl sulfate (DMS) diluted in an assay buffer consisting of 50 mM sodium cacodylate (pH 8.0) and 50 mM EDTA. The reaction will be stopped by adding 1.5 M sodium acetate containing 1 M 2-mercaptoethanol. After ethanol precipitation (1), the methylated DNA will be mixed with purified SarA and the mixture allowed to equilibrate for 15-30 minutes at room temperature. The SarA-DNA complexes will then be resolved using native polyacrylamide gels. The bands corresponding to SarA-DNA complexes and to free DNA will be purified and subjected to piperidine cleavage. The reaction products will then be resolved using polyacrylamide sequencing gels. After exposure to X-ray film, protected bases will be detected by comparison to the free DNA lane; fragments corresponding to protected bases will be absent in the lane derived from the SarA-DNA complexes (1). It is anticipated that the DNase I footprinting will allow us to define the region containing the SarA-binding site and that the methylation interference assays will provide an indication of the relative

significance of specific guanine and adenine residues within the protected region.

Methods #3: Defining the sequence characteristics of the SarA-binding site upstream of cna. The final set of experiments in this section (Specific Aim #2, Part B) is directed toward confirmation of the SarA-binding site(s) and identification of the specific sequence requirements necessary for SarA binding. Specifically, once we have identified the protected region in our footprinting experiments, we will generate sequence variants by synthesizing complementary pairs of oligonucleotides that span the protected region. The same approach was taken with the 45-bp fragment synthesized for the preliminary EMSA experiments done to confirm the identity of our purified SarA (Fig. 9). However, in this case, we will design the oligonucleotides to introduce specific nucleotide substitutions within the protected region. If the protected region does, in fact, represent a specific SarA target, then it should be possible to generate variants that are not bound by SarA. Although it is not possible to anticipate the nature or extent of all substitutions, the methylation interference assays should provide some information with regard to critical adenine and/or guanine residues. Moreover, it should be emphasized that the Molecular Biology Core Facility at the PI's institution is located in the Co-I's laboratory and includes an oligonucleotide synthesizer that can be used to generate as many sequence variants as necessary. The results will be evaluated by comparing the sequence of DNA fragments that are bound by SarA with the sequence of DNA fragments that are not bound by SarA. To more accurately assess the relative affinity of SarA for different DNA targets, all experiments will be done using SarA at various concentrations. The specificity of binding will subsequently be confirmed by repeating the EMSA experiments using combinations of labeled and unlabeled target DNAs. Taking this approach, it should be possible to identify those base pairs that are necessary for SarA binding (i.e. those base pairs that cannot be changed without eliminating the SarA DNA-binding target).

C. Correlation of SarA binding with regulation of *cna* transcription.

Methods. The functional significance of the putative SarA-binding site(s) will be assessed by *in vivo* experiments in which different versions of the *cna* target(s) are cloned into the *xylE* reporter gene vector and introduced into *S. aureus*. The activity of different mutagenized versions of the binding site(s) will be assessed by performing CATase assays using cells taken at various stages of growth and comparing the results to CATase assays done using the *xylE* reporter construct containing the wild-type binding site. To reduce interference from an endogenous binding site, these experiments will be done in a *cna*-negative strain of *S. aureus* (e.g. RN6390). The specific DNA fragments to be introduced into the *xylE* reporter vector will be determined based on the results of the EMSA and footprinting experiments decribed in Part B. Most of the other methods required for these experiments have already been described. Specifically, the *xylE* reporter vector, the methods to be employed for the synthesis of different versions of the SarA target upstream of *cna*, and the CATase assays to be used to assess expression from different versions of the SarA-binding site are all described in other sections of this proposal. However, it should be emphasized that we do not intend these experiments to be as all inclusive as those described in Section B. More specifically, the objective of these experiments is to confirm that the putative SarA binding site identified in Part B is functional in terms of serving as a SarA target in *S. aureus*. Based on that, we will only examine a restricted set of fragments using the *xylE* reporter vector. Specifically, we will identify a subset of fragments that retain the ability to bind SarA and a subset of fragments that have lost the ability to bind SarA and will then clone these fragments into the promoter-less *xylE* vector. To introduce the target region into the *xylE* vector, appropriate restriction site linkers will be added to each fragment using standard procedures. Plasmids carrying different variants of the SarA target(s) upstream of *cna* will be introduced into RN6390 by electroporation. The results will be

evaluated by comparison to the results obtained with a *xylE* fusion containing the wild-type promoter region and will be reported as the ratio of CATase activity observed with each variant versus the activity observed with the wild-type sequence (CATase activity of variant/CATase activity of wild-type). Because mutation of *sar* results in overexpression of *cna*, elimination of the SarA binding site should result in increased *xylE* expression. For that reason, any increase in this ratio will be taken as an indication that the putative SarA-binding site is functional *in vivo* in *S. aureus*.

Specific Aim 3: Identification of S. aureus virulence factors under the direct control of SarA.

Rationale. We believe the preceding discussion establishes the fact that *sar* modulates expression of *S. aureus* genes through both *agr*-dependent and *agr*-independent regulatory pathways. In most cases, the existence of the *agr*-independent pathway can only be inferred from the comparison of *sar* and *agr* mutants. On the other hand, our results with regard to the regulation of *cna* transcription establish the existence of additional SarA targets within the *S. aureus* genome and suggest that the regulation of other genes that are differentially regulated by *sar* and *agr* may also involve a direct interaction between SarA and *cis* elements upstream of the target genes. Additionally, the characterization of *sar* mutants has been limited to a relatively small set of genes defined by the availability of appropriate gene probes and/or quantitative phenotypic assays. For that reason, our understanding of the *agr*-independent pathway is incomplete both in terms of the mechanism by which *sar* regulates expression of different target genes and the scope of SarA targets. However, the increasing size of the *S. aureus* genomic database and the availability of computer programs capable of the detailed analysis of megabase quantities of DNA have set the stage for detailed studies aimed at the direct identification of genes under the direct regulatory control of SarA. Our successful purification of SarA in a form capable of binding diverse DNA targets (e.g. *cis* elements upstream of the *agr* and *cna* promoters), together with the established expertise of the Co-I in the use of PCR-assisted binding site selection (PCR-ABS) (15), places us in a unique position with regard to the identification of additional SarA targets within the *S. aureus* genome. Finally, the molecular expertise of the PI in the molecular biology of *S. aureus* and the successful production of appropriate *sar* and *agr* mutants will allow us to confirm the results of our PCR-ABS experiments by directly testing for *sar*-mediated regulation of the genes *cis* to putative SarA-binding sites. The ability to pursue these experiments to completion will be greatly facilitated as information regarding the complete nucleotide sequence of the *S. aureus* genome becomes available.

Experimental design. The PCR-ABS technology is based on the synthesis of a complex pool of DNAs that contain a random sequence across a target region large enough to bind the transcription factor of interest (15). Based on our experience with other bacterial transcription factors (15), as well as the data of Morfeldt et al. (27) suggesting that SarA binds to a DNA target contained within a 28 bp region defined by heptanucleotide repeats, we have chosen to utilize a 30 bp target region for our first round of experiments. Synthesis of the appropriate pool of DNAs is accomplished by synthesizing three oligonucleotides (Fig. 14). One of these will be a 70-mer that includes defined 20 bp sites at each end. The only requirements for the design of these defined sites are 1) that they have no obvious similarity to the binding site of interest and 2) that they are distinct by comparison to each other. The first of these requirements is somewhat difficult to predict since the consensus binding site for SarA has not been defined (indeed, that is the point of these experiments). At present, all we can do is design the 20 bp sites such that they do not contain any form of the *agr* heptads (AGTTAAG, AGTTAGG, CTTAACT, and CCTAACT) or any sequence longer than a few base pairs (~5-6) that matches any part of the sequence upstream of *cna*. The second requirement (that the 20 bp sites be unique with respect to each other) is important because the PCR-ABS technology utilizes PCR to generate a random sequence across the region between the 20 bp sites (Fig. 14). It is therefore important that the 20 bp sites not contain sequence elements that would allow the primers to anneal to each other. The sequence of the 30 bp region between the 20 bp sites is irrelevant (see below). The identity of the 2nd and 3rd oligonucleotides required for PCR-ABS is based on the identity of the 20 bp sites at each end of the 70-mer. Specifically, one of these (primer 1) is identical to the 20 bp site at one end of the 70-mer while the other (primer 2) is complementary to the 20 bp site at the other end of the 70-mer (Fig. 14). The complementary primer is used in an initial extension to generate a double-stranded DNA molecule. The two primers are then used together to amplify the 70 bp fragment. Importantly, this amplification is done using an equimolar mixture of all four nucleotides such that the 30 bp region between the primer sites has a random sequence. The PCR-ABS technology is based on binding of the protein of interest to this region of random sequence. The basic steps in the PCR-ABS protocol are 1) amplification of a pool of fragments containing random sequences across the 30 bp target site, 2) EMSA experiments in which the pool of fragments is allowed to bind to the transcription factor of interest (e.g. SarA), 3) purification of bound fragments and 4) cloning and sequencing of fragments selected for the ability to bind SarA (Fig. 14). Once a sufficient number of bound fragments have been identified, they are sequenced and the 30 bp regions aligned to reveal the

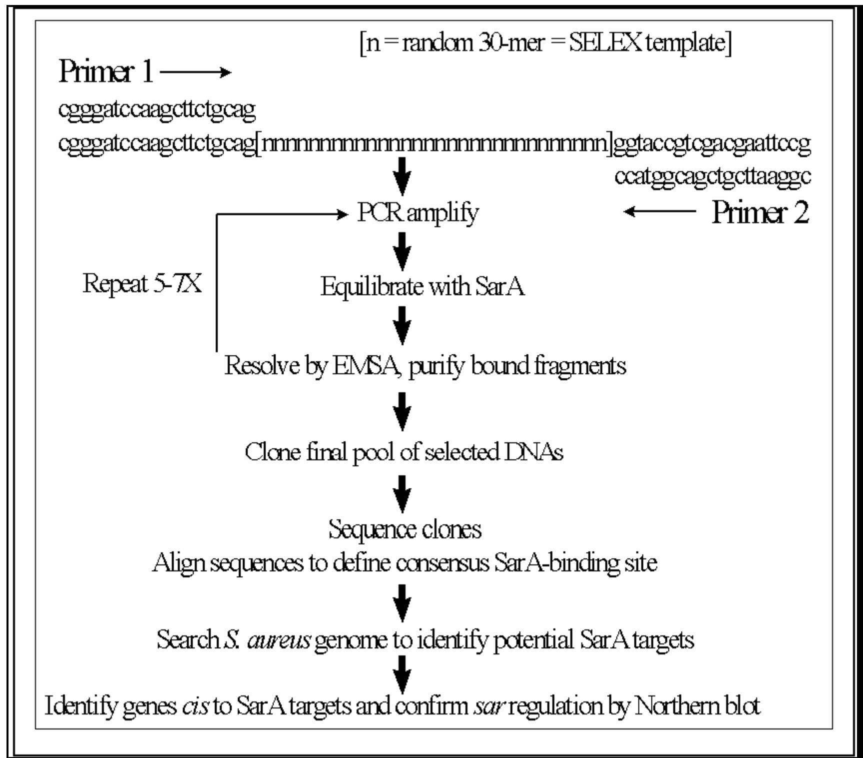

Fig. 14. Flow-chart for for PCR-assisted binding site selection (PCR-ABS).

Figure illustrates experimental steps.

consensus SarA-binding site. An example of such an alignment is shown below (Fig. 15). Once the consensus binding site is established, its presence within the *S. aureus* genome will be assessed by computer-assisted homology searches. In those cases in which a potential binding site is appropriately placed with respect to a potential *S. aureus* promoter, the *sar*-mediated regulation of the gene *cis* to this binding site will be examined by the Northern blot analysis of wild-type strains and their corresponding *sar* mutants. Importantly, the PCR-assisted binding site selection protocol, which is also known by various other acronyms (e.g. SELEX), is well-established in the Co-I's laboratory as evidenced by the successful identification of *trp* repressor binding sites within the *E. coli* genome (15).

	-10	**-9**	**-8**	**-7**	**-6**	**-5**	**-4**	**-3**	**-2**	**-1**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
G	2	96	4	4	0	0	0	100	4	23	9	47	33	22	51	76	17	25	4	41
A	4	4	2	96	0	0	100	0	0	51	2	43	6	0	37	10	10	53	4	26
T	43	0	79	0	0	100	0	0	87	11	8	6	18	78	7	5	63	13	0	9
C	51	0	15	0	100	0	0	0	9	15	81	4	43	0	5	10	10	10	93	24
	C	***G***	***T***	***A***	***C***	***C***	***A***	***G***	***T***	***A***	***C***	***G***	***C***	***T***	***G***	***G***	***T***	***A***	***C***	***G***

Fig. 15. Derivation of a consensus sequence for a DNA-binding protein by PCR-ABS. The top row indicates the relative position of each base. The numbering scheme is based on the *trp* operator sequence and reflects the fact that the binding site is symmetrical. Each possible bases is indicated along the left with the numbers in each cell indicating the percentage of clones that contain the corresponding base at each position. The consensus sequence (defined as the most likely base to be found at each position across the entire target region) is indicated along the bottom.

Figure illustrates how conclusions will be reached on consensus sequence.

Methods #1: Selection of SarA-binding targets. The success of the experiments described in Specific Aim #3 is dependent on the efficiency of our selection for SarA-binding targets. We have chosen the PCR-ABS approach based on the successful utilization of this technology by the Co-I for the characterization of binding sites for the *E. coli trp* repressor (15). The three DNA fragments required to implement the protocol (Fig. 14) were synthesized using an oligonucleotide synthesizer in the Molecular Biology Core Facility at the University of Arkansas for Medical Sciences. The sequence of the 20 bp termini of the 70 bp oligonucleotide and the corresponding 20 bp primers is illustrated in Fig. 14. These sequences were chosen because 1) their utility as PCR primers has already been confirmed by the Co-I in his analysis of *trp* repressor-binding sites (15), 2) each primer includes three restriction sites that can be used to clone selected DNAs (Fig. 16) and 3) neither primer has significant similarity to DNA sequences upstream of the *agr* and *cna* promoters. Based on these characteristics, it should be possible to amplify appropriate target DNAs without introducing an "artificial" SarA-binding site. The central region of the template DNA will be randomized and simultaneously labeled using a mixture of the four bases that includes [^{32}P]dCTP. To generate the starting

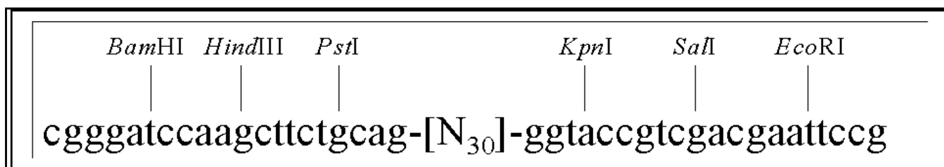

Fig. 16. Restriction sites within the 20 bp primers synthesized for use in the SELEX selection of DNAs containing SarA-binding sites.

DNA and maintain the complexity of the fragment pool, 40-50 μg of synthetic template will be used for a single round of primer extension using the complementary primer (primer 2). The resulting double-stranded DNA fragment will then be gel-purified and mixed with an appropriate concentration of SarA (the appropriate concentration of SarA will be determined empirically according to Pollock (32)). The bound DNAs will then be separated from the unbound DNAs using one of two methods. In the first, we will utilize our affinity-purified SarA antibody and a protein A-sepharose slurry to immunoprecipitate bound DNAs. After elution of bound DNAs from protein A-sepharose pellets, the DNA will be purified by extraction with phenol/chloroform and ethanol precipitation. The second is EMSA, in which case the bound-DNA fraction will be visualized by autoradiography and excised from the gel. The DNA's will then be purified by elution from the gel followed by phenol/chloroform extraction and ethanol precipitation. In either case, DNAs purified from the bound fraction will be subjected to a second amplification using both primer 1 and primer 2. To enhance the selectivity of the protocol, this cycle of binding, purification and amplification will be repeated approximately 5-7 times (see below). In all amplifications, the number of cycles will be kept relatively low (~14-20) to avoid depletion of the nucleotide pool and generation of artifactual PCR products. Additionally, even when we use the immunoprecipitation protocol to purify bound DNAs, we will do EMSA using both bound and unbound DNA fragments to ensure that 1) the size of the amplification product (as reflected by the size of the unbound fraction) is appropriate and 2) the size of the bound fraction is relatively constant (as expected based on the binding of the same protein to a 70 bp DNA fragment). The fraction of bound DNAs should increase with respect to the total DNA loaded on the EMSA gels until a steady state level is achieved. At that point, the DNAs will be cloned by blunt-end cloning into a *SmaI* site and by taking advantage of the restriction sites engineered into the 20 bp flanking regions (Fig. 16). The latter will be employed to facilitate the generation of multiple clones while the former will be used to minimize the bias associated with the removal of fragments that contain the restriction enzyme site. Approximately 100 - 150 of the cloned DNAs will be sequenced using an ABI377 automated DNA sequencer located in the Department of Microbiology and Immunology. DNA sequences will be aligned using the PILEUP program resident in the Genetics Computer Group's software package. A consensus sequence will be derived based on the relative abundance of each nucleotide at each position (Fig. 15).

Once we have derived the consensus sequence defining a SarA-binding site, we will test the specificity of our selection protocol by comparing the affinity of different DNAs for SarA. Specifically, the sequence alignment from the PILEUP analysis will be arranged in a gradient ranging from those sequences that are most homologous to the consensus sequence to those sequences that are least homologous to the consensus. To determine the contributions of specific base pairs in the selected DNA to SarA binding, we will choose representative DNAs that are highly homologous, moderately homologous and somewhat homologous to the consensus and then determine the affinity (equilibrium dissociation constants) for SarA-binding using quantitative EMSA. The test DNAs will include pairs of fragments that have differences in a single base pair that is implicated as important based on its abundance in the pool of selected DNAs. For example, if the third position in the 30-mer is a guanine in a high proportion (e.g. 80%) of selected DNAs, we will test the contribution of that guanine by synthesizing a corresponding fragment that differs only at the third position. For quantitative EMSA, a limiting amount of ^{32}P-labeled DNA (<10 pM) will be equilibrated with various concentrations of SarA and subjected to native gel electrophoresis. The gels will be fixed, dried and exposed to X-ray film. Laser densitometry will be used to quantitate the amount of unbound DNA and the concentration of bound DNA will then be calculated by subtracting the amount of unbound DNA from the total concentration of input DNA. Bound DNA will be plotted against the concentration of SarA protein and an apparent K_D value derived from the concentration of protein that results in 50% complex formation. This procedure was successfully used by the Co-I in studies of the

trp repressor (15) and the myogenic transcription factors MyoD-E12 and myogenin-E12 (14).

Methods #2: Identification of S. aureus genes cis to SarA-binding sites and confirmation of SarA-mediated regulatory control. The primary goal of Specific Aim #3 is to identify *S. aureus* genes under the regulatory control of SarA. From the experiments described above, we will obtain important information regarding the DNA sequence that constitutes a SarA-binding site. This information will be used to search the *S. aureus* genomic database. While the utility of this approach is currently limited by the availability of sequence data, it is anticipated that the results of ongoing efforts to sequence the entire *S. aureus* genome will be publicly available in the near future. As this information is made public, we will use the FASTA searching algorithm resident in the GCG package to search the sequence database for sites that closely match the consensus SarA-binding site. The search parameters will be made degenerate based on the results of the PCR-ABS experiments (i.e. we will include alternative bases at sites that show no clear base preference) and will include any architectural parameters that are found to contribute to binding affinity. An example of such an architectural parameter is the spacing of the heptanucleotide repeats upstream of the *agr* P_2 and P_3 promoters). Putative binding sites will be considered likely candidates for SarA-mediated regulatory control only if they are *cis* to known or reasonable promoters. The genes *cis* to these promoters will then be identified in order to directly test for SarA-mediated regulatory control as described below. Again, our ability to identify target genes will be greatly facilitated by the availability of sequence data derived from the *S. aureus* genome. To test for SarA-mediated regulation of the genes identified by homology searches, we will do Northern blots using total cellular RNA isolated from wild-type *S. aureus* strains and their corresponding *sar* mutants. Northern blots will be done using DNA probes generated by PCR. To ensure that any regulatory effect is detected, RNA for Northern blot analysis will be isolated at various stages of *in vitro* growth (e.g. exponential *and* post-exponential growth). All of the protocols required to carry out the Northern blot analysis of individual target genes have already been described in earlier sections of this proposal. Although not included as part of this proposal, it is anticipated that we will eventually undertake experiments aimed at mutagenesis of genes found to be under SarA-mediated regulatory control and evaluation of the mutants using various animal models of staphylococcal disease.

The text of this research plan is copyrighted.
See the other sections of NIAID's Annotated R01 Research Plan and more advice here:
http://www.niaid.nih.gov/ncn/grant/app/default.htm