Lukomski, Slawomir, Ph.D.

The Moran Foundation 2000-2001 Annual Report

Our proposal entitled: "Functional and Pathological Relevance of the *scl* Gene Encoding Group A *Streptococcus* Extracellular Protein with Similarity to human Collagen" was approved for funding (project number 00-0107) by the Scientific Advisory Committee on behalf of the Moran Foundation's Board od Directors.

REPORT

Group A *Streptococcus* (GAS) is a major cause of human morbidity and mortality worldwide. GAS strains produce numerous cell surface proteins, called adhesins, that help colonize the host by interacting with the host's cellular receptors and extracellular matrix components. Recently, a new class of adhesins of GAS has been discovered in the PI's laboratory. Because the central part of these proteins is composed of a repeated collagen-like motif (Gly-X-X), the proteins have been given the name "Scl" and their genes, the name "*scl*" for "streptococcal collagen-like."

Our hypothesis was: "Streptococcal collagen-like protein is an important pathogenicity factor for group A *Streptococcus* which participates in host-pathogen interactions and induces an autoimmune response against collagen in humans."

To test this hypothesis we proposed following aims ([1]the budget was approved towards the completion of Aims 1 & 3):

Aim 1. We will determine the distribution of the *scl* gene among GAS isolates.

Aim 2. The expression of the *scl* gene by *S. pyogenes* will be studied to determine its biological importance in human GAS pathogenesis.

Aim 3. We will study the Scl protein contribution to host-pathogen interactions in mouse infection models.

Aim. 4. Anti-Scl humoral response during streptococcal infection may be directed against human collagen: autoimmunity issue.

After the proposal was approved for funding, a second related gene, *scl2* (the original *scl* gene was designated *scl1*), was discovered and included into the study.

Ad **Aim 1**

We have screened 50 genetically diverse and epidemiologically unrelated strains by PCR with *scl*-specific primers to assess the distribution of the *scl* and *scl2l* genes. PCR products were obtained from all 50 strains. In addition, considerable size variation was identified between strains expressing different or the same M protein serotypes. The *scl* genes were sequenced to determine the molecular basis of size variation. Complete sequence data were obtained for the *scl1* gene in all 50 strains representing 21 M types, and for the *scl2* gene from 25 GAS isolates of 13 M types.

When comparing the 37 distinct Scl1 variants identified, well conserved and highly polymorphic regions were found. The aminoterminal V (variable) region of the mature Scl1 proteins was highly polymorphic in number of amino acids and primary sequence. The CL (collagen-like) region, located adjacent to the V region, also was highly variable in length as a consequence of the different number of

GXY motifs. For example, 14 GXY motifs were identified in M6 organisms studied, whereas M41 isolate had 62 GXY motifs. The GXY motifs present in the 50 GAS isolates also differed in primary amino acid sequence. In the aggregate, 50 distinct GXY amino acid motif sequences were identified in the CL regions of the 50 Scl1 proteins studied.

In Scl2, the length of the V region also varied from 61 amino acids in a serotype M9 strain to 77 residues found in M3 GAS. As identified for Scl1, the CL region of Scl2 was located C-terminal to the V region. It contained a variable number of GXY motifs ranging from 33 in an M1 strain to 116 triplet repeats in M3-serotype GAS strain. Interestingly, in scl2 gene of many GAS strains the presumed GTG-Val start codon was out of frame with DNA located immediately downstream. CAAAA nucleotide sequence repeats were identified between the presumed GTG-start codon and a CAT (histidine) codon adjacent to the CAAAA repeat region. The number of CAAAA repeats varied greatly among the GAS strains studied, ranging from two (in M77) to 17 (M9). Two CAAAA repeats is the minimal number that would permit correct translation of Scl2. Similarly, the addition of three CAAAA repeats (total of 5) or multiple of three repeats (n = 8 repeats, 11, 14, 17, etc.) should result in in-frame and full-length Scl2 protein translation. Three contiguous CAAAA nucleotide repeats would encode the pentapeptide QNKTK, whereas other numbers of CAAAA repeats should cause premature translation termination. Hence, Scl2 production could be controlled at the translation level by variation in the number of CAAAA repeats, a mechanism documented in gram-negative but not gram-positive bacteria.

Sequence data identified a potential Mga (positive transcriptional activator of virulence genes in GAS) binding site in the promoter region of the *scl1*. Therefore, we hypothesize that *scl1* gene expression is regulated at the level of transcription by Mga, whereas Scl2 protein production is regulated at the level of translation and depends on the number of CAAAA repeats. The latter mechanism has not been previously identified in GAS. Comparison of the *scl* genes and presumed Scl proteins is summarized in table below.

Characteristics of two *scl* genes and Scl proteins in GAS[a]

Gene/ protein	*scl* gene			No. of amino acids in Scl protein				
	Distribution	Locus (frame)	Expression	Ss	V	CL	L	WM
scl1/Scl1	All GAS	1,651,591 -1,652,637 (-3)	Mga-regulated transcription	37	66-85	42-186	30-135	63
scl2/Scl2	All GAS	863,909 -864,860 (+3)	CAAAA-regulated translation	32-57	61-77	99-348	Not present	100

[a]*scl1* gene from 50 GAS strains and *scl2* gene from 25 GAS strains were sequenced.

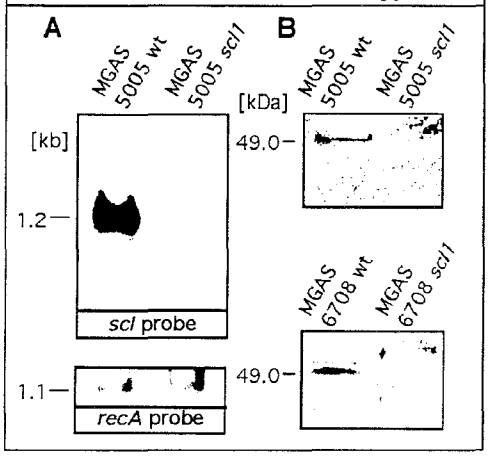
Ss, signal sequence; V, variable region; CL, collagen-like region; L, linker region; WM, cell wall/membrane region

Ad **Aim 3**

To facilitate assessment of the role of Scl1 in host-pathogen interaction, isogenic nonpolar *scl1* mutant derivatives were constructed from serotype M1 strains because this is the most common M type recovered from invasive infections in many case series.

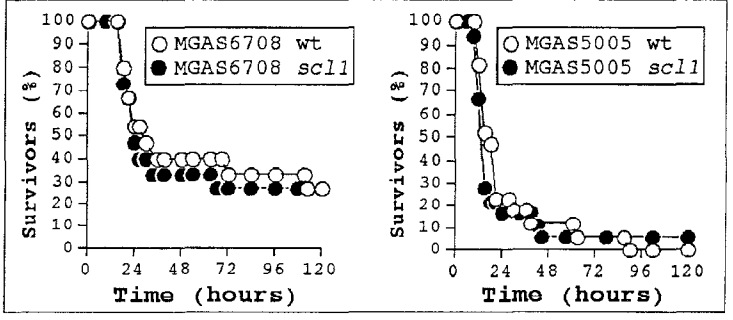
To construct the Scl1-deficient GAS, the *scl1* gene and flanking sequences were cloned into the *E. coli* vector. Next, the nonpolar spectinomycin resistance, spc2, cassette was inserted in-frame, resulting in deletion of virtually the entire *scl1*-open reading frame. This construct was electroporated into the competent cells made from MGAS6708 and MGAS5005, the M1 serotype GAS strains. MGAS6708 was used for the Streptococcal Genome Sequencing Project, whereas MGAS5005, which is representative of the M1 strains commonly recovered from invasive infections. Mutants were analyzed by Southern hybridization and DNA sequencing. Lack of *scl1* expression was confirmed by northern and western blot analyses (Figure to the right).

Fig. 4. Northern and western blot analysis of the *scl1* mutants of M1-type GAS.



We compared the capacity of the wild-type parental organisms and isogenic mutant strains to cause mouse death after i.p. inoculation (Figure below). We inoculated mice intraperitoneally with the wild-type and *scl1* mutant strains (~2.2 x 10^7 CFU were used for the MGAS5005 strain pair and ~1.0 x 10^9 CFU for the MGAS6708 strain pair). No significant difference in mouse mortality was observed for either pair of isogenic GAS strains. Hence, we concluded that Scl1 does not contribute significantly to the virulence of the M1 serotype GAS in this mouse model.

Fig. 5. Mouse lethality after inoculation with isogenic M1 GAS.



Certain extracellular GAS products also have been reported to participate in the pathogenesis of invasive skin disease when mice are injected subcutaneously with GAS. Therefore, we next investigated the effect of inactivation of the *scl1* gene after s.c. inoculation with the wild-type MGAS5005 and isogenic mutant derivative. Mice inoculated with both wild-type and mutant strains developed skin lesions that healed over 3 weeks. Similarly, after an initial weight loss, mice in both experimental groups regained weight. When the weight loss and the severity of the skin lesions in the two experimental groups were compared on days 2 and 4 after inoculation, the differences were not statistically significant (*t* test), probably due to the large variation within animal groups. However, when those parameters were compared over time, the differences in both weight loss ($P < 0.0001$) and severity of soft-tissue pathology (abscess area, $P < 0.0001$; abscess volume, $P < 0.0001$) between mice injected with the wild-type GAS and *scl1* isogenic mutant were statistically significant (mixed-model repeated measures). Hence, inactivation of the *scl1* gene was associated with decreased morbidity in this model of GAS pathogenesis.

PUBLICATIONS

1. Research papers

Lukomski, S., Nakashima, K., Abdi, I., Cipriano, V. J., Shelvin, B. J., Graviss, E. A., and Musser, J. M. (2001) Identification and characterization of a second extracellular collagen-like protein made by group A *Streptococcus*: control of production at the level of translation. Infect. Immun. **69:** 1729-1738.

Lukomski, S., Nakashima, K., Abdi, I., Cipriano, V. J., Ireland, R. M., Reid, S. D., Adams, G. G., and Musser, J. M. (2000) Identification and characterization of the *scl* gene encoding a group A *Streptococcus* extracellular protein virulence factor with similarity to human collagen. Infect. Immun. **68:** 6542-6553.

2. Communications

Lukomski, S., Nakashima, K., Abdi, I., Shelvin, B. J., Graviss, E. A., and Musser, J. M. (2001) Streptococcal collagen-like proteins: gene distribution, expression, and role in pathogenesis. 101th General Meeting of American Society for Microbiology, Orlando, FL

3. Presentations

"Streptococcal Collagen-like Proteins: Biology and Function", presented at the Center for Extracellular Matrix Biology, Institute of Biosciences and Technology, Texas A & M University System Health Science Center, Houston, TX

PROJECT ROLE AND PRESENT STATUS

Funding for the continuation of the study entitled "Streptococcal collagen-like proteins (Scl): Function and Relevance in the Pathology of Diseases Caused by Group A *Streptococcus*." was awarded for the 2001-2002 period.

Preliminary studies obtained through this funding were used for an RO1 grant application by the PI.

ENCLOSURES

Enclosed are reprints of two research papers that acknowledge funds provided by the Moran Foundation, and the copy of an Abstract presented at the meeting.