Review

Toward a Genome-Wide Systems Biology Analysis of Host-Pathogen Interactions in Group A Streptococcus

James M. Musser* and Frank R. DeLeo†

From the Center for Molecular and Translational Human Infectious Diseases Research, The Methodist Hospital Research Institute, and Department of Pathology, The Methodist Hospital, Houston, Texas,* and the Laboratory of Human Bacterial Pathogenesis,† Rocky Mountain Laboratories, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Hamilton, Montana

Genome-wide analysis of microbial pathogens and molecular pathogenesis processes has become an area of considerable activity in the last 5 years. These studies have been made possible by several advances, including completion of the human genome sequence, publication of genome sequences for many human pathogens, development of microarray technology and high-throughput proteomics, and maturation of bioinformatics. Despite these advances, relatively little effort has been expended in the bacterial pathogenesis arena to develop and use integrated research platforms in a systems biology approach to enhance our understanding of disease processes. This review discusses progress made in exploiting an integrated genome-wide research platform to gain new knowledge about how the human bacterial pathogen group A Streptococcus causes disease. Results of these studies have provided many new avenues for basic pathogenesis research and translational research focused on development of an efficacious human vaccine and novel therapeutics. One goal in summarizing this line of study is to bring exciting new findings to the attention of the investigative pathology community. In addition, we hope the review will stimulate investigators to consider using analogous approaches for analysis of the molecular pathogenesis of other microbes. (Am J Pathol 2005, 167:1461–1472)

Introduction and Overview of Group A Streptococcus (GAS)

Diseases

GAS (also known as Streptococcus pyogenes) is a gram-positive bacterium that causes many types of human disease, including pharyngitis and/or tonsillitis, skin infections (impetigo, erysipelis, and other forms of pyoderma), acute rheumatic fever (ARF), poststreptococcal glomerulonephritis, scarlet fever, severe sepsis, puerperal sepsis (childbed fever), meningitis, necrotizing fasciitis, and toxic shock syndrome.1–3 For reasons that remain obscure, ARF and acute glomerulonephritis sometimes occur as late noninfectious sequelae in certain individuals. Globally, GAS causes immense human morbidity and mortality. For example, annual direct costs associated with pharyngitis in the US are thought to be in the range of 1 billion dollars. An estimated 10,000 to 15,000 invasive infections caused by GAS occur each year in the US, and a recent study highlighted the importance of this pathogen as a cause of childhood bloodstream infections in developing countries.4,5 Antibiotic treatment has essentially eliminated rheumatic fever and subsequent rheumatic heart disease in the US and Europe. However, these crippling forms of heart disease remain common in the developing world and in some populations in developed countries.6–8

The Infection Cycle

Because GAS is a human specialist bacterium, transmission occurs mainly person-to-person. Outbreaks of pharyngitis caused by ingestion of contaminated food have been reported, but these are rare. Study of the early molecular events resulting in GAS colonization of the host have focused predominantly on analysis of extracellular adhesin molecules made by the pathogen, showing that many distinct molecules participate.9 The relative contribution of each GAS adherence molecule seems to depend in part on the target host cell and the experimental system used. Some of the GAS outer surface molecules

Supported by the National Institute of Allergy and Infectious Diseases, National Institutes of Health (grant UO1 AI60595 and the Intramural Research Program).

Accepted for publication August 12, 2005.

Address reprint requests to James M. Musser, M.D., Ph.D., Center for Molecular and Translational Human Infectious Diseases Research, The Methodist Hospital Research Institute, and Department of Pathology, The Methodist Hospital, 6565 Fannin St., F816, Houston, TX 77030. E-mail: jmmusser@tmh.tmc.edu.
implicated in adherence include M protein, M protein-like molecules that bind immunoglobulins, lipoteichoic acid, fibronectin- and vitronectin-binding proteins, streptococcal pyrogenic exotoxin B, pullulanase, hyaluronic acid capsule, and C-carbohydrate.\(^2,3,9\) Host molecules implicated as receptors for GAS include fibronectin, vitronectin, fibrinogen, collagen, laminin, CD44, CD46, and galactose- and fucose-containing glycoproteins. However, this area of GAS pathobiology has been explored less than the bacterial products that contribute to adherence. A full discussion of these important topics is outside of the realm of this review.

Although adherence to host cells is a crucial early event in the infection cycle, the human oropharynx is bathed in saliva, which is the first substance with which GAS interacts. Saliva contains many components of the innate and acquired immune systems and plays a pivotal role in defending the oropharynx against invading microorganisms.\(^10\) Hence, infection and subsequent transmission to new hosts depends on the ability of GAS to survive and proliferate in saliva (see below).

After initial host interaction, GAS may either remain at the site of colonization in a relatively quiescent state or proliferate to cause local disease such as pharyngitis or pyoderma. In a few individuals, the pathogen avoids clearance or control by host immune mechanisms and initiates a more invasive infection characterized by extensive local tissue destruction and spread and/or intravascular entry and dissemination. The tissue destruction in many patients can be fulminant, resulting in complete destruction of the extracellular matrix and death. The exact molecular mechanism(s) responsible for tissue destruction are unknown, but production of substances that directly degrade host molecules or subvert normal host defenses is known to participate.\(^11,12\)

**Complex Regulatory Networks Participate in Virulence**

All bacterial pathogens, including GAS, have molecular mechanisms dedicated to coordinately regulating expression of virulence factors in response to environmental stimuli (Figure 1).\(^13\) This coordinate regulation provides a strategy for bacteria to adapt very rapidly to changes in their extracellular milieu and thereby enhance survival. A key feature of many of these networks is control by a two-protein interaction cascade involving a sensor and response protein, commonly referred to as a two-component gene regulatory system (TCS). These TCSs control expression of unlinked genes encoding toxins, adhesins, and other virulence-associated molecules that promote bacterial survival and pathogenesis.\(^13\) As an initial activation step, a transmembrane sensor protein, usually a histidine kinase, detects extracellular stimuli and undergoes autophosphorylation. The histidine kinase subsequently phosphorylates a response regulator, which in turn acts as a transcription factor for specific target genes. GAS has 12 conserved TCSs, but only 4 of these 12 (CovR-CovS, FasBCAX, Ihk-Irr, and Spy0874/0875, also known as SptR/SptS) have been well studied (see below).\(^14\) In addition, a GAS gene called mga (multiple gene activator) also serves as a crucial gene regulator.\(^14\) The mga gene encodes a trans-acting positive
regulatory protein (Mga) that activates transcription of several genes encoding proven or putative GAS virulence factors, including M protein, M-like proteins, the streptococcal inhibitor of complement (C5a peptidase), and a lipoprotein lipase (opacity factor). Interestingly, although sequence homologies suggest that Mga is the transducer protein of a TCS, the putative sensor protein that responds to environment perturbations and subsequently activates mga transcription has not been identified. Several other important regulatory systems have been described in GAS, but space constraints do not permit us to review their contributions to host-pathogen interactions. We refer the interested reader to a recent review of this topic.14

Rapidly Emerging Discipline of Systems Biology Research

An investigative discipline now referred to as systems biology research has emerged rapidly in recent years.15,16 One publication15 defines systems biology as "...a scientific discipline that endeavors to quantify all of the molecular elements of a biological system to assess their interactions and to integrate that information into graphical network models that serve as predictive hypotheses to explain emergent behaviors." Inherent in this definition is that genome-wide strategies will be used to generate very large multiparameter datasets, including but not limited to comparative genome sequences, transcriptome, proteome, metabolome, and so forth. The datasets are systematically integrated and interrogated to achieve the particular goal of interest. Much of the effort thus far has focused on cancer biology or analysis of model organisms such as yeast, with relatively little attention given to problems in infectious diseases research. Analysis of the molecular pathogenesis of infectious disease by a systems biology approach is especially complicated, in part because pathogens are highly diverse genetically, multiple phases of the infectious process can be prolonged and anatomically distinct (eg, skin, tissue, blood, cerebrospinal fluid, individual and multiple organs), and host immune processes are multiphasic. Thus, in the case of bacterial pathogenesis, there exists concern about parameters such as the genotype of the infecting strain and host, phase of infection, immune status, and treatment modality, to name a few variables. Clearly a systems biology analysis of this type of problem is quite complicated. However, it is reasonable to suspect that steps toward this goal will yield new insights about host-pathogen interaction at each juncture. Currently, we think of systems biology research of infectious agents with a holistic view of integrating molecular events (broadly defined) occurring in the pathogen with those occurring in the host. Initial steps toward this type of analysis in our laboratories involve a highly integrated, discovery-enabling infrastructure coupled with animal infection models and human patient studies (Figure 2).

The systems biology approach toward understanding GAS-host interactions and pathogenesis has included sequencing the genome of multiple GAS strains. The wealth of resulting genome data were subsequently used to produce expression and DNA-DNA microarrays, from which many dozens of target genes of interest were identified. The role of specific GAS target molecules in pathogenesis was then examined in animal models and in vitro host-pathogen assays and was coupled in multiple instances with a repository of human clinical data (Figure 2). Proteomics also contributed significantly to the identification of GAS virulence molecules, although the topic is not discussed extensively in this review. This integrated methodology has led to a significant number of new discoveries in GAS pathogenesis, many of which are generally applicable to mechanisms of virulence in other human pathogens.

GAS Genome Sequences: an Embarrassment of Riches

Introduction and Background

We believe a systems biology approach to analysis of host-pathogen interaction must start with a reasonably complete description of the components of the system. From the pathogen side of the equation, an excellent starting point is the genome sequence of strains of interest. The first bacterial genome sequence was reported in 1995,17 and now more than 150 bacterial genome sequences have been published, including at least one strain of every major human pathogen (http://www.tigr.org/tigr-scripts/CMR2/CMRHomePage.spl and http://www.genomesonline.org/). Until relatively recently, it was common to have only one genome sequence available for a pathogenic bacterial species. However, it has long been known that the amount of genetic diversity present in most bacterial species far exceeds that for eukaryotic organisms. That knowledge, coupled with well-recognized differences in clinically relevant behavior between strains of individual bacterial species, relatively small genome size, decreasing DNA sequencing costs, therapeutics research, and concerns about illegitimate release of certain pathogens, fueled the effort to sequence the genome of multiple isolates of many important human pathogens. Thus, more than one genome sequence is now publicly available for many bacterial pathogens, and for several species, including GAS, more than three genomes have been sequenced. The availability of multiple genome sequences has been a tremendous asset to pathogenesis research.

For more than 50 years, GAS strains have been classified based on serological diversity in M protein, a major surface antigen and virulence factor. However, serological typing of strains has been replaced by sequencing of the hypervariable part of the enm gene encoding M protein. More than 125 enm types are recognized, and this allelic diversity has been useful for categorizing strains for epidemiological studies.4,18–21 Although no one M serotype or enm type is solely responsible for any single type of GAS infection, strains expressing certain M types have been repeatedly associated with specific dis-
eases. For example, serotype M1 and M3 strains are the leading cause of invasive infections in many western countries, and M18 strains have been commonly associated with rheumatic fever.\textsuperscript{19–21} These epidemiological observations have led to the concept that certain serotypes are differentially capable of causing specific clinical diseases. A corollary of this idea is that genetic differences between GAS strains account for some of these nonrandom disease associations. Importantly, this line of thinking has stimulated much of the interest in sequencing the genome of GAS strains with distinct disease associations.

Complete genome sequences are available for eight GAS strains (serotypes M1, M3, M5, M6, M18, and M28) commonly causing pharyngitis and invasive disease (http://www.sanger.ac.uk/Projects/S_pyogenes/).\textsuperscript{22–28} The availability of genome sequences of M protein serotypes causing distinct diseases has yielded a tremendous amount of new information important for a systems biology approach to pathogenesis research.

**General Concepts Revealed by Comparative Genomics**

The eight genome sequences have provided extensive data regarding the breadth of strain variation within and between serotypes, molecular correlates of disease specificity, and many unexpected leads for pathogenesis research. All strains studied thus far have a genome size
of ~1.8 to 1.9 Mb, five or six highly conserved rRNA operons, and a core group of proven and putative virulence genes. Approximately 90% of gene content is shared among strains and constitutes the core GAS genome. Importantly, the genomes of all strains sequenced are polylysogenic, that is, they contain multiple prophages or prophage-like elements. (Prophages are bacterial viruses that are integrated into the chromosome.) Virtually all GAS prophages encode one or two proven or putative extracellular virulence factors such as pyrogenic toxin superantigens (PTSAgs), DNases, novel phospholipase A2 (SlaA), efflux pump responsible for resistance to macrolide antibiotics, and a novel cell wall-anchored protein hypothesized to be an adhesin. Many of these prophage-associated virulence factor genes were unknown before their discovery by GAS genome sequencing projects. Additional specific discoveries revealed by the genome sequencing projects are described below.

**Genome Sequence and Population Genomics of ARF-Associated GAS Serotype M18 Strains**

ARF is the leading cause of global preventable pediatric heart disease. This disease is a devastating sequela of GAS infection that usually occurs several weeks after an antecedent throat infection. The molecular pathogenesis of ARF and subsequent rheumatic heart disease has not been fully delineated, but autoimmune mechanisms likely play a prominent role. It is believed that only a relatively small number of the more than 125 M protein serotypes have rheumatogenic potential in humans who are predisposed, presumably for host genetic reasons. For example, serotype M18 GAS strains have been repeatedly implicated in ARF outbreaks in the US. As a first step toward gaining new insight into ARF pathogenesis, we sequenced the genome of a M18 strain cultured from a patient with ARF. In addition, we used DNA microarray analysis to compare the genomes of 36 M18 strains, including organisms cultured from patients with ARF and pharyngitis during two epidemics of ARF in Salt Lake City, UT.

Many genes are unique to serotype M18 strains. Importantly, this serotype has genes encoding unique secreted proteins involved in human-GAS interactions, including two previously uncharacterized novel PTSAgs that are encoded by a prophage. Whole-genome DNA-DNA microarray analysis of 36 serotype M18 strains, collected from diverse localities throughout 6 decades, revealed that most regions of variation corresponded to prophages or prophage-like elements. Two epidemics of ARF occurring 12 years apart in Salt Lake City, UT, were caused by serotype M18 strains that were genetically identical, or nearly so. The analysis provided a critical foundation for accelerated research into ARF pathogenesis and a molecular framework to study the plasticity of GAS genomes.

**Insight into the Emergence of Drug-Resistant GAS Strains: an Outbreak of Pharyngitis Caused by a Serotype M6 Macrolide-Resistant Clone**

GAS infections are commonly treated with penicillin or a related β-lactam antibiotic. Strangely, despite decades of intense selective pressure, all GAS strains remain very susceptible to this class of antibiotics. Erythromycin and related macrolide antibiotics sometimes are used as an alternative treatment for GAS pharyngitis. In contrast to β-lactams, resistance of GAS to macrolide antibiotics was described as early as the late 1950s and has increased markedly worldwide in the last decade.

Martin and colleagues reported an outbreak of pharyngitis caused by erythromycin-resistant serotype M6 GAS strains among schoolchildren in Pittsburgh, PA. This event was of particular concern because within a few months the frequency of macrolide-resistant GAS increased rapidly, and drug-resistant strains spread widely to surrounding communities. Molecular studies implicated serotype M6 strains of a single clone, and drug resistance was due to the presence of the mefA gene encoding a macrolide efflux pump.

Inasmuch as serotype M6 strains are one of the more common causes of pharyngitis and invasive infections and there was considerable conflicting information in the literature about the molecular mechanism of acquisition and transfer of the mefA gene in GAS, we chose to sequence the genome of a genetically representative strain. Interestingly, we found that the mefA gene in the Pittsburgh clone was encoded by an unusual 58.8-kb foreign genetic element with characteristics of both a transposon and a prophage. It is likely that acquisition of the mefA element by lateral (horizontal) gene transfer and subsequent expansion of the drug-resistant clone were key contributors to the outbreak.

The chimeric element also encoded a novel extracellular protein that is expressed during GAS-human interaction. The protein is related to molecules that facilitate adhesion to host cells in other microbial pathogens, suggesting that it assists the organism in colonizing the posterior pharynx. Taken together, sequencing of this genome and the studies it facilitated shed new light on the important mechanism of increased antimicrobial resistance in GAS.

**New Understanding about the Pathogenesis of Puerperal Sepsis Revealed by the Genome Sequence of a Serotype M28 GAS Strain**

Puerperal sepsis, also known as childbed fever, killed up to one in every six mothers who delivered babies in some European hospitals in the 19th century. Clinical investigations conducted by Semmelweis in the 1840s determined that puerperal sepsis was transmitted to pregnant women in labor by attending doctors. Semmelweis’ studies eventually led to widespread implementation of infection control practices such as routine hand washing,
Molecular Mechanisms Underlying the Emergence of Severe GAS Infections: the Key Role of Lateral Gene Transfer (LGT)

The molecular events contributing to the evolution and emergence of unusually virulent pathogens are poorly understood. Bacteria can evolve slowly by accumulation of point mutations or rapidly in a quantum leap by acquisition of new genetic material through lateral (also known as horizontal) gene transfer (LGT). New combinations of genes or allelic variants can increase the fitness of a bacterium by conferring unique properties such as immune evasion, antimicrobial agent resistance, and the ability to colonize a previously unexploited ecological niche. Three well-described processes known to promote LGT are transduction, conjugation, and transformation.

In the mid-1980s, the frequency and severity of invasive infections caused by serotype M1 GAS precipitously increased in the US and many other countries. Although serotype M1 GAS are usually the predominant serotype recovered from invasive disease episodes, the increase in disease frequency and severity (especially necrotizing fasciitis) caused by M1 strains was noteworthy. However, no information was available about the molecular processes responsible for this phenotypic shift at the time of the episodes.

To better understand the molecular events involved in the origin of new pathogenic bacteria, we studied the evolution and development of a virulent clone of serotype M1 GAS that emerged in the mid-1980s. This study was greatly facilitated by the availability of the genome sequence of one serotype M1 GAS organism (strain SF370). Inasmuch as strain SF370 is genetically distinct from the great majority of strains responsible for contemporary infections caused by M1 strains, and is less virulent, we hypothesized that acquisition of novel genetic elements had created an unusually successful clone of serotype M1. To test this hypothesis, we sequenced the genome of a strain that is genetically representative of organisms responsible for contemporary episodes of M1 infections. Genome sequencing, DNA-DNA microarray analysis, and high-throughput single nucleotide polymorphism analysis revealed that this clone evolved from an ancestral strain through a complex series of LGT episodes. These events included acquisition of prophages encoding SpeA and extracellular DNase virulence factors. Another LGT event involved reciprocal recombination of a 36-kb chromosomal segment encoding the extracellular toxins NAD-glycohydrolase (NADase) and streptolysin O. Virtual identity in the 36-kb region present in contemporary M1 isolates and serotype M12 isolates suggested a model in which a serotype M12 strain served as the donor of this region. The most likely mechanism involved generalized transduction, a process in which a bacteriophage inadvertently packages a segment of GAS chromosomal DNA, rather than the phage genome, into its phage head. This LGT event was associated with significantly increased streptolysin O and NADase production. These findings have implications for understanding the molecular events underlying the emergence of other bacterial clones that are unusually virulent or have other characteristic disease phenotypes; they also stress the importance of using highly integrated genome-wide analyses to study clone emergence and abrupt changes in infection character.

Genome-Wide Dissection of the Molecular Events Underlying GAS Epidemics: the Serotype M3 Model

Molecular factors that directly contribute to bacterial epidemics remain elusive. GAS is an ideal pathogen to investigate this topic because the organism undergoes rapid changes in disease frequency and severity caused by strains expressing certain M protein serotypes. We conducted a molecular genetic dissection of two epidemic waves by using multipronged approach that involved sequencing the genome of a relevant strain, performing functional analysis of novel extracellular proteins encoded by the genome, and then using comparative genomic analysis of the strains.

To initiate this line of investigation, the genome of a strain isolated from a patient with invasive disease was sequenced. As found for other GAS strains, prophage-like elements accounted for the great majority of variation in gene content relative to other GAS genomes. The strain had phage genes that encoded proven and putative virulence factors, including PTSAGs, and a novel phospholipase A1 virulence factor (designated SlaA) with similarity to snake venom toxins. Importantly, we
discovered that serotype M3 strains with the prophage-encoded slaA gene increased dramatically in frequency late in the 20th century, commensurate with the rise in invasive infections caused by serotype M3 strains. The study showed that phage-mediated recombination involving virulence factor genes played a critical role in the emergence of a new, unusually virulent clone of GAS. Japanese investigators subsequently confirmed our findings.25

The availability of the genome sequence of a serotype M3 strain opened the door to a detailed analysis of molecular factors contributing to disease epidemics. We hypothesized that analysis of a comprehensive population-based strain sample recovered from patients with invasive infections together with genome-wide investigative methods would provide new insight into this fundamental infectious disease problem. To test the hypothesis, 255 contemporary M3 strains obtained from an 11-year population-based surveillance study of invasive disease in Ontario, Canada, were analyzed.49 Genetic diversity in these strains was investigated by several state-of-the-art methods, including DNA-DNA microarray and whole-genome PCR tiling. The core genome of all strains was identical in gene content. That is, variation in prophage content accounted for all differences in gene content between strains.

Sequence analysis of the emm gene and single nucleotide polymorphism analysis were used to extend analysis of genetic relationships among the 255 M3 strains. The strains were pauciclonal, with a limited number of subclone groups identified. Interestingly, statistically significant associations were discovered between prophage genotypes and GAS disease types. Hence, by integrating genome-wide molecular genetic analysis techniques with detailed epidemiological and clinical information, temporal changes in the gene content of M3 subclones were linked to disease phenotype. Collectively, acquisition or loss of prophages, allelic variation in chromosomal genes, expansion of subclone populations, and introduction of new subclone variants were shown to contribute to peaks of infection and different infection types.49

One finding from this study was particularly noteworthy and provided key insight into a potential mechanism accounting for clone emergence. Molecular genetic analysis, combined with immunological studies, implicated a 4-amino acid duplication in the extreme N-terminus of M protein as a factor contributing to rapid clone emergence. Linear epitope mapping with antisera raised against M protein-based synthetic peptides revealed differential reactivity to the peptides representing the extreme N-terminus of M protein. In addition, the difference in M protein sequence resulted in significantly decreased ability of human polymorphonuclear leukocytes (PMNs) to phagocytose the GAS strain that expressed the variant M3 protein. Taken together, the findings indicated that the 4-amino acid duplication conferred altered immune recognition to the M3 protein. Thus, it is possible that GAS strains with the variant M3 sequence rose to prominence as a consequence of host immunological selective pressure.

Expression Microarray Analysis of the GAS Transcriptome

Expression microarray (transcriptome) analysis has dramatically accelerated the pace of discovery in every area of biomedical research, and microbial pathogenesis is no exception. Several expression microarray studies have examined GAS, and extensive new information has been obtained about the nature and extent of changes in the pathogen’s transcriptome in response to various perturbations.28,51–56 This global approach to understanding pathogenesis was facilitated by the information obtained from GAS genome sequencing projects, as described above. Initial studies used conventional spotted microarrays and focused on in vitro analysis, but more recently a custom Affymetrix high-density oligonucleotide array has been used for ex vivo and in vivo studies.55,56 The result has been substantial new understanding about how GAS reacts to its environment and the mammalian host. Here we summarize key findings from several of these investigations.

Virulence Control and the covR-covS Gene Regulatory System

The CovR-CovS (Cov, control of virulence; also known as CsrR-CsrS) TCS has been investigated extensively.14,52,55,57,58 Genetic inactivation of this TCS produces a mutant strain that is hypervirulent in mice, has a mucoid colony phenotype associated with overexpression of hyaluronic acid capsule, and has increased resistance to in vitro killing by human PMNs. In addition, increased transcripts for proteins likely to promote survival and virulence, including streptokinase, streptolysin S, and a DNase, were identified in the isogenic mutant strain by conventional Northern blot analysis. However, expression microarray analysis afforded assessment of the full impact of the CovR-CovS TCS on GAS gene expression.52 In vitro, this TCS influences (directly or indirectly) transcription of 15% of all GAS genes, including many that encode surface and secreted proteins mediating host-pathogen interactions. CovR also plays a central role in gene regulatory networks by influencing expression of genes encoding transcriptional regulators, such as other TCSs. The key findings of the microarray analysis were confirmed in a mouse soft-tissue model of invasive infection and by ex vivo analysis of GAS growth in human blood (see below).52,55

Molecular Interactions of GAS with Human Saliva

The ability of microbes to adapt successfully to specific host environments by differential gene expression is essential for survival and causing human disease. Although many human bacteria persist for extended periods in distinct anatomical niches, relatively little is known about the molecular mechanisms that facilitate such persistence. The human oropharynx is the predominant site of
GAS infections. Although saliva is a vital component of host defense in the human oropharynx, little is known about how GAS responds to interaction with this substance. To increase our understanding of this interaction, we investigated the growth of GAS in saliva obtained from healthy volunteers. We found that GAS responds in a complex manner to growth in human saliva. Importantly, two abundantly secreted proteins made by GAS (Sic, streptococcal inhibitor of complement, and SpeB, a potent cysteine protease) were discovered to contribute to prolonged survival in saliva.

To gain enhanced understanding of the molecular mechanisms used by GAS to interact with saliva, we recently analyzed the transcriptome of a contemporary serotype M1 GAS strain (sequenced strains, see above) during growth in this fluid. An iterative expression microarray method was used. This powerful strategy involves expression array analysis of a wild-type organism to generate an initial dataset. The information that is revealed is then exploited to identify genes potentially critical to pathogen-host interaction under the condition of interest. Subsequently, a target gene of interest is genetically inactivated, and a second (iterative) expression microarray experiment is performed to compare the transcriptome of the wild-type and mutant organisms grown under identical conditions. Thus, the datasets obtained from these experiments rapidly yield information about complex gene regulatory networks operating in distinct settings. With respect to interaction of GAS with saliva, one example illustrates the power of this strategy. Expression microarray analysis of wild-type GAS revealed that a TCS (designated Spy0874/0875, also known as SptR/SptS) of unknown function had very high transcript levels in human saliva, suggesting a contribution to growth and persistence in this fluid. Consistent with this idea, an isogenic nonpolar mutant strain (ΔSptR) was dramatically less able to persist in human saliva compared to the parental strain. Transcriptome analysis revealed that compared to the parental strain, the isogenic mutant strain had decreased transcript levels of multiple genes encoding several well-known virulence factors and proteins involved in complex carbohydrate acquisition and utilization pathways. Thus, iterative microarray analysis identified a genetic program that significantly increased persistence in saliva of a major human pathogen, thereby providing new insight into how bacteria survive in the host. Taken together, the data suggest that the molecular processes that enhance human colonization and survival in the upper respiratory tract are underway well before the pathogen contacts the epithelial cell surface. From the standpoint of translational research, the data suggest that using novel inhibitors to target this particular two-component system would be fruitful.

Transcriptome Dynamics in the Upper Respiratory Tract

To what extent, if any, do molecular events that occur during in vitro growth reflect in vivo processes? This important question has long plagued investigators in all areas of biomedical research, including that of microbial pathogenesis. Identification of the genetic events that contribute to host-pathogen interactions is important not only for understanding the natural history of infectious diseases but also for developing new therapeutics. Central to this goal in recent years have been transcriptome studies conducted on pathogens, although most have focused on specific end points or disease phases rather than analysis of the entire time course of infection. To gain a more complete understanding of how bacterial gene expression in vivo changes throughout time, the transcriptome of our contemporary serotype M1 GAS strain was analyzed during an 86-day infection protocol in 20 cynomolgus macaques with experimental pharyngitis. A key discovery was that the temporal pattern of GAS gene expression in pharyngitis is intimately linked to three distinct phases of infection. For example, successful colonization and severe inflammation were significantly correlated with superantigen gene expression. Similarly, differential expression of TCS regulators CovR and SptR was significantly associated with GAS colony-forming units, inflammation, and phases of disease. It is important to note that SptR was also identified as a key gene regulator by the ex vivo saliva transcriptome studies described above. This nonhuman primate study confirmed the in vivo relevance of many of the in vitro observations described recently. The study also provided extensive new knowledge of the gene expression patterns characterizing each clinical phase of pathogen-host interaction. The detailed understanding of GAS gene transcript patterns throughout time in an infected host provides new avenues for targeted investigation of proven and putative virulence factors and genes of unknown function and will assist therapeutics research. An analogous study of the changes in the monkey transcriptome occurring throughout the 86 days of this experiment is currently underway (K. Virtaneva and J. M. Musser, unpublished data).

GAS Transcriptome Dynamics during Growth in Human Blood

During the transition from a throat or superficial skin infection to an invasive infection, GAS must adapt to changing environments and host factors. Recently, we used transcript profiling and functional analyses to investigate the transcriptome of a serotype M1 GAS strain in human blood ex vivo. This was a particularly pertinent investigation because GAS sepsis is a devastating infection with high morbidity and mortality rates. Hence, insight into molecular events transpiring in blood is crucial. Using an Affymetrix custom microarray, we discovered that global changes in GAS gene expression occurred rapidly in response to human blood exposure. Increased transcription was identified for many genes that are likely to enhance bacterial survival, including those encoding PTSAgs and host-evasion proteins. For example, on blood exposure, expression increases for GAS genes encoding proteins that interact with host cell surfaces.
(adhesins, such as M protein; collagen-binding proteins; and capsule) and that contributes to the evasion of the host innate defenses (Sic, Mac, and SpeA). The microarray analysis confirmed and extended evidence that the CovR-CovS TCS enhances bacterial survival during disseminated host infections.57,58 This study provided crucial insights into strategies used by a bacterial pathogen to thwart host defenses and to survive in human blood and suggested new vaccine and therapeutic strategies.

Molecular Mechanisms Used by GAS to Circumvent Destruction by the Innate Immune System

Identification of a GAS Genetic Program that Protects against Killing by Human PMNs: the Ihk-Irr Two-Component System

Most microbes are destroyed by PMN-derived reactive oxygen species and/or microbialicidal granule products, which are released into pathogen-containing phagosomes. However, GAS is relatively resistant to killing by these anti-bacterial components.53,54,61–64 Although progress has been made, the molecular basis for this resistance is incompletely understood. Two recent transcriptome-based studies shed new light on the gene regulation mechanisms used by GAS to avoid destruction by human PMNs.

To begin analysis of the gene regulatory pathways used by GAS to circumvent the human innate immune system, we conducted a comprehensive analysis of changes in GAS gene expression during phagocytic interaction with human PMNs.53,54 Genes related to prophage, oxidative stress, cell-wall biosynthesis, virulence, and novel extracellular secreted proteins of unknown function were up-regulated during phagocytic interaction (Figure 3). Importantly, an essential role in immune evasion was discovered for the Ihk-Irr TCS using an isogenic mutant strain in which this gene regulatory system was inactivated. That is, the mutant strain was killed far more rapidly by human PMNs than the wild-type parental strain. Differences in survival were attributed to increased susceptibility of the irr-mutant strain to microbicidal components present within the PMN phagosome.53,54 Thus, the Ihk-Irr TCS controls expression of genes that facilitate evasion of PMN-mediated bacterial killing, thereby assisting GAS survival and pathogenesis (Figure 4). Consistent with these findings, the irr gene is highly expressed in human GAS pharyngitis.65

A subsequent study investigated the molecular basis for the inability of the irr-mutant strain to withstand destruction by neutrophils. Iterative expression microarray analysis revealed that Ihk-Irr significantly influenced 20% of all transcripts in the GAS genome. Importantly, genes implicated in cell wall metabolism, transcription, oxidative stress, and virulence were significantly down-regulated in the isogenic irr-mutant strain.54 Compared with the wild-type strain, the irr-mutant strain had significantly decreased resistance to killing by H2O2 and neutrophil primary granule components. Moreover, expression of ihk and irr were increased significantly in wild-type GAS strains exposed to H2O2 and PMN primary granules. Collectively, the data indicated that Ihk-Irr functions as a sensor for components of human innate host defense. Thus, these iterative expression microarray studies helped to identify a global pathogen-protective transcrip-
tome that is engaged during interaction with the innate immune system.

**GAS Modulates Human PMN Gene Expression and Function**

Inasmuch as PMNs produce an array of cytotoxic molecules that can cause host tissue destruction, neutrophils typically undergo apoptosis after phagocytosis as a means to facilitate healthy turnover of effete or spent cells.64,66–68 This process is linked to the production of NADPH oxidase-derived reactive oxygen species.69–72 Using human oligonucleotide microarrays throughout the course of several independent studies, Kobayashi and colleagues64,67,71 discovered an apoptosis differentiation program in human PMNs that likely facilitates the resolution of infection. The apoptosis differentiation program is comprised of a core set of genes commonly induced or repressed after uptake of bacteria.71

During phagocytic interaction with human neutrophils, GAS modulates expression of the core set of PMN genes comprising the common apoptosis program.64 However, the pathogen also triggers changes in gene expression not observed with other bacteria tested, including repression of genes involved in specific cell survival pathways.64 Changes in the expression of these genes significantly parallels accelerated PMN apoptosis and host cell lysis (Figure 4). Based on these studies, we proposed that there are two fundamental outcomes of the interaction of neutrophils with bacteria.64,68 On one hand, neutrophils are activated, bacteria are killed, and there is induction of apoptosis leading to the resolution of infection. Alternatively, bacterial pathogens such as GAS circumvent killing by neutrophils and thus alter the apoptosis program to survive and cause disease. The ability of GAS to survive and cause eventual PMN lysis is in part regulated by the Ihk-Irr TCS as described above (Figure 4).

**Genome-Wide Analysis and Therapeutics Research: Vaccines**

In recent years, a paradigm shift in bacterial vaccine research has occurred due to the development of a strategy now referred to as “reverse vaccinology.”73–76 The basic concept behind the reverse vaccinology approach is to use the complete genome sequence of a pathogen to identify a large number of novel protective antigens, express each candidate in recombinant form, and test each protein in one or more assays for capacity to confer immunological protection. Potential candidates are initially identified by computer analysis of the bacterial genome sequence with algorithms that predict secreted or surface-exposed proteins. Testing of these antigens requires high-throughput cloning and expression of recombinant antigens as well as the use of *in vitro* and/or *in vivo* assays of protective capacity. Pioneering studies were performed using *Neisseria meningitidis*, an important cause of sepsis and meningitis worldwide.73,74 The complete genome sequence of a serogroup B strain of this pathogen was determined, 344 candidate proteins were overexpressed in *Escherichia coli*, and 28 novel proteins were ultimately identified as protective antigens.74 Subsequently, a similar strategy, with some improvements, was used to identify novel protective antigens against group B *Streptococcus*76 and GAS.77–79 The GAS efforts have involved an integrated systems biology approach that uses diverse techniques such as proteome, immunome, and *in vitro* and *in vivo* transcriptome analyses; comparative genomics and bioinformatics analysis; and experimental infection models in mice and monkeys.56,65,77–79 Importantly, this integrated strategy has enabled far faster identification of conserved candidate antigens than otherwise possible by conventional vaccine research methods.

**Summary**

The goal of this review was to summarize some of the key recent advances that have been made in GAS pathogenesis research by deploying a multipronged genome-wide systems biology strategy. The progress made thus far has provided many new avenues for basic pathogenesis research and translational research that is focused on developing novel therapeutics. Although many gaps in our knowledge of GAS-human interactions remain to be filled, we hope that our review stimulates other investigators to consider using analogous integrated strategies for studying the molecular basis of microbial pathogenesis.

**Acknowledgments**

We thank all members of our laboratories who provided thoughtful suggestions to improve the manuscript.

**References**

8. Veasy LG, Wiedmeier SE, Orsmond GS, Ruttenberg HD, Boucek MM, Roth SJ, Tait VF, Thompson JA, Daly JA, Kaplan EL: Resurgence of
Staali L, Morgelin M, Bjorck L, Tapper H: Streptococcus pyogenes


