ISSN : 0974 - 7435 Volume 9 Issue 6

BioTechnology
An Indian Journal

FULL PAPER

BTAIJ, 9(6), 2014 [242-247]

A neoteric approach: Designing the virulence regulation model for *Staphylococcus aureus* by contemplating boolean network

Amogh Jalihal, V.Praveen Krishna, S.Adline Princy*
Quorum sensing Laboratory, School of Chemical and Biotechnology,
SASTRA's Hub for Research and Innovation (SHRI), SASTRA University, Thanjavur, Tamil Nadu, (INDIA)
Email : adlineprinzy@biotech.sastra.edu

ABSTRACT

The virulence regulation of *Staphylococcus aureus* is controlled by a complex set of genetic-level interactions which can be modeled using existing mathematical frameworks. The different levels of regulation can be modeled using different systems, and are thus classified as Motifs, Modules and Games. In this review, an overview of the regulatory sequences and the frameworks to model them have been consolidated.

KEYWORDS

Mathematical modeling;
S. aureus;
Boolean network.

INTRODUCTION

Staphylococcus aureus has been of immense interest largely due to the development of drug-resistant strains. The regulatory networks and various virulence factors of *S. aureus* are being characterized, and the complex interconnected pathways are only now being understood. These regulatory mechanisms present a unique opportunity for mathematical modeling; such a model will not only explain the existing behavior of the system, but will also help us predict new functions and capabilities along with novel regulatory pathways at the genetic as well as organismal levels, within certain realistic constraints. Also, such mathematical models will aid in the correlation of population wide behavior to organism-specific processes within the species i.e. genetic regulatory pathways. On a general, non-specific scale, the elements of the complex regulatory system in *S. aureus* can be classified as the virulence module. This module can be expressed as a series of motifs which constitute the overall system. These motifs are the individual molecular level systems, for example, the *agr* locus forms a motif for a classical TCSTS (two component signal transduction system), or the motif for a global virulence regulator, a model of the *sar* system. The interaction between *S. aureus,* the environment and other competing organisms can be expressed as a population level interaction which relies on a large number of external factors. These interactions can be modeled as fixture which includes a framework for evolutionary interactions and host pathogen interactions. This system of motifs, modules and games[1] provides an overall framework for considering behavior at many different organizational levels. It is of our interest to find correlations between the different known regulatory pathways and the target genes computationally, with a deterministic approach with the purpose of achieving certain stage where its predictability can be measured for the system

under consideration[2].

COMPONENTS OF THE VIRULENCE MODULE

During *S.aureus* infection, genes related to virulence expression or biofilm formation are expressed and regulated with respect to the cell concentration at the site of infection[3]. For instance, at low cell densities, genes for adhesion and for biofilm formation are manifested, whereas at increased cell densities, genes for virulence factors are expressed. The main system responsible for the cell density dependent regulation is the *agr* regulon. As a whole, the virulence module of *S.aureus* contains regulatory sequences, genes encoding α-toxin (*hla*), toxic shock syndrome toxin-1 (*tst*), fibrinogen binding protein A (*fnbA*), the *sar* system with SarA and its homologs and regulation by RNAIII[4]. In this section, the functioning of certain components of this module will be explained in brief to introduce the components that will be explained in brief to introduce the components that will be taken for consideration while expressing these systems in terms of mathematical "motifs".

The *agr* regulon

The *agr* locus of *S.aureus* which is responsible for the accessory gene regulators, represents a Quorum Sensing Circuit[5-7]. The locus driven by 4 genes *agrBDCA* has two divergently transcribed promoter regions. Agr AC represents the TCSTS, AgrC which is the histidine kinase sensor and AgrA being the cognate response regulator. *agrB* and *agrD* produce, modify and export the AIP (Auto Inducer Peptide). At higher cell densities, the AIP binds to the transmembrane AgrC which leads to autophosphorylation of AgrC. The phosphate group is then relocated to the AgrA which initiates transcription of the promoters P2 and P3.

The transcripts of P2 and P3 form the RNAII and RNAIII respectively which help in regulation of wide range of virulence factors. The expression of *agrBDCA* is maintained at a constant basal level to produce the AIP molecules during early exponential phase. Binding of phosphorylated AgrA to the intergenomic region between P2 and P3 promotes transcription of the agr locus. The intergenomic region contains binding sites for SarA. Expression of *agr* is also dependent on glucose concentration and pH.

Sar locus- SarA and homologs

Sar or the Staphylococcal Accessory Regulator can be considered as the global virulence regulator in *S.aureus*[8,9]. SarA, which was first identified in this locus, has been shown to have *agr*-dependent as well as *agr*-independent regulation of virulence factors. SarA activates the *agr* system by binding to the intergenic region between the *agr* P2 and P3 promoters. The SarA up regulates and down regulates various virulence factors. For example, studies of SarA mutants have shown that SarA directly or indirectly transcriptionally activates coagulase, fibrinogen binding protein A, fibronectin binding protein and represses α-toxin, serine protease and lipase production.

In an *agr*-independent manner, it represses extracellular proteases [*ssp,aur,scp*], collagen binding protein (*cna*) and activates α-hemolysin[*hla*], toxic shock syndrome toxin 1(*tst*) and fibrinogen binding protein A (*fnbA*). Homologs of SarA have been identified with conserved regions. The homologs SarS, SarT, SarU, SarX and SarV, along with the *sar* repressor SarR help in regulating the production of virulence factors. This regulation is achieved as these homologs can regulate a particular virulence factor and can also sometimes repress another *sar* homolog.

Regulation by RNAIII

The transcript of the *agr* P3 promoter is RNAIII which has been shown to be the main regulating molecule of virulence factors[10,11]. It has been proposed that RNAIII works as anti-repressor, binding with various Sar homologs to either activate or repress various virulence factors. For example, it has been predicted that RNAIII transcriptionally activates *hla* expression by binding to its promoter. Here *Rot* [repressor of toxin] a protein similar to SarA, represses expression of *hla*. It is hypothesized that RNAIII binds to *Rot* thereby leading to expression of *hla*[11].

Target genes

The actual targets of the series of regulatory networks described above are various proteins that aid in virulence, cell recognition, biofilm establishment, biofilm degradation, and host evasion. Virulence factors are

FULL PAPER

expressed during high cell concentrations whereas cell-cell recognition and adhesion factors that lead to biofilm establishment are expressed at low cell concentrations. Factors that lead to biofilm degradation resulting in release of cells are also regulated. Some virulence factors are α-hemolysin(*hla*), toxic shock syndrome toxin 1 (*tst*), α-toxin, δ-toxin etc. Expression of coagulase, serine protease, fibronectin binding protein, protein A(*spa*), extracellular proteases (*ssp,aur,spa*) are also targeted by the regulatory processes.

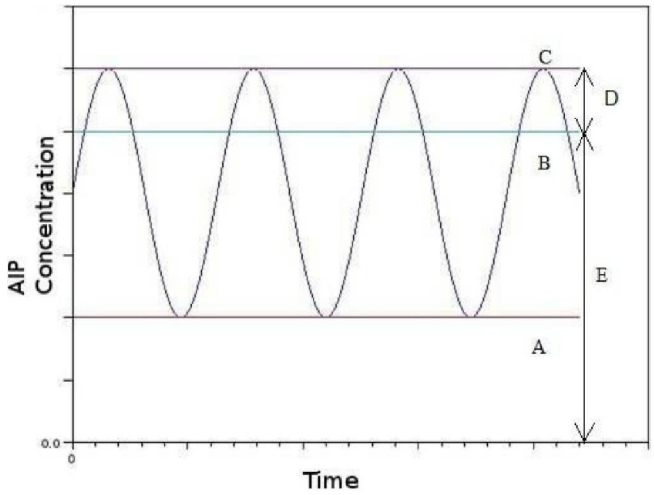
REGULATORY PROCESSES AS MOTIFS

The virulence module as described in the previous section can be represented as a network of functionally distinct elements called motifs. Motifs can be regarded as equivalents of discreet logic components.

In the context of a genetic network with complex interactions between DNA, RNA and protein, the process of definition of a motif can be a hard task due to the very fine distinction in the functionality of each component. In this study, the focus will lie on three largely distinct molecular systems representing all three levels of biological information: the *agr* regulon (DNA) and the function of RNAIII (RNA), and the *sar* locus (protein). The choice of components reflects the basic question about the need for such a model of virulence expression. By comparing the motifs with similar functions in other related bacterial species, we can predict evolutionary relationships, population interactions, possible strategies of survival, etc. In this study, the motifs described will not be analyzed with similar components in other species. Rather, a possible model with respect to its function will be described. By comparing motifs based on the functionality, components of unrelated species having common functions can also be studied. In the description of the motifs, the focus of the model (i.e. the variable under consideration), the equilibrium state (if any) and the dynamic properties (if any) will be included.

Agr regulon - an oscillating bitable switch

The *agr* TCSTS represents the quorum sensing component, which makes the virulence model behave in a cell density dependent fashion. Due to the continuous, dynamic switching between the planktonic state and the biofilm state, we hypothesize that there occurs a periodic oscillation in the products of the virulence modules, as illustrated in Figure 1.



A- Basal expression of AIP. B- Threshold AIP concentration. C- Saturation Concentration. D- Planktonic State. E-Biofilm state.

Figure 1 : Representation of hypothetical oscillations in the states of an S.aureus population assuming infinite resource availability

A differential equation model for such an oscillator can consider the oscillations of different 'states' as a function of variation of concentration of AIP with time. In a simplified model disregarding cellular noise and idealizing the oscillations to be sinusoidal, the system when represented graphically shows the distinct switching between the expression of cell surface proteins (at the lower AIP concentrations) and expression of soluble proteins like virulence factors (the peaks of the sinusoids at high AIP concentrations). This representation completely neglects interference by other environmental signals and fluctuations. A model for the quorum sensing circuit in *P.aeruginosa* has already been proposed by Dockery and Keener[12]. A similar framework can be adopted for the modeling of the *agr* TCSTS.

The graph shows an intermediate concentration of AIP where the switching action takes place. In biological terms, at the critical concentration, transcription of RNAIII is upregulated. As RNAIII increases expression of proteases, lipases, hemolysins and virulence factors and decreases expression of cell surface proteins, fibronectin binding proteins, etc., RNA III acts as the reverse switch, forcing cells back into the planktonic state.

Mathematically, RNAIII is modeled on rate of transcription of P3, which is directly proportional to an exponential order of AIP concentration, above a lower cut off concentration. In function, it behaves as a module wide switch, enforcing expression of factors which force the system into its initial state.

Here, additional regulatory factors like RIP (RNAIII inhibiting peptide) and RAP (RNAIII activating protein) and Rot (Repressor of toxin) will not be included as they do not fall under the control of the *agr* system.

This system can also be expressed in simpler terms as a Boolean network. Consider the class of virulence factors and biofilm degradative enzymes to represent state V. Also consider the class of cell surface proteins and other binding factors to represent state B. Let E be the END state of the population meaning either inactivation or death.

Considering purely the switching action and ignoring transcriptional and translational processes, the relation between the state variable U (either B or V) and R (the RNAIII 'switch') will be:

U= R? V: B

And as R can itself be related to concentration of AIP, we can set the critical concentration of AIP as A. Then

U= {B, V, E, B ″, V ″} // U represents the variable states B, V or E or transitions B″ and V″

X= {0, 1}//The set of environmental factors required for the two states

```
U=B
        A=0 ″
Cyclebegin:
If (U=B)
        { U=B ″
        A=1
        R = A? 1: 0
        U=V
        }
If (U=V)
        { If(X)
                { U=V ″
                A=0
                R=A? 0:1
                U=B
                }
                else
                U=E
        }
```

Thus the action of the *agr* regulon as a continuously oscillating bi-stable 'switch' can be modeled using a system of differential equations, and its working can be represented as a Boolean network.

Sar homologues: cascading switches

The function of the Sar system is a highly regulatory one, and it directly regulates virulence expression in an *agr*-independent as well as dependent manner[13,14]. In an *agr* dependent manner, SarA binds to the area between *agr* promoter P2 and P3 binding site, maintaining the basal level expression of *agr*, and repressing particular virulence factors. In an *agr*-independent system, the Sar homologs regulate the virulence factors and some repress another Sar homolog leading to a complex interconnected network that regulates expression of virulence.

The functioning of such a cascade switch is not unlike the working an integrated circuit. Like the Integrated Circuits (IC), the *sar* locus functions as a network of coordinated switches forming a specific logic network. Depending on the transcription of the locus and the relative concentrations of each homologue, the pattern of virulence expression can be altered.

Thus, it can be thought of as a multi input/output IC, where the input is the specific relative concentration of each homologue, the network logic being determined by specific functioning of each protein, and the output being the regulated expression of virulence genes.

A Boolean network similar to that described for the previous motif can be applied to the Sar system as well, but the target genes are diverse, and the mode of regulation (i.e. whether activation or repression) is far more complicated. In such a system, feedback and feedforward are both crucial and highly sensitive. Modeling of the *sar* locus as a logical switching cascade with distinct high and low states presents a unique challenge. Further complications of the model can be introduced by taking into consideration the action of SarR, environmental fluctuations and interactions with other regulatory functions.

Motif interactions

The functioning of each motif described above can result in very complex expression patterns. The basis for this is the extreme state dependence of each motif.

FULL PAPER

Now, when virulence factors are regulated as a result of the motifs interacting in sync with each other, the resulting levels of regulation involve very high complexity. The intricacies of such an extensive regulatory architecture provide a system with a robust mechanism of virulence mechanism over a very large range of in-vivo environmental conditions. The modeling of motif interactions can be carried out only with a large set of experimentally determined data. Interactions on this scale can be modeled as stochastic processes[15], considering the probabilities of the various states with the input parameters being the initial state of each motif as a function of environmental fluctuations.

Noise generation

Any cellular system is highly prone to noise mainly due to the presence of a complex, diverse and constantly fluctuating environment both within the cell and outside of it. In the case of a regulatory network, the concentrations of various RNAs along with their respective translated proteins can vary drastically due to the combined effects of all other cellular processes that may or may not directly affect the network under consideration. When data sets are analyzed for obtaining variations of levels of factors with time, the inherent error in the data due to noise cannot be neglected. The prediction of this "background noise" in any system can again be modeled stochastically, but the parameters under consideration will be many including temperature, pH, basal transcription rate and translation rate, presence of specific inhibitors and/or activators, interactions with other molecules etc. In our system, the major source of noise can be the random bursts of AIP production which can affect other networks in unpredictable ways. The transcription rate of all genetic factors, including the transcriptional rate of the various Sar homologs and virulence transcripts are under very tight regulation. Unanticipated induction of any of the interconnected paths will push the system into states that might not even be allowed in a given model. Phenomenon like Stochastic Resonance[16,17] has to be studied and incorporated to understand better the features and effects of sub-cellular noise. A model that incorporates a correction factor to reduce the effect of noise can provide a clearer understanding of the working of network, along with reproducible results.

POPULATION WIDE BEHAVIOR-GAMES

It is evident that the behavior of cultured organisms *in vitro* and *in vivo* will be very different based on the complexity and extent of environmental interactions. In a disease causing organism like *S.aureus*, the environment determines the state of the system. Modeling an *in vivo* system must take into account the extreme high levels of extracellular noise, interspecies interactions, host interactions as well as resource limitations.

A model for the regulation of virulence will incorporate the influence of individual motifs combined to form distinct modules. The cross activation of various pathways seen in the commonly occurring *Staphylococcus species* forms the basis for setting the parameters for interspecies interactions. Since cross activation of conserved loci takes place, it can be safe to assume that the homologous module in each species has a similar underlying architecture. Thus an interspecies interaction model can simply find the influence of each module over the other module, under the given constraints considering the following input parameters such as:

- Cell densities of each interacting species.
- Initial states of each motif and module in each species.
- Presence of specific activation or inhibition factors in the environment for each species.
- The species-specific values for parameters of each module (and each motif of each module)

  Including transcriptional and translational rates, etc.

The output parameters should include:

- Final population densities of each species.
  1. Specific concentrations of the products of each motif, with a comparison with the initial conditions.
  2. Finally, a comparison with the initial conditions and a representation of the effect of each species on the other including type and degree of interactions.

Such a model will require a greater insight into the possible interactions in *in vivo* systems, but if such a model is indeed possible, a far greater predictability of population interactions can be achieved.

CONCLUSION

The governance of virulence in *Staphylococcus aureus* is determined by composite interaction of various genetic elements which form a specific genetic network, which can be modeled by using existing computational tools. Such a model can be used not only to predict the molecular genetics level regulation, but can also be used to correlate behavior on the population scale with the molecular behavior on the intracellular level. Interaction between the various infection-causing species *in vitro* can also be simulated. The theoretical predictions of this model can then be verified experimentally, whose results can then be used to strengthen the model further. Further, the scheme to be used, namely the classification of different levels as Motifs (at the basic genetic level), Modules (at the level of interacting gene elements) and Games (at the level of interacting populations) will help in formulating comparative analyses of intracellular pathways in the context of multi-species interactions. The next stage will be to carry out pertinent experiments to validate the utility of this hypothesis, which will in turn lead to a strengthening of the model, lending it a greater predictive capability.

ACKNOWLEDGEMENTS

The authors earnestly appreciate Dr. S. Swaminathan, Dean Sponsored Researcher, SASTRA University for constant encouragement and perpetual support rendered to develop this model. Dr. S. Adline Princy gratefully acknowledges Mr. Amogh Jalihal for his unyielding dedication towards the development of the model. She also wishes to thank the co- workers who have helped in bringing out this paper propitiously.

REFERENCES

[1] D.M.Wolf, A.P.Arkin; Curr.Opin.Microbiol., **6**, 125-134 **(2003)**.
[2] J.M.Bower, H.Bolouri; Computational Modeling of Genetic and Biochemical Networks, MIT Press, London, **(2001)**.
[3] S.Arvidson, K.Tegmark; Intl.J.Med.Microbiology, **291**, 159-170 **(2001)**.
[4] G.Sakoulas; Drug Discov.Today, **3**, 287-294 **(2006)**.
[5] S.P.Brown, R.A.Johnstone; Proc.R.Soc.B, **268**, 961-965 **(2001)**.
[6] C.M.Waters, B.L.Bassler; Annu.Rev.Cell Dev.Biol., **21**, 319-346 **(2005)**.
[7] P.Williams, K.Winzer, W.C.Chan, M.Camara; Phil.Trans.R.Soc.B, **362**, 1119–1134 **(2007)**.
[8] J.H.Heinrichs, M.G.Bayer, A.L.Cheung; J.Bacteriol., **178**, 418-423 **(1995)**.
[9] J.Pratten, S.J.Foster, P.F.Chan, M.Wilson, S.P.Nair; Mic.Inf., **3**, 633-637 **(2001)**.
[10] Y.Gov, A.Bitler, G.Acque, J.V.Torres, N.Balaban; J.Pept.Sci., **22**, 1209-1620 **(2001)**.
[11] Y.Q.Xiong, W.V.Warnel, C.C.Nast, M.R.Yeaman, A.L.Cheung, A.S.Bayer; J.Infect.Dis., **186**, 668–77 **(2002)**.
[12] J.D.Dockery, J.P.Keener; Bull.Math.Biol., **63**, 95-116 **(2001)**.
[13] P.F.Chan, S.J.Foster; J.Bacteriol., **180**, 6232-6241 **(1998)**.
[14] T.R.Rechtin, A.F.Gillaspy, M.A.Schumacher, R.G.Brennan, M.S.Smeltzer, B.K.Hurlburt; Mol.Microbiol, **33**, 307-316 **(1999)**.
[15] J.Paulsson; Phys.Life Rev., **2**, 157–175 **(2005)**.
[16] L.Gammitoni, P.Hanggi, P.Jung, F.Marchesoni; Eur.Phys.J.B., **69**, 1–3 **(2003)**.
[17] P.Hanggi; Chem.Physchem., **3**, 285-290 **(2002)**.