Molecular BioSystems

RSC Publishing

PAPER

View Article Online
View Journal | View Issue

Cite this: *Mol. BioSyst.*, 2013, **9**, 2665

A semi-quantitative model of Quorum-Sensing in *Staphylococcus aureus*, approved by microarray meta-analyses and tested by mutation studies†

Christof Audretsch,[ab] Daniel Lopez,[c] Mugdha Srivastava,[d] Christiane Wolz*[b] and Thomas Dandekar*[ae]

Staphylococcus aureus (SA) causes infections including severe sepsis by antibiotic-resistant strains. It forms biofilms to protect itself from the host and antibiotics. Biofilm and planktonic lifestyle are regulated by a complex *quorum sensing* system (QS) with the central regulator *agr*. To study biofilm formation and QS we set up a Boolean node interaction network (94 nodes, 184 edges) that included different two component systems such as *agr*, *sae* and *arl*. Proteins such as *sar*, *rot* and *sigB* were included. Each gene node represents the resulting activity of its gene products (mRNA and protein). Network consistency was tested according to previous knowledge and the literature. Regulator mutation combinations (*agr*$^-$, *sae*$^-$, *sae*$^-$/*agr*$^-$, *sigB*$^+$, *sigB*$^+$/*sae*$^-$) were tested *in silico* in the model and compared regarding system changes and responses to experimental gene expression data. High connectivity served as a guide to identify master regulators, and their detailed behaviour was studied both *in vitro* and in the model. System analysis showed two stable states, biofilm forming *versus* planktonic, with clearly different sub-networks turned on. Predicted node activity changes from the *in silico* model were in line with microarray gene expression data of different knockout strains. Additional *in silico* predictions about node activity and biofilm formation were compared to new *in vitro* experiments (northern blots and biofilm adherence assays) which confirmed these. Further experiments *in silico* as well as *in vitro* showed the *sae* locus as the central modulator of biofilm production. *Sae* knockout strains showed stronger biofilms. Wild type phenotype was rescued by *sae* complementation. The *in silico* network provides a theoretical model that agrees well with the presented experimental data on how integration of different inputs is achieved in the QS of SA. It faithfully reproduces the behaviour of QS mutants and their biofilm forming ability and allows predictions about mutations and mutation combinations for any node in the network. The model and simulations allow us to study QS and biofilm formation in SA including behaviour of MRSA strains and mutants. The *in vitro* and *in silico* evidence stresses the role of *sae* and *agr* in fine-tuning biofilm repression and/or SA dissemination.

Received 22nd March 2013,
Accepted 29th July 2013

DOI: 10.1039/c3mb70117d

www.rsc.org/molecularbiosystems

Introduction

Staphylococcus aureus (SA) is an important pathogen causing many diseases, especially nosocomial infections. However SA is also part of the natural flora of many people without causing any infections (commensal lifestyle). Diseases caused by SA have a wide range from skin and wound infection over endocarditis and osteomyelitis to potentially fatal systemic disorders, such as the toxic shock syndrome.[1] The high virulent potential of SA is partly due to its *Quorum-Sensing* (QS) ability, because it enables SA to adapt perfectly to its surrounding and coordinates the gene expression within a SA population.[2] QS is a common cell-to-cell communication mechanism in bacteria which enables cells to monitor their population density and activate a concrete pool of

[a] *Department of Bioinformatics, Biocenter, Am Hubland, University of Würzburg, 97074 Würzburg, Germany. E-mail: christofaudretsch@gmx.de, dandekar@biozentrum.uni-wuerzburg.de; Fax: +49 931 888 4552; Tel: +49 931 318 4551*

[b] *Microbiology, Auf der Morgenstelle, University of Tübingen, 79104 Tübingen, Germany. E-mail: christiane.wolz@med.uni-tuebingen.de*

[c] *ZINF, Josef-Schneider Str. 2/D15, University of Würzburg, 97080 Würzburg, Germany. E-mail: daniel.lopez@uni-wuerzburg.de; Fax: +49 931 201 645 001; Tel: +49 931 888 3171*

[d] *Department of Botany, Atarra P.G. College, Bundelkhand University, Bundelkhand, India. E-mail: mugdha.srivastava@gmail.com*

[e] *EMBL Heidelberg, Meyerhofstraße 1, 69117 Heidelberg, Germany*

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c3mb70117d

genes when the bacterial population has reached a certain density threshold.[3,4] SA secretes an autoinducing peptide, the *AIP*, which acts as the signalling peptide of a two component system composed of *agrC* and *agrA* (accessory gene regulator). *AIP* is a product of the *agr*-locus from which two RNAs, *RNAII* and *RNAIII*, are transcribed, driven by two promoters P2 and P3. *RNAII* codes for the proteins *AgrB*, *AgrD*, *AgrC* and *AgrA*. *AgrB* exports and processes *AgrD* to *AIP*; *AIP* accumulates externally and its concentration is proportional to bacterial density. *AIP* in turn activates *AgrC* which is then able to phosphorylate *AgrA* that acts as a transcription factor and hence facilitates transcription of *RNAII* and *RNAIII*. This feedback loop leads to an increase in *AIP* expression with rising density of the colony. *RNAIII* on the other hand simultaneously activates a complex gene regulon that, in turn, is responsible for changing the bacterial lifestyle of the microbial community from a sessile, biofilm producing style to a mobile, planktonic, invasive and more virulent phenotype. Additional regulatory circuits like *SarA* (staphylococcal accessory regulator) and other two-component-systems like *Sae* (SA exoprotein expression)[5] or *SigB* (Sigma factor B) are also involved in this transition.

The above mentioned sessile lifestyle of SA implies the formation of surface-associated microbial communities commonly known as biofilms. In biofilms, the microbial community is encased in a self-produced extracellular matrix that generally protects the cells from external insults, such as antibiotic treatments and the immune system.[6,7] Moreover, biofilms prevent the bacteria from mechanical damage and from being washed away for example by the bloodstream.[8] Consequently biofilm formation is an important stage in the development of an infection because it provides, to the community of infective bacteria, a more resistant niche to spread the infection. SA biofilms require the synthesis of a polysaccharide intercellular adhesin (*PIA*, also known as PNAG for poly-*N*-acetyl-β-(1,6)-glucosamine) but also some extracellular proteins and/or secreted DNA. These substances form an extracellular matrix in which the bacteria are embedded. When the cell density is rising, the *AIP* concentration also rises, concomitant with the increase in bacterial cell density. The bacteria develop an invasive toxic phenotype that is characterized by the reduced expression of surface adhesins such as staphylococcal protein A (*spa*) and fibronectin binding protein (*fnb*) and/or by the up-regulation of the expression of proteases like *splA-F* or nucleases. The nucleases help to dissolve the biofilm and enable the cells to disseminate. Moreover the bacteria start to produce different virulence factors such as γ-haemolysins (*hlg*) that enable them to invade new tissue.[9]

To gain better insight, understanding and an integrated view on how the many proteins, RNAs and other molecules influence the biofilm forming capability and pathogenicity, we model here the whole QS system. One approach is to model the QS mathematically.[10] This is very exact, yet also very complex and not easy to work with. Applying changes in this model is usually a large effort as well. Instead, we use a Boolean network of nodes that activate or inhibit each other which can easily be manipulated and studied. Analysis of the interaction of this network can be done very easily by using SQUAD.[11] We first tested the network against previous knowledge by comparing its node activation pattern to the gene activation pattern obtained from SA microarray experiments. Then we compared *in silico* predictions with *in vitro* results concerning the QS, gene regulation and biofilm building ability of SA by northern blot and biofilm adherence assays. Finally, various mutant outcomes were predicted and compared with experimental results. With this model we would like to present a tool for easy testing of theories about mutant strains and effects of substances by manipulating the nodes of the simulated network. This tool can be used to get a better insight into the SA QS and regulation of SA behaviour.

Methods

Network setup and simulation

For setting up the network, information about different nodes and their interaction was collected from different databases such as KEGG (http://www.genome.jp/kegg/) and STRING (http://string.embl.de) and a first basic network model with many of the nodes was created. However, additional input from the literature and expert knowledge is necessary to retrieve and verify all involved specific genes and their products. Moreover, the work of correctly setting up the network is just the beginning: the interactions in the network, its size, master and hub nodes need to be defined and collected, and conflicts in interactions as well as the relevance of biologically important nodes had to be completed by comparing different databanks and literature resources together with expert curation so that a reliable and useful network is compiled. For setting up the network in CellDesigner we adhered to the following four rules: (1) Every connection we could find concerning the *agr*-locus and the QS system of SA was implemented. (2) When there was conflicting information on one interaction the most frequently found type of connection was chosen; when there was no difference we either excluded this connection or we chose the most coherent one. (3) When we were not sure if the node or the connection was of importance for the QS system of SA we chose to implement it to make sure not to miss any biologically relevant nodes or interactions. (4) When there is a biological fact which cannot just be simplified and adapted to the activating inhibiting network model we prioritised to reproduce the biological fact correctly (*e.g.* AgrC activates AgrA if AIP is there) and not to implement all available connections.

The combination of all these data sources allowed the construction of a node–node interaction network that included all biologically important aspects of a gene, its products and activities as well as its interactions regardless of whether this included RNA or protein expression. One node thus does not represent just a gene or its protein, it rather represents the effect the gene locus has as a whole, including the protein or other effector molecules taking part in the network interaction. At the same time we tried to keep the model as simple as possible. Further activation or inhibition was thus only modelled in a simplified way: in particular, all relevant interactions were considered and included but no biophysical or biochemical details were considered (*e.g.* affinity constants, on and off kinetics).

The nodes themselves were not modelled in their activation strength, except for the input nodes (*SaeS, AIP*, *RsbP* and *ArlS*). The network model was first established as a graph (applying CellDesigner software). The activating as well as inhibiting interactions were taken from the literature following the rules mentioned above. Further activation or inhibition of non-input nodes is a result of the activating and inhibiting inputs simulated, following the kinetics calculated and provided by SQUAD (see ESI† for detailed calculations). In the end we obtained a comprehensive network in which all the nodes that are known to be important for the QS process around *agr* as well as all the known interactions between these nodes are included (see ESI† for a detailed list, data, interactions and references on all nodes of the network). The construction and visualisation of the network was done using a computer program named CellDesigner Version 3.5.1[12] (www.celldesigner.org). With this program interacting networks can be built just by creating a node and connecting it with either inhibiting or activating links to other nodes. The result is a Boolean network with the Boolean states (either 0 = off or 1 = on) for each node. This output of Celldesigner can be stored in the SBML format.

Network dynamics was investigated using SQUAD Version 2.0.[13] First, we, analyzed and calculated the number of stable steady states in this node–node interaction network. To achieve this, SQUAD converts the network (drawn by CellDesigner) into a Boolean graph: the logical connectivity between input interactions was considered, noting combinations of interactions and the resulting logical output as well as source input signals for activation and/or inhibition. This model was next transformed and used for the semi-quantitative simulations. For this, e-functions interpolate between full on and off states of network nodes, considering again the logical connectivity of the network (using catenated e-functions according to logical connectivity; Di Cara *et al.*, 2007[13]). The e-functions approximate the kinetic parameters in a simplified manner, but they allow to closely mirror the logical succession of events as well as dominance of inhibition over activation and signal transfer and succession within the network (semi-quantitative model; see ESI† for details). Specific properties of each network state (mutational background, biofilm or planktonic state) are described in each of the corresponding sections.

To search for stable network states we used SQUAD and its fast heuristic algorithm to sample over the different network states. Interestingly, the network showed only two steady states (see below). Further modelling work examined the qualitative agreement of the simulation with the *in vitro* experiments. This is done in SQUAD by providing it with a pattern of external stimuli considering involved key regulatory nodes. This allowed to determine which activation strength of the node in the *in silico* simulations did fit the *in vitro* experiments best. The up-regulation or activation values were as follows: T1: *SaeS* = 0.35; *AIP* = 0.4; *RsbP* = 0.2; *ArlS* = 0.35 and T3: *SaeS* = 0.35; *AIP* = 0.9; *RsbP* = 0.21; *ArlS* = 0.35. The values for these simulated external stimuli were not obvious, but according to the network topology, the data from the literature and the experimental data, this pattern of activation corresponds best to the observed behaviour and data. The fit was an iterative process where we first incorporated information that is available from the literature, such as that the *AIP* level in T1 should be considerably lower than that in T3. Second we used a basic activation level of 0.25 for nodes known to perceive external stimuli. With this basic activation distribution we ran the simulation in SQUAD and then changed this basic activation level of 0.25 in incremental small steps until we reached the qualitatively best results.

In a Boolean network there are a large number of possible states: in our network with 94 nodes there are 2^{94} states possible considering that each node can be either activated/active or inhibited. However, in most cases they are unstable as the network logic rapidly transforms the state into the next state. Only very few states remain stable and remain in equilibrium. These are the stable states for the node–node interaction network and evolution and selection made sure that these correspond to clear biological functions. To calculate these, one first has to consider that the more precise the simulation should be, the more the increments that are needed between 0 and 1, and thus the amount of possible network states tends to become infinite. From this large amount of system states the interesting ones need to be selected. The interesting states are generally the stable ones to which the network aspires to return to and to which it always comes back after a temporally limited external input. In a stable state in equilibrium the expression of each node is constant. For analysing the existence of steady states SQUAD converts the Boolean network into a dynamic system by replacing the Boolean states (either 0 = off or 1 = on) by a heuristic curve (catenated e-functions according to the network topology, see above and ref. 13) to interpolate all intermediate states for each node. For this procedure SQUAD first uses a fast heuristic search algorithm to identify all stable steady states existing in the discrete node–node interaction network (Fig. 1). Then it calculates the steady states in the continuous dynamical model with activity states between 0 and 1 for each node. All equations needed for this process are generated automatically by SQUAD upon reading the given SBML file. This allows SQUAD to identify steady states in complex networks even in the absence of kinetic parameters. In the case of our network model the SQUAD simulation identified two steady states. These states correspond very well (regarding the activity of individual nodes as well as the resulting network behaviour) to two biological states (planktonic state *versus* biofilm producing state) that are actually observed in SA. SQUAD includes dynamic simulations where external parameters can be changed in an ongoing simulation. The initial activation strength of each node is selected and thus also the influence of each node is reflected, with its inhibiting or activating connections, on other nodes and on the system as a whole. The strength of activating and inhibiting connections corresponds with the activity of the node they originate from. The activity of the other nodes will then be changed during the course of the simulation and according to their activating or inhibiting inputs. Finally a state will be reached in which the activating and inhibiting inputs to

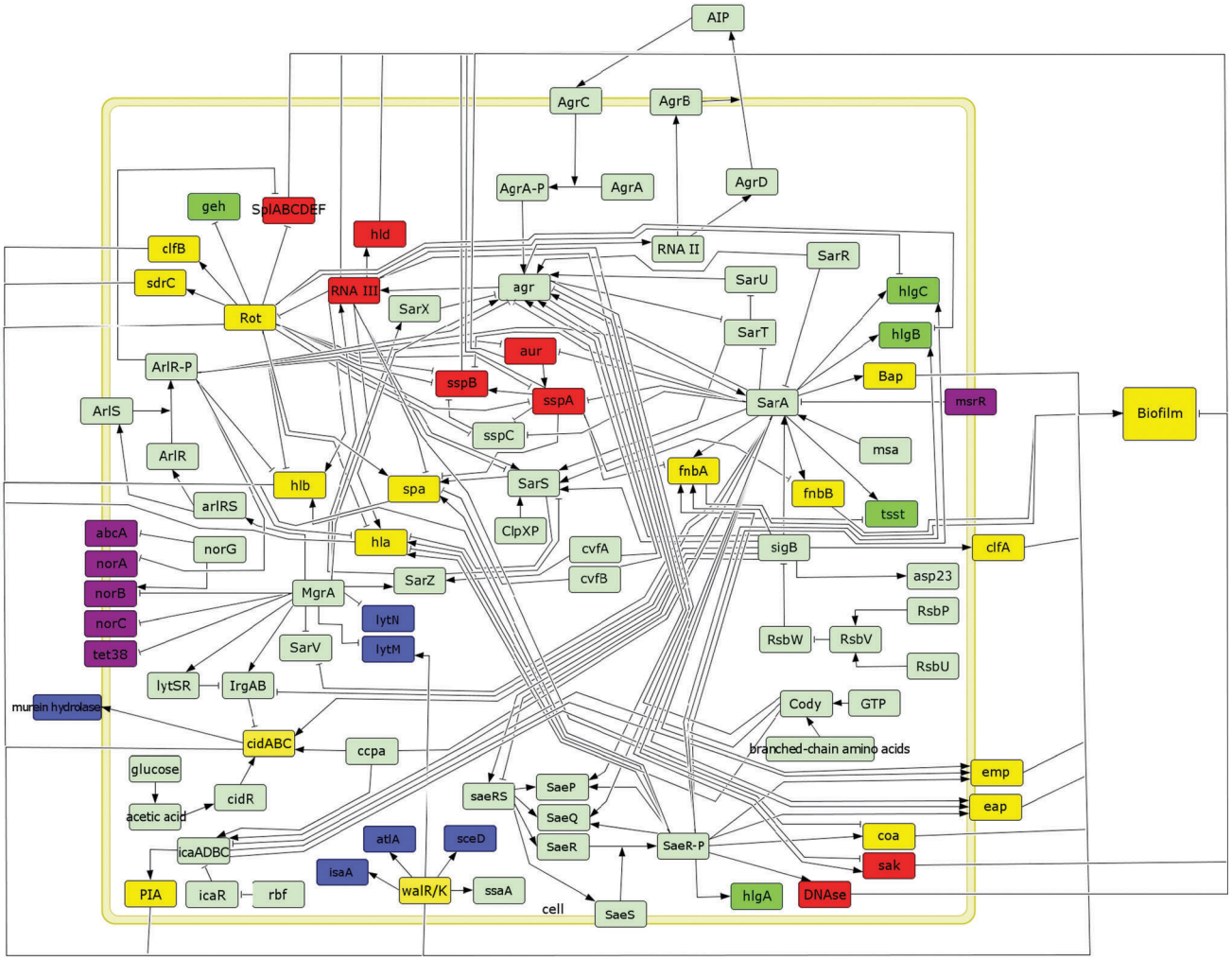

Fig. 1 Node–node interaction network around the *agr* locus: the different nodes and their activating (respectively inhibiting) interactions are shown. The colours of the different nodes show the effect they have on the phenotype of SA. Red = biofilm inhibiting; yellow = biofilm promoting; red and green = important for invasion of new tissue and production of toxins; blue = lysines; purple = multi-drug resistance efflux pumps.

each node are balanced and thus the activity of the nodes is not changing anymore. This is then the steady state. Of course which of the existing steady states the system acquires depends on the initial conditions. The system state can also be transferred from one existing steady state to another by external inputs. This is done by changing the node activity balance in a way that leads, with ongoing simulation, to a balance representing one of the other existing steady states. Furthermore, the activity of some nodes can also be set to a fixed value independent of their simulated inputs for example to represent a knockout mutant strain or any other constant external stimulus to the system. This changes the whole system and may in consequence change the amount and the node activation distribution of the steady states. In this way a stable external milieu and the knockout mutants were simulated in our *in silico* experiments.

The protocol and combination of software (CellDesigner and SQUAD) given here have also been successfully applied to other systems, *e.g.* plant immune defense (Naseem *et al.*, 2013[14]) as well as liver cells, apoptosis and viral infection (Philippi *et al.*, 2009;[11] Schlatter *et al.*, 2011[15]). All simulations were done on a computer with Windows Vista Home Premium SP2 (32 Bit) and Intel core 2 duo CPU with 2.53 GHz and 4 GB RAM.

Transcriptome analysis

As one node represents the resulting overall activity of a gene and its products (see Methods) and this incorporates in large part the gene expression strength, we decided to validate the network by using microarray gene expression data. We compared the *in vitro* changes in gene expression due to environmental or genetic changes to corresponding *in silico* node activation changes.

Therefore we collected microarray data of different *in vitro* scenarios, using the SAMMD database (www.bioinformatics.org/sammd/). We investigated the concordance between *in silico* and *in vitro* in three different scenarios: (A) agr^+ *vs.* agr^-, (B) $sarA^+$ *vs.* $sarA^-$ and (C) $biofilm^+$ *vs.* $biofilm^-$. In scenario A and B, all the values that were at least three times higher in the wild type

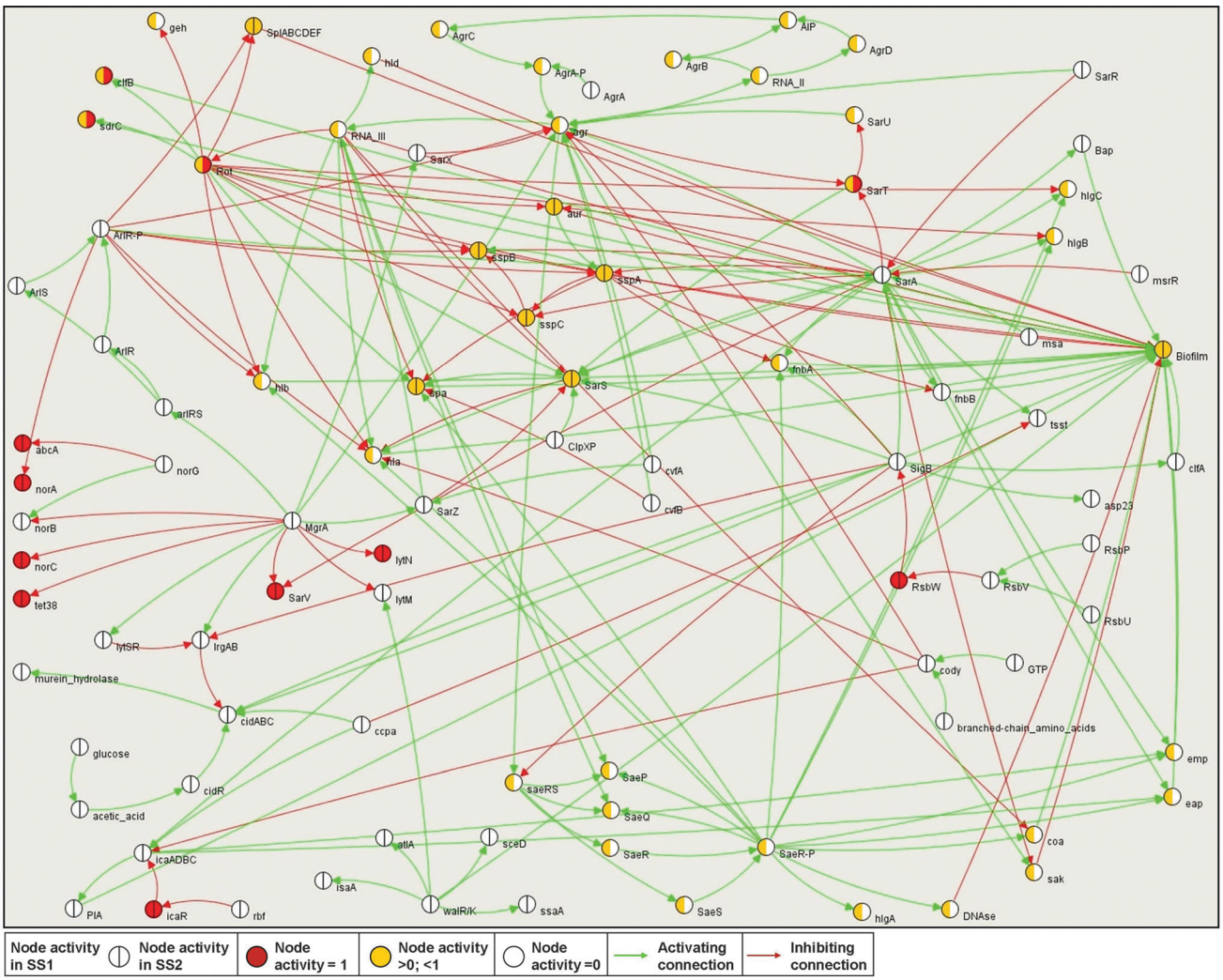

Fig. 2 Steady states of the network: the two different steady states and the corresponding activity of nodes are shown graphically. Here the left halves of the circles representing the different nodes stand for the activity of the particular node in steady state 1. The right halves of the circles represent the activity in steady state 2. The first steady state (SS1) represents a more invasive, toxic phenotype. The second steady state (SS2) represents a biofilm producing phenotype.

than in the mutant strain were noted as up-regulation. On the other hand, when the expression of a gene in the mutant strain was at least three times higher than in the wild type strain, it was noted as down-regulation of this gene. In scenario C, a gene was noted as up-regulated when its expression under the biofilm forming condition was 2.5 times higher than in the not biofilm forming condition. A gene was declared as down-regulated in the biofilm forming condition when it was at least 2.5 times stronger expressed in the not biofilm forming condition compared to the biofilm forming condition.

From the SAMMD database (see above) we gathered data from different experiments according to the three mutation scenarios. For scenario A and B, we used microarray analyses conducted by Cassat *et al.*[16] with UAMS-1 SA strains. This analysis compares the microarray gene expression data of agr^+ strains to the expression data of agr^- strains. This was done in the exponential phase (OD 1.0 at 560 nm) and in the post exponential phase (OD 3.0 at 560 nm). Moreover this microarray analysis compares in the same way $sarA^+$ and $sarA^-$ strains at OD 1 and OD 3. So in the end we collected four different datasets: (A1) agr^+ *vs.* agr^- *OD1*, (A2) agr^+ *vs.* agr^- *OD3*, (B1) $sarA^+$ *vs.* $SarA^-$ *OD1*, (B2) $sarA^+$ *vs.* $SarA^-$ *OD3*. Moreover we used microarray analyses conducted by Dunman *et al.*[17] with RN27 SA strains. This analysis compares the microarray gene expression data of agr^+ strains to the expression data of agr^- strains and the expression data of $sarA^+$ strains to the expression data of $sarA^-$ strains at OD 3. We collected two more datasets from this analysis: (A3) agr^+ *vs.* agr^- *RN27*, (B3) $sarA^+$ *vs.* $SarA^-$ *RN27*.

For scenario C, we collected three different datasets from SAMMD. For dataset C1, we used microarray analyses conducted by Brady *et al.*,[18] where the gene expression of a late exponential phase (6 h) planktonic culture is compared to the gene expression of a maturing (48 h) biofilm culture. For dataset C2, we used microarray analyses conducted by Resch *et al.*,[19] who compared the gene expression of *SA113* grown in a biofilm (for 24 h) with its gene expression in planktonic growth (for 24 h). For dataset C3, we used microarray analyses

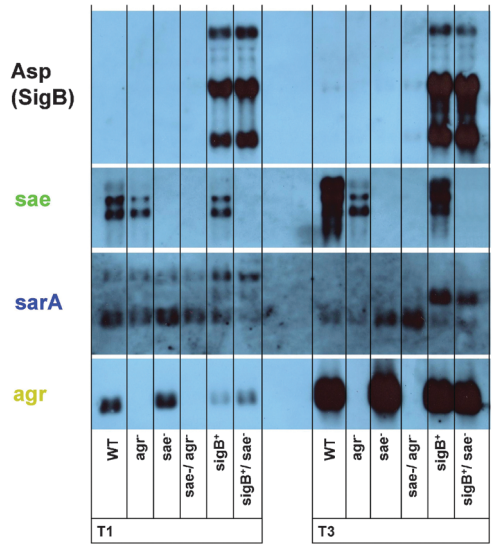

Fig. 3 Strain comparisons in experimental northern blots: northern blots were made with the *WT* strain and the five different strains with mutations in important nodes. In all strains the expression of *agr*, *sarA*, *sae*, and *asp* (*sigB*) mRNA is detected in T1 and T3.

conducted by Beenken *et al.*,[20] where the gene expression of a one week old biofilm forming *UAMS-1* colony grown in a flow cell was compared to the gene expression of a planktonically grown *UAMS-1* colony in the stationary phase (OD 3.5 at 560 nm).

All these *in vitro* datasets were compared to corresponding *in silico* datasets. Datasets from the A and B scenarios were compared to datasets created by comparing the node activities in SQUAD simulations using the settings for the wild type strain as well as for the *sarA*⁻ and the *agr*⁻ strains, under T1 and T3 conditions (OD1 exponential growth *vs.* OD3 stationary phase), respectively. For this the same settings were used as for the northern blot simulations. We calculated four different datasets: *agr* sim T1, *agr* sim T3, *sarA* sim T1 and *sarA* sim T3. The datasets from scenario C were each compared with two different *in silico* datasets created by SQUAD simulations. For the first *in silico* dataset just the activation strength of the nodes in the biofilm forming steady state 2 was compared to the not biofilm forming steady state 1 (sim 1: SS2 *vs.* SS1). For the second *in silico* dataset just the *AIP* concentration was changed to simulate QS influence on the biofilm formation ability. Therefore the activation strength of the nodes under biofilm promoting circumstances with low AIP concentrations was compared to the activation strength of the nodes under biofilm inhibiting circumstances with high AIP concentrations (sim 2: *AIP* low *vs. AIP* high).

To obtain a value for the concordance between *in silico* and *in vitro* in the three different scenarios, we considered and counted all the nodes that qualitatively reacted in the same way in the model and in the experiment (*in vitro* and *in silico* either up-regulated, down-regulated or no change in the gene expression/node activation strength). Then we calculated for each compared dataset the percentage of all compared nodes that by this qualitative measure showed the same reaction. This percentage value was called "*in vitro* to *in silico* consistency".

For further analysis of the reactions of the network model and the similarity of our model in layout and reaction to

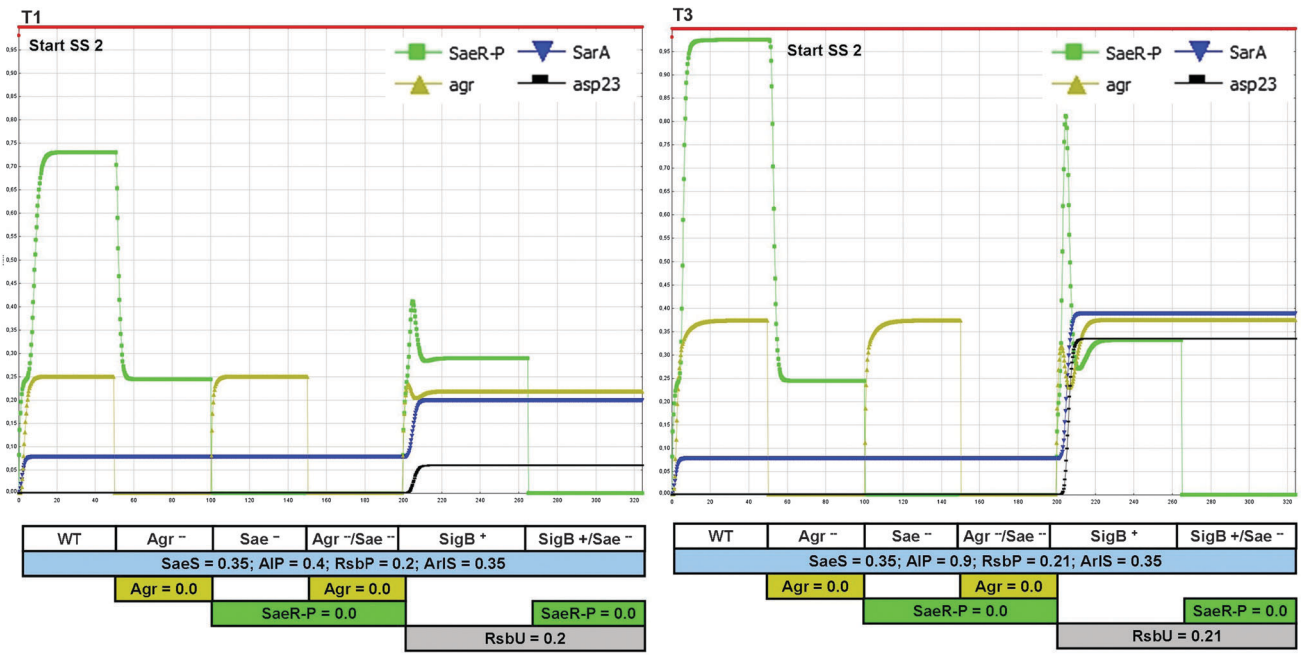

Fig. 4 Strain comparisons in virtual northern blots. Top: *in silico* expression of *sarA*, *agr*, *saeR-P*, the phosphorylated (and thus activated) *saeR* as well as *asp23*, the indicator for the *sigB* expression, in the six tested strains: *WT* [0–50], *agr*⁻ [50–100], *sae*⁻ [100–150], *sae*⁻/*agr*⁻ [150–200], *sigB*⁺ [200–265], *sigB*⁺/*sae*⁻ [265–320]. Data are shown comparing the T1 and T3 growth phase. The SQUAD perturbations, meaning constant- and range pulses applied on the system during the simulation, are shown by bars underneath the graph.

experimental and already published data as well as to make the results more comprehensive, we conducted further studies. First of all we tested which node is affected in which way when *agr* or *saeRS* is knocked out *in silico* and also if these reactions fit what we know up to now about the behaviour of the different nodes from experiments already published. We prepared figures in which the node activity in agr^- was compared to that in agr^+ and the node activity in $saeRS^-$ was compared to that in $saeRS^+$ (see Fig. 6 and 8). In the figures we show nodes that are either down- or up-regulated in the mutant strain. To emphasize that these results are not exclusive to our *in silico* modelling experiments but are already described in the literature we highlighted additionally all such confirmed changes in the figures. Furthermore to get a more detailed insight into how prominent nodes (*ArlR*, *hla*, *icaA-C*, *RNA III*, *Rot*, *SaeR*, *SarA*, *sigB* and *sspA*) are affected in *agr* or *saeRS* knockout strains, we show the *in silico* activation change of these different prominent nodes as a result of knocking out *agr* (Fig. 7) or *saeRS* (Fig. 9). Furthermore we highlighted nodes down-regulated under knockout conditions. To show in which reference a corresponding node behaviour under these conditions is already described we added little flags in different colours to the nodes indicating the different references. A change between the wild type and the mutant was assumed *in silico* and *in vitro* when the expression strength in the mutant was 2.5 fold stronger or weaker than in the wild type. In all *in silico* SQUAD simulation scenarios the different two component systems (TCS) incorporated in the simulation were up-regulated a little bit to simulate an *in vitro* like surrounding where the TCS are putatively stimulated to some extent.

Bacterial strains and growth conditions

The strains are listed in Table 1. Strain ISP546-29 was obtained by transduction using a ϕ11 phage lysate of strain ISP546 and strain ISP479C-29 as recipient. Bacteria were grown on tryptic soy agar supplemented with the appropriate antibiotics for strains carrying resistance genes (kanamycin at 50 μg ml^{-1}, erythromycin at 10 μg ml^{-1}, and tetracycline at 5 μg ml^{-1}). Bacteria from an overnight culture were diluted to an initial optical density at 600 nm (OD600) of 0.05 in fresh medium without antibiotics and grown with shaking (222 rpm) at 37 °C to the mid-exponential (OD600 = 0.5) and post exponential growth (T3: 1.5 h after the probes reached OD600 = 1.5 (T1: OD600 = 0.5)) phase in CYPG (Casamino acids [10 g per liter], yeast extract [10 g per liter], NaCl [5 g per liter], 20% glucose, 1.5 M phosphoglycerate) for northern analysis or TSB with 0.5% glucose for biofilm assays.

Northern hybridisation

Bacteria were harvested by centrifugation and immediately lysed in 1 ml Trizol (Invitrogen/Life Technologies, Grand Island, NY, USA) with 0.5 ml zirconia-silica beads (diameter, 0.1 mm) (Carl Roth, Karlsruhe, Germany) in a high-speed homogenizer (Fastprep, MP Biomedicals, Irvine, California, USA). Subsequently RNA was isolated following the directions in the instructions of the manufacturer of Trizol (Invitrogen). 2 μg of RNA was electrophoresed through a 1% agarose–0.66 M formaldehyde gel. Before transfer the intensities of the 23S and 16S rRNA bands stained by ethidium bromide were checked to be equivalent in all the samples. The gel was then blotted by alkaline transfer (Turbo Blotter, Schleicher and Schuell, Dassel, Germany) onto a positively charged nylon membrane (Biodyne, Pall Corporation, Port Washington, New York, USA). The blot was hybridized with digoxigenin-labeled DNA probes against *asp* (a prototypic sigma factor B regulated gene) *sarA*, *RNAIII* (of the *agr* system) and *sae* following the instructions given by the manufacturer of the DIG detection system (Roche, Mannheim, Germany).

Biofilm adherence assay

A biofilm adherence assay was conducted as described by Christensen *et al.*[21] Bacteria from the exponential growth phase were diluted to an OD of 0.05, and 1 ml per well was used to inoculate a 24-well plate (Greiner Bio-One, Frickenhausen, Germany). Plates were incubated for 24 h at 37 °C and washed three times with 1 ml of PBS (phosphate buffered saline). The biofilm was fixed with 1 ml of 50% methanol for approximately 30 minutes and stained with 200 μl of Carbolgentiana violet for one minute. Excessive dye was removed by washing three times with water.

Results

Setting up the model

For setting up the network we included different two component systems (TCS) proposed to be important in SA for detecting the colony density and the environmental conditions (*e.g. agr*, *sae*)[22–24] as well as different nodes proposed to be important for the quorum and environment sensing network (*e.g. sar*, *sigB*).[22,23] We collected data on these nodes using interaction databases such as STRING, as well as gene expression data and already existing network models from different sources (see references summarized in Table S2, ESI†). All data were compiled and integrated into one complete network with 94 nodes and 184 edges using Celldesigner v.3.5.1. The network includes different TCS and also signalling cascades that connect these TCS and lead to either up-regulation or repression of the biofilm forming capability of the SA colony. The network is an overview of the knowledge we have today about the *agr*-locus and its signalling cascade and its influence on different important nodes in the neighbourhood of it. If our current knowledge about the *agr*-locus and its signalling cascade is correct, the network reactions should be consistent with the accumulated knowledge on alterations in the signalling cascade and resulting changes of the phenotypic output, depending on changes of different nodes in the network. Moreover this network can be used for predictions about alterations in the network signalling cascade, as well as for predictions about changes in the phenotypic reactions of SA colonies, again dependent on changes of different nodes in the network.

The computer program SQUAD[13] samples over all nodes and connected network states to determine which network states are stable. We found that the network can have two different

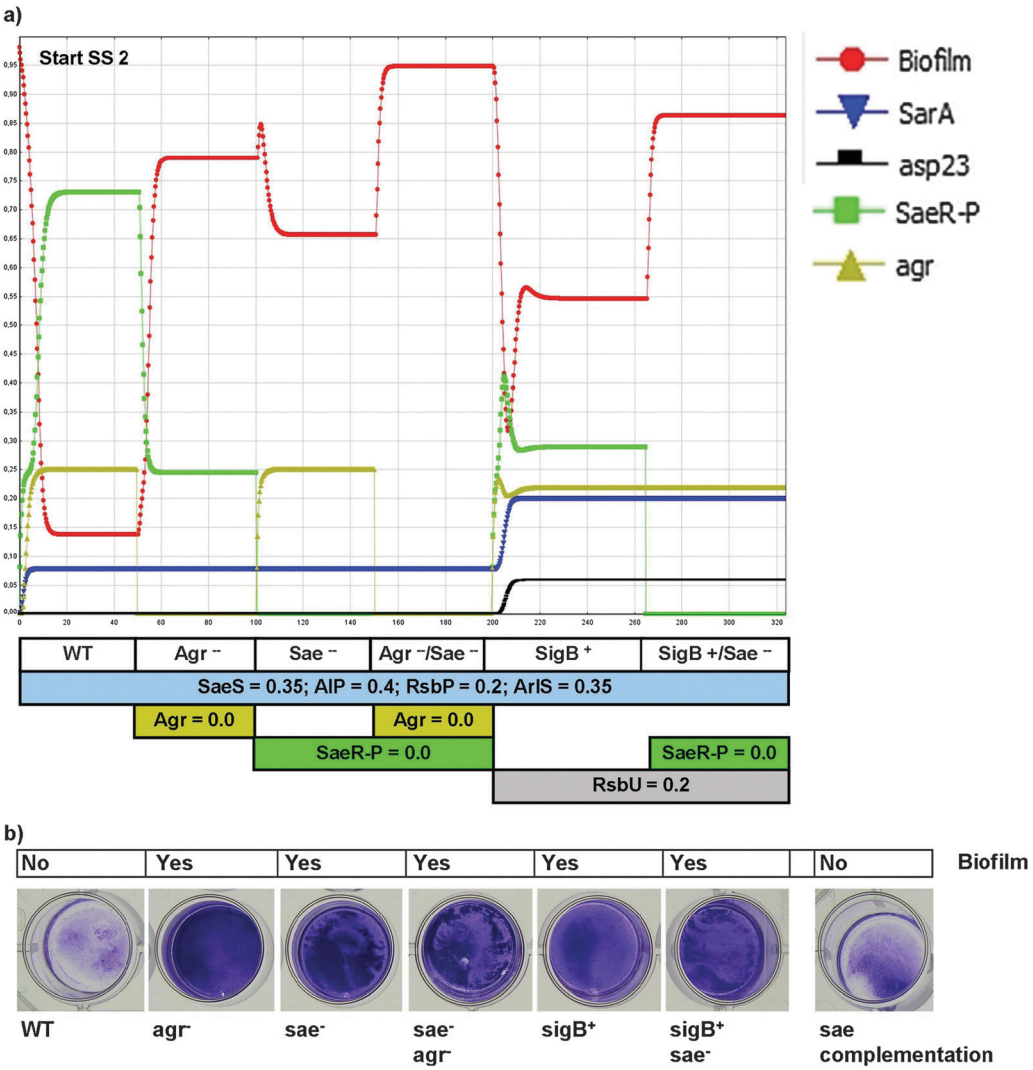

Fig. 5 Biofilm comparison for different strains. (a) *In silico* expression data: compared are *sarA*, *agr*, *saeR-P*, the phosphorylated (and thus activated) *saeR* and *asp23*, the indicator for the *sigB* expression, as well as the biofilm expression in the six tested strains (*WT* [0–50], agr^- [50–100], sae^- [100–150], sae^-/agr^- [150–200], $sigB^+$ [200–265], $sigB^+/sae^-$ [265–320]) at the T1 and T3 growth phase. The SQUAD perturbations, meaning constant- and range pulses applied on the system during the simulation, are shown by bars underneath the graph. (b) *In vitro* biofilm results of the different strains: strain names are as in panel a and are given at the bottom. The biofilm forming ability is qualitatively comparable to the *in silico* experiments. Right: biofilm forming ability of the *sae* complementation. The biofilm intensity is comparable to the *WT*, hence the biofilm up-regulation in sae^- can *de facto* be ascribed to the *sae* mutation.

steady (and stable) states (Table 2): the first steady state (SS1) represents a more invasive, toxic phenotype (no biofilm) in which for example different haemolysins (*hla*, *hlb*, *hld*, *hlgA*, *hlgB*, *hlgC*), proteases (*splA-F*, *aur*), a *DNAse*, and *geh*, a glycerol ester hydrolase, are all up-regulated. These nodes are known for producing toxins which are able to destroy erythrocytes and cleave proteins and DNA as well as esters of membrane-lipids and fat in adipose tissue. This makes these substances very tissue destructive and hence they are important for SA in invading new tissue. The second steady state we detected (SS2) represents a biofilm producing phenotype in which for example *rot* (repressor of toxins), the clumping factor B (*clfB*), and the binding proteins *sdrC* and *spa* are up-regulated. These nodes are known to enhance the biofilm forming ability of the colony[9,22,25] (see Fig. 2; Table 3). A switch between these two steady states, the invasive (SS1) and the biofilm building state (SS2), happens *in silico*, mimicking the situation *in vivo*, simply by up-regulating or down-regulating the activity of the *agr*-locus by increasing or decreasing the *AIP* level, simulating a surrounding with a high and with a low population density of SA, respectively.

In summary, all important nodes for signalling were integrated in the model as well as their type of interaction. System analysis showed that there are only two stable states, biofilm forming *versus* planktonic, with clearly different subnetworks turned on. Validation according to gene expression data confirmed this finding.

Network regulation and validation by transcriptome data

The model was set up with the data mentioned in Table S1 in the ESI.† To test model performance and to validate predictions we compared the model with experimental data by conducting

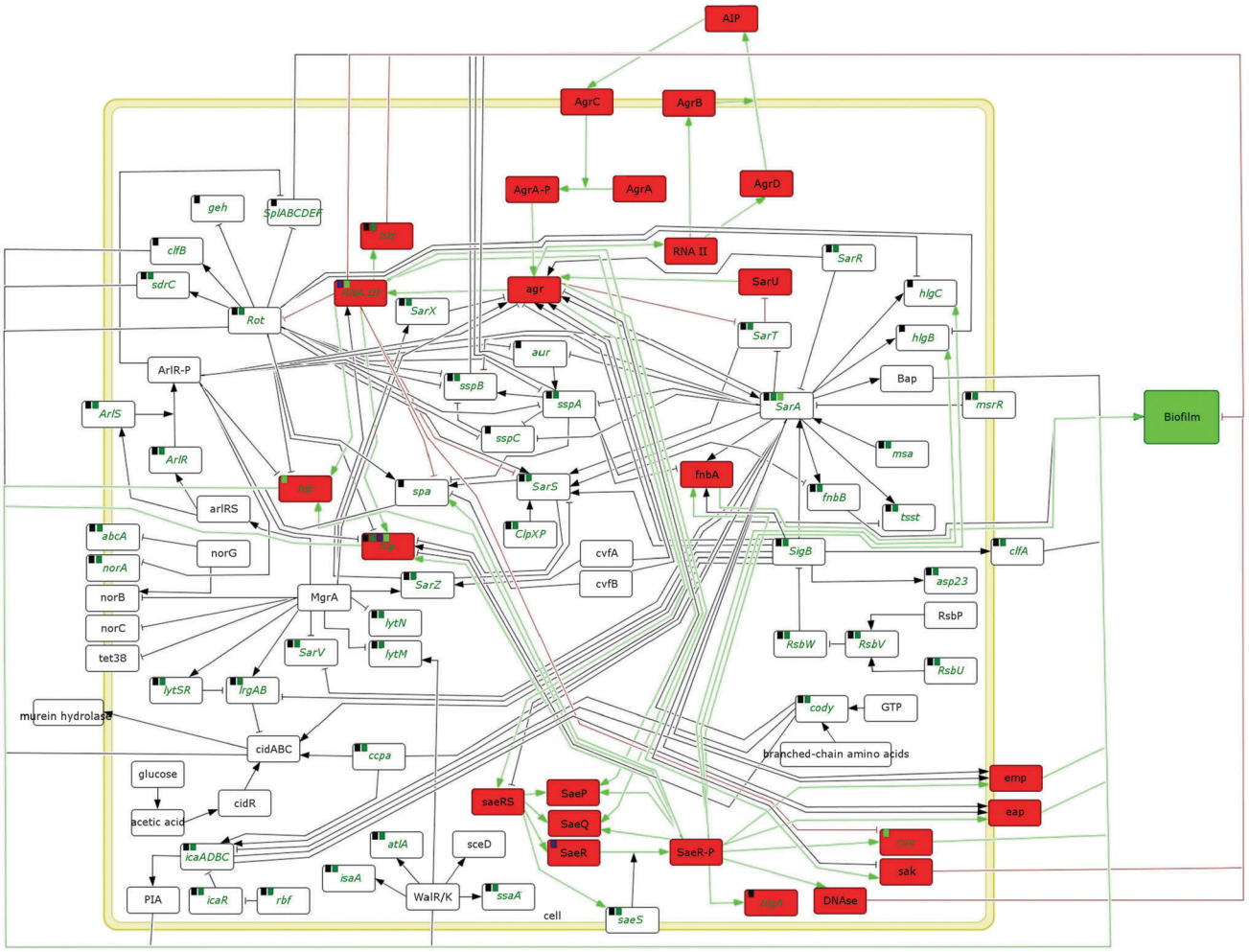

Fig. 6 Impact of *agr*$^-$ on the network: we compared the activity of every node comparing a *wild type* and an *agr*$^-$ mutant strain. A change between *wild type* and mutant was assumed when the expression strength in the mutant was 2.5 fold stronger or weaker than in the *wild type*. Red: all nodes which are down-regulated in the mutant compared to the *wild type*. Green: up-regulated nodes. Unaffected nodes are shown in white. Moreover the figure shows with little flags in different colours in which reference an equal regulation of the corresponding node under the same circumstances can be found (black;[16] dark green;[17] light green;[26] dark blue[27]). In addition the names of all the nodes whose reaction is verified by reference are written in green and bold italics.

an analysis of raw data from published microarray experiments (www.bioinformatics.org/sammd/).[16–20] We investigated the consistency between *in silico* and *in vitro* experiments in an *agr*$^+$ *versus agr*$^-$ genetic background as well as in a *sarA*$^+$ *versus sarA*$^-$ genetic background. The different experimental scenarios were constructed by comparing biofilm *versus* planktonic developmental backgrounds. First of all we compared *in silico agr*$^+$ *versus agr*$^-$ in the exponential and the post exponential growth phase with the *in vitro agr*$^+$ *versus agr*$^-$ of a *UAMS-1* strain either in the exponential phase (OD 1.0 at 560 nm) or in the post exponential phase (OD 3.0 at 560 nm).[16] Here we noted consistencies of 71.43% and 75.71% for the *in vitro* data *versus* the *in silico* predictions regarding the exponential and the post-exponential phase, respectively, *i.e.* the percentage of all the compared nodes that showed qualitatively the same reaction. Furthermore, *in silico agr*$^+$ *versus agr*$^-$ modelling data were also compared to *agr*$^+$ *versus agr*$^-$ gene expression experiments in an *RN27* strain in the stationary growth phase.[17] Here we observed 80.00% consistency between *in vitro* and *in silico* results.

Moreover, we compared *in silico sarA*$^+$ *versus sarA*$^-$ model predictions again in the exponential and the post exponential growth phase with *in vitro* experiments of *sarA*$^+$ *versus sarA*$^-$ mutation in a *UAMS-1* strain in the exponential phase (OD 1.0 at 560 nm) and in the post exponential phase (OD 3.0 at 560 nm).[16] Here we noted consistencies of 71.43% and 75.71% for the *in vitro* data *versus* the *in silico* predictions regarding the exponential and the post-exponential phase, respectively. In addition, the *in silico* predictions for *sarA*$^+$ *versus sarA*$^-$ were also compared to *sarA*$^+$ *versus sarA*$^-$ in an *RN27* strain in the stationary growth phase.[17] Here we found an *in vitro* to *in silico* consistency of 68.57%. Note that we compare here, for each node, the *in vitro* activation change between the particular knockout strain and the *wt* mutant with the *in silico* activation change in the corresponding simulated strain (a statistical analysis of these comparisons is given in ESI†). A more general comparison below shows that regarding system behaviour the consistency is much higher. In a third validation comparison, two *in silico* situations were compared to *in vitro* data from three strains.

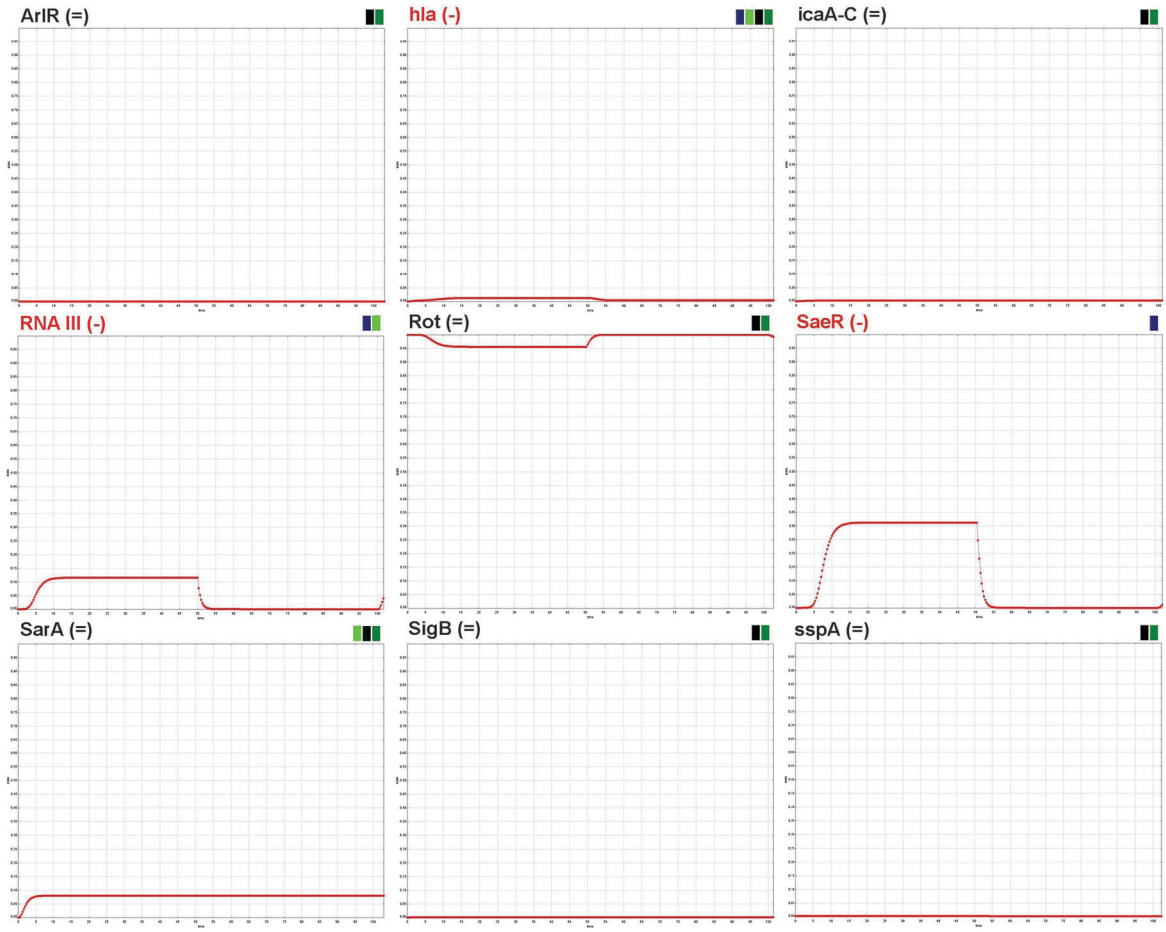

Fig. 7 *In silico* reaction of different nodes after *agr* knockout: here we show the *in silico* reaction of *ArlR*, *hla*, *icaA-C*, *RNA III*, *Rot*, *SaeR*, *SarA*, *SigB* and *sspA* when *agr* is knocked out at 50 s. The *x*-axis reflects the time in seconds and the *y*-axis reflects the activity of the node. The name of the node is written in red when the activity of the node in the mutant was more than 2.5 times lower than in the *wild type* and thus was assumed to be down-regulated *in silico*. The names of unaffected nodes are written in black. Moreover the figure shows with little flags in different colours in which reference an equal regulation of the corresponding node under the same circumstances can be found (black;[16] dark green;[17] light green;[26] dark blue[27]).

The first *in silico* situation was the no biofilm forming system state 2 (SS2) *versus* biofilm forming SS1. This was first compared to a late exponential phase (6 h) planktonic culture *versus* a maturing (48 h) biofilm culture[18] (yielding an *in vitro* to *in silico* consistency of 57.75%). Second we looked at the *SA113* strain, grown as a biofilm (for 24 h) *versus* planktonically grown (for 24 h).[19] Here we noted *in vitro* to *in silico* consistency of 56.34%. Finally, the *in silico* system states (SS1 *vs.* SS2) were compared to a *UAMS-1* colony grown in biofilm (one week old, grown in a flow cell) *versus* a stationary phase (OD 3.5 at 560 nm) planktonically grown *UAMS-1* colony[20] (*in vitro* to *in silico* consistency of 57.75%).

The second scenario compared the *in silico* situation of a biofilm forming phenotype with low *AIP* concentrations *versus* a biofilm negative phenotype induced by high *AIP* concentrations. Thus obtained *in silico* modelling data were compared again to experimental data regarding all three strains with *in vitro* to *in silico* consistencies of 76.06%, 71.83% and 77.46% respectively (see also Table 3 with selected nodes or the full table, Table S3, with all compared nodes in the ESI†).

The consistency of overall systems behaviour is much higher than that obtained when comparing *in silico* simulation with *in vitro* measurements in one scenario: in all scenarios we observed some nodes for a given strain that did not show the same reaction *in vitro* as in the simulation. However, there is no single node that, regarding its system state changes, is not supported by any of the *in vitro* comparisons. Furthermore, we can demarcate the differences in system behaviour to specific nodes and strain specific differences. Concerning the *agr*$^-$ scenarios the only nodes that were inconsistent and different in all three strains were *AgrB*, *fnbA*, *hlb*, *saeR*, *sak* and *sarU*. In contrast, *Aur*, *fnbB*, *isaA-D*, *sspA-C* and *tsst* were the only inconsistent nodes with strain-specific responses in all three *sarA*$^-$ datasets. The only nodes that reacted inconsistently with strain-specific responses in the different strains in all biofilm$^+$ *vs.* biofilm$^-$ datasets were *agrB*, *agrD*, *aur*, *saeR*, *sak*, and *splABDEF*. Moreover one can notice on the basis of the two different *in silico* biofilm$^+$ *vs.* biofim$^-$ datasets and the different *in vitro* sarA$^-$, agrA$^-$ and biofilm datasets that small changes in the experimental conditions may add to the inconsistent behaviour

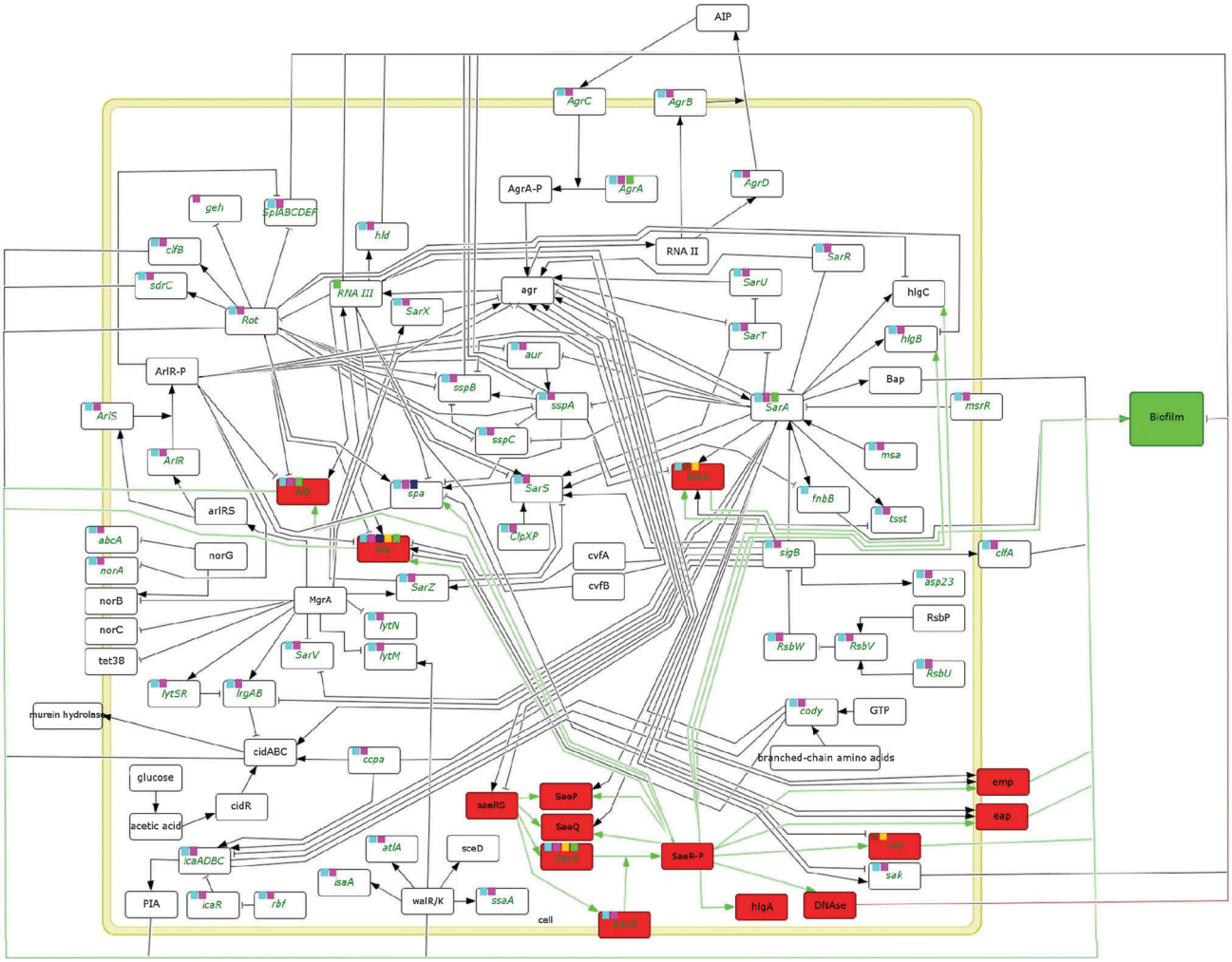

Fig. 8 Impact of *saeRS*$^-$ on the network: for this figure we compared the activity of every node when simulating a *wild type* and an *saeRS*$^-$ mutant strain. A change between *wild type* and mutant was assumed when the expression strength in the mutant was 2.5 fold stronger or weaker than in the *wild type*. In red we show all the nodes which are down-regulated in the mutant compared to the wild type. In green on the other hand we show the up-regulated nodes. In white unaffected nodes are shown. Moreover the figure shows with little flags in different colours in which reference an equal regulation of the corresponding node under the same circumstances can be found (light green;[26] dark blue;[27] yellow;[28] light blue;[29] pink;[30] red[31]). In addition the names of all the nodes whose reaction is verified by reference are written in green and bold italics.

of some nodes. Hence with more knowledge about the actual scenario (*e.g.* growth conditions, nutrient availability, pH) regarding the experimental testing of a specific strain an even higher consistency with the simulation could be achieved, given that the simulation has the ability (see Methods) to incorporate the variables.

Table 3 and Table S3 (ESI†) show the *in silico versus in vitro* correlation of wild type *vs.* *agrA*$^-$ and wild type *vs.* *sarA*$^-$. Furthermore, comparison of biofilm$^+$ *vs.* biofilm$^-$ graphics shows the expression of most of the nodes under *saeRS*$^-$ and *agr*$^-$ conditions according to the model. Small coloured flags show support by observed experimental data from previous publications (black,[16] dark green,[17] light green,[26] dark blue,[27] yellow,[28] light blue,[29] pink,[30] red[31]). For different nodes figures were also created where either *saeRS* (see Fig. 9) or *agr* (see Fig. 7) was knocked out in perturbation experiments. Little flags show support for the behaviour of different nodes from our simulation by experimental observations from previous publications (black,[16] dark green,[17] light green,[26] dark blue,[27] yellow,[28] light blue,[29] pink,[30] red[31]). Statistical analysis of the significance values for the different predictions made is given in ESI† (Table S1); all prediction scenarios show significant results and the detailed values are given.

These data support that the model faithfully reproduces the behaviour of QS signalling mutants. The integrated model comprising 94 nodes and 184 edges allows prediction of various other network mutations and is supported by experimental data from different strains. Furthermore, the well connected hub nodes show how integration of different inputs is achieved by the QS network.

Mutation tests of central network nodes and effects

To test for the robustness of our network we tested whether it could be used for predictions about alterations in the signalling

Fig. 9 Reaction of different nodes when knocking out *saeRS*: here we show the *in silico* reaction of *ArlR*, *hla*, *icaA-C*, *RNA III*, *Rot*, *SaeR*, *SarA*, *SigB* and *sspA* when *SaeRS* is knocked out. The *x*-axis reflects the time and the *y*-axis reflects the activity of the node. The name of the node is written in red when the activity of the node in the mutant was more than 2.5 times lower than in the *wild type* and thus was assumed to be down-regulated *in silico*. The names of unaffected nodes are written in black. Moreover the figure shows with little flags in different colours in which reference an equal regulation of the corresponding node under the same circumstances can be found (light green;[26] dark blue;[27] yellow;[28] light blue;[29] pink;[30] red[31]).

Table 1 Bacterial strains

Strain no./type	Strain[a]	Description	Ref.
1. WT	ISP479C	Derivative of 8325-4	47
2. agr^-	ALC14	ISP479C Δagr::tetM	48
3. sae^-	ISP479C-29	ISP479C ΔsaePQRS::kan	35
3b. sae^- rescue	ISP479C-29(b)	ISP479C-29 ΔsaePQRS::kan, with sae restored by plasmid pCWSAE28 (pCL84 with saePQRS L [nucleotides 1 to 3515], subclone from pCWSAE32)	This study
4. sae^-/agr^-	ISP546 (Tn551e)	ISP479C Δagr::Tn551; ΔsaePQRS::kan	This study
5. $sigB^+$	ISP479r	ISP479C with rsbU restored	49
6. $sigB^+/sae^-$	ISP479r(h)	ISP479r ΔsaePQRS::kan, with rsbU restored	This study

[a] These SA strains were used for northern blot analysis and biofilm formation experiments.

cascades and for predictions about changes in the phenotypic reactions of SA colonies, dependent on changes of different nodes in the network. We decided to use the network to make predictions about the results of northern blot experiments as one node represents not just the protein activity, but much more than that. It reflects the activity of the named locus as a whole which is obviously also strongly affected by the level of gene expression. We used five different knockout mutants (agr^-, sae^-, sae^-/agr^-, $sigB^+$, $sigB^+/sae^-$) and investigated the expression of different prominent nodes (*asp* (*sigB*), *sae*, *sarA*, *agr*) by conducting a northern blot analysis (see Fig. 3). We selected these nodes for the knockout mutants because of their proposed importance for QS[22–24] and considered them as important nodes for the whole network we set up. Literature data point out key regulatory nodes for QS. However, it turns out that network connectivity is a strong predictor to identify the regulatory master nodes just by their high connectivity index.

Table 2 Activity of different nodes in the two steady states[a] (SS1 and SS2)

Node	Activation in SS1	Activation in SS2
abcA	1.000000000	1.000000000
agr	0.929121724	0.000000000
AgrA-P	0.929896264	0.000000000
AgrB	0.999898610	0.000000000
AgrC	0.999999923	0.000000000
AgrD	0.999898610	0.000000000
AIP	0.999997722	0.000000000
aur	0.999567678	0.070103715
biofilm	0.000103676	0.913598828
clfB	0.000582424	1.000000000
coa	0.004164738	0.000000000
DNAse	0.999999923	0.000000000
eap	0.845758097	0.000000000
emp	0.845758097	0.000000000
fnbA	0.000008953	0.000000000
geh	0.999417576	0.000000000
hla	0.981777638	0.000000000
hlb	0.998347767	0.000000000
hld	0.995835256	0.000000000
hlgA	0.999999923	0.000000000
hlgB	0.925356954	0.000000000
hlgC	0.925356954	0.000000000
icaR	1.000000000	1.000000000
lytN	1.000000000	1.000000000
norA	1.000000000	1.000000000
norC	1.000000000	1.000000000
RNA_II	0.997017139	0.000000000
RNA_III	0.907882820	0.000000000
Rot	0.004164744	1.000000000
RsbW	1.000000000	1.000000000
SaeP	0.999430482	0.000000000
SaeQ	0.999430482	0.000000000
SaeR	0.999898610	0.000000000
SaeR-P	0.999997722	0.000000000
saeRS	0.997017139	0.000000000
SaeS	0.999898610	0.000000000
Sak	0.997017139	0.000000000
SarS	0.002366794	0.958714341
SarT	0.092117180	1.000000000
SarU	0.971181167	0.000000000
SarV	1.000000000	1.000000000
sdrC	0.000582424	1.000000000
Spa	0.008757900	0.984250423
SplABCDEF	0.999567678	0.070103715
sspA	0.999601718	0.003676761
sspB	0.970896564	0.000165225
sspC	0.152550870	0.152590863

[a] In this table the activity of the different nodes of our network in the two different steady states (steady state 1 [SS1] and steady state 2 [SS2]) is shown. Here only nodes are shown in which the activity in one of the steady states differs from zero. In the result section there is a detailed description of the two steady states.

The *agr*-locus has 24 connections to other nodes in the network. This means that *agr* together with *sarA* is the node with the most connections. Through these connections for example the *agr*-locus up-regulates itself in a positive feedback loop.[23] *Sae*[23] and different haemolysins (*hla*, *hlb*, *hld*) are also up-regulated.[22,32] The *agr*-locus on the other hand down-regulates for example *rot*.[22] When the *agr*-locus is knocked out, an enhanced biofilm forming ability results.[9]

Like the *agr*-locus, *SarA* has 24 connections to other nodes in our network. This also reflects the importance of this node. Nodes that are up-regulated by *SarA* are for example the *agr*-locus[32] and the different haemolysin toxins whose expression is positively controlled in an *agr*-dependent manner (*hla*, *hlgB*, *hlgC*).[22] Furthermore, *SarA* up-regulates the toxic shock syndrome toxin (*TSST*)[22] and the intracellular adhesion proteins A–C (*icaA, icaB, icaC*) that are responsible for the production of the extracellular polysaccharide matrix of the biofilm in which cells are embedded.[33] The serine proteases *sspA*, *sspB*, and *sspC* are down-regulated by *SarA*.[34]

Rot, the abbreviation for repressor of toxins, is a node with sixteen interactions with other nodes in our network and thus has rank two when listing the nodes by the amount of interactions with other nodes. Many haemolysins (*hlgb*, *hlgc*, *hla*, *hlb*)[22] and the serine proteases *sspA*, *sspB* and *sspC*[34] are down-regulated by *rot*. The biofilm forming ability on the other hand is up-regulated by rot.[22]

The *sae*-locus has many up-regulating connections. For example many haemolysins (*hla*, *hlb*, *hlgA*, *hlgB*, *hlgC*)[24,33] and the fibronectin binding protein A (*fnbA*) are up-regulated.[35] Altogether the *sae*-locus has twelve connections and is thus the locus with the third most connections in our network.

SigB, an alternative sigma factor of the SA RNA polymerase, is known to be of great importance for the stress response of SA.[36] In our network *sigB* has ten connections to other nodes and is thus the node with the fourth most connections. *SigB* for example up-regulates *sarA*[37] and the murein hydrolase activators *cidA*, *cidB* and *cidC*,[38] which are also known to contribute to biofilm formation.[39]

With nine connections to other nodes the *arl*-locus has rank five when estimating the importance of nodes by the amount of interactions. The *arl*-locus contributes for example like many other nodes to up-regulation of *sarA*.[32] Down-regulated by the *arl*-locus are for example again many haemolysins (*hla*, *hlb*) as well as the serine proteases *splA-F*[32] and the *agr*-locus.[32]

The expression data obtained from the northern blots were compared to the corresponding simulated knockout mutants created by down-regulating the knocked out nodes in SQUAD, leading as in the *in vitro* knockout mutants to a near to zero activity of the node. Moreover the different TCS incorporated in the simulation were up-regulated a little bit (see Fig. 4) to simulate an *in vitro* like surrounding, where all TCS are putatively stimulated to some extent. The activity of these prominent nodes was qualitatively comparable in the *in vitro* mutants and in the *in silico* simulations.

To test more directly whether the biofilm forming capability of these mutants is affected in the same way as the biofilm forming capability of the simulated mutants, a biofilm adherence assay was conducted (see Fig. 5). All simulated mutants with biofilm intensity above 0.5 were rated as biofilm forming positive phenotypes. Similar to the RNA expression data from the northern blot the biofilm forming capability was qualitatively affected in the same manner in both scenarios. All mutants were able to form a strong biofilm but in the wild type strain the biofilm forming capability was weakened. The high biofilm forming capability of the sae^- mutant strain which can be seen in the simulation as well as in the biofilm adherence experiments was very unexpected and according to our knowledge this has not been shown elsewhere. To confirm that this result is not an artefact of the sae^- mutation construction, biofilm adherence assays

Table 3 Microarray meta-analysis

	In vitro results	*In silico* results	No. of nodes compared	No. of concordant nodes	No. of non-concordant nodes	Consistency between *in vitro* and *in silico* results (%)
agrA$^+$ *vs.* *agrA*$^-$	agr$^+$ *vs.* agr$^-$ OD1	*agrA* up/down simulation T1	70	57 *e.g.*: arlR; Rot; SaeS; SarA; SigB	13	81.43
	agr$^+$ *vs.* agr$^-$ OD3	agrA up/down simulation T3	70	50 *e.g.*: arlR; Rot; SaeS; SarA; SigB	20	71.43
	agr$^+$ *vs.* agr$^-$ RN27	agrA up/down simulation T3	70	56 *e.g.*: arlR; Rot; SaeS; SarA; SigB	14	80.00
sarA$^+$ *vs.* *sarA*$^-$	sarA$^+$ *vs.* SarA$^-$ OD1	sarA up/down simulation T1	70	50 *e.g.*: AgrA; arlR; Rot; SaeS; SigB	20	71.43
	sarA$^+$ *vs.* SarA$^-$ OD3	sarA up/down simulation T3	70	53 *e.g.*: AgrA; arlR; Rot; SaeS; SigB	17	75.71
	sarA$^+$ *vs.* SarA$^-$ RN27	sarA up/down simulation T3	70	48 *e.g.*: arlR; Rot; SaeR; SaeS; SigB	22	68.57

				No. of nodes compared		No. of concordant nodes		No. of non-concordant nodes		Consistency between *in vitro* and *in silico* results (%)	
	In vitro results	*In silico* results		Sim 1 SS2 *vs.* SS1	Sim 2 *AIP* low *vs. AIP* high	Sim 1 SS2 *vs.* SS1	Sim 2 *AIP* low *vs. AIP* high	Sim 1 SS2 *vs.* SS1	Sim 2 *AIP* low *vs. AIP* high	Sim 1 SS2 *vs.* SS1	Sim 2 *AIP* low *vs. AIP* high
Biofilm *vs.* planktonic	Biofilm *vs.* planktonic (maturing)	Biofilm *vs.* planktonic (SS) sim 1	Biofilm *vs.* planktonic (AIP) sim 2	71	71	41 *e.g.*: *AgrA*; *arlR*; *SaeS*; *SigB*	54 *e.g.*: *AgrA*; *Rot*; *SaeS*; *SigB*	30	17	57.75	76.05
	Biofilm *vs.* planktonic 24 h	Biofilm *vs.* planktonic (SS) sim 1	Biofilm *vs.* planktonic (AIP) sim 2	71	71	40 *e.g.*: AgrA; arlR; SaeS; SigB	51 *e.g.*: AgrA; Rot; SaeS; SigB	31	20	56.34	71.83
	Biofilm *vs.* planktonic OD 3.5	Biofilm *vs.* planktonic (SS) sim 1	Biofilm *vs.* planktonic (AIP) sim 2	71	71	41 *e.g.*: AgrA; arlR; SaeS; SigB	55 *e.g.*: AgrA; Rot; SaeS; SigB	30	16	57.75	77.46

We collected microarray data of different *in vitro* scenarios, using the SAMMD database (www.bioinformatics.org/sammd/). We investigated the concordance between *in silico* and *in vitro* in three different scenarios: (A) *agr*$^+$ *vs.* *agr*$^-$, (B) *sarA*$^+$ *vs.* *sarA*$^-$ and (C) biofilm$^+$ *vs.* biofilm$^-$. In addition the same *in silico* comparisons are given (*AgrA*$^+$ *vs.* *AgrA*$^-$, *SarA*$^+$ *vs.* *SarA*$^-$, biofilm forming *vs.* non biofilm forming scenarios). Detailed information on these three scenarios is listed in Table S3 in the ESI.

with the *sae*$^-$ complementation strain were conducted and the biofilm forming capability was impaired again, suggesting that *sae* plays an important role in biofilm regulation. More precisely, *sae* plays an important role in preventing SA from forming biofilms in inappropriate situations and may prove to be important together with *agr* for the dissemination of the bacteria from the existing biofilm.

In conclusion, five key regulatory mutation combinations (*agr*$^-$, *sae*$^-$, *sae*$^-$/*agr*$^-$, *sigB*$^+$, *sigB*$^+$/*sae*$^-$) were directly tested in the model and also in experiments. High connectivity is a good guide to identify master regulators and their detailed behaviour was studied both *in vitro* and in the model. Together, both lines of evidence support in particular a refined regulatory role for *sae* regarding involvement in biofilm repression and/or SA dissemination.

Discussion

QS is a fascinating feature of bacterial adaptation and has been well studied, including first modelling and simulation efforts.[10,40,41] QS can be seen as a very basic model for swarm intelligence (emergent collective intelligence of groups of simple agents[42]) and as a model for basic decision making processes where the regulatory networks of the individual bacterial cell as modelled here is the basic unit for the emergent collective behaviour. It is a main regulator for SA to change between planktonic and biofilm state of living.[5] As this change is accompanied by enormous changes in pathogenicity and resistance to antibiotics and the host's immune system, controlling this state change has the potential to be used for treatment of SA infections.[9] Realizing the strong influence of QS in the infective behavior of many bacterial species, there is now a promising ongoing effort to develop new small molecules that target QS. Anti-biofilm agents such as QS-supportive drugs may be beneficial in order to facilitate bacterial clearance by the immune system and/or antibiotics. Anti-QS drugs known as quorum quenchers,[43] on the other hand, could prevent SA from changing from the relatively harmless, biofilm forming to the invasive, toxic phenotype and hence prevent SA from invading new healthy tissue and expressing toxic factors such as the toxic shock syndrome toxin (*TSST*). Potentially all this could be achieved by manipulating for example the QS of SA. Most importantly, all the factors leading to biofilm or the expression of virulence factors are potential target sites for new anti-staphylococcal agents.

On the other hand controlling biofilm building behaviour and QS in bacteria has widespread scientific implications,

such as building of microstructures or exploiting the decision making ability of the QS process to implement it in biological computing systems such as logic operators.

In this concrete example the QS network we have built is a model which focuses on several key regulatory nodes such as the two component systems *agr*, *arl* or *sae* and other different important nodes and signalling cascades such as *SigB*, *Rot* or *Sar*. The model provides a lot of detail for example when looking closely at the modeled regulation of the *agr*-locus itself or at all the nodes and their interactions included in the periphery. This model is of course still simplified regarding the exact interaction of the different nodes on a molecular level. For example, we assumed no difference in whether an inhibiting interaction influences the transcription, the translation or the efficiency of the final gene product. We considered only whether the node is impaired, regardless of how it is affected. However, an advantage is that the model works independent of detailed kinetic data. Our model not only provides a basis to simulate QS but it also, in combination with our own experiments and the comparison to published data, helps to better elucidate the function and interactions around the central nodes of regulation. The semi-quantitative model established predicts system changes and their succession. These predictions can be compared to experimental data. In particular, gene expression data (array experiments, northern blot) also monitor system changes. They focus on transcription changes and quantitative gene expression differences. Thus, for comparison of model predictions to gene expression data, statistical analysis has to first filter out the noise from experimental data. Furthermore, gene expression changes monitor only indirectly changes in protein activity which often reflects the node activity. However, they allow to monitor all occurring system changes on the level of gene expression for the modeled QS network. Because the gene expression obviously is a strong impact factor for the protein expression and thus also for the protein activity, it can be used as a surrogate parameter for the node activity that includes gene expression, protein activity as well as all other parameters possibly influencing it. Hence, these gene expression changes can be used for a comparison of the predicted system changes for the node interaction network comparing the simulation as well as the transcription data. In cases where the effect of the node is conveyed by the protein activity the gene expression changes give only a lower limit for the accuracy of the network predictions as the latter pertain to the activity of the proteins, but the actual protein activity is hard to assess and obviously cannot be determined for nodes like RNAII, GTP or Biofilm. Moreover the results show that network activity changes were in most cases accompanied by similar changes in gene expression and transcription. Furthermore it was successfully applied to make predictions about the outcome of northern blots and biofilm adherence assays. The network can now be used for quick and easy testing of predictions around the *agr* and the *sae* locus as well as the QS of SA. Moreover the interaction of different nodes can be seen very easily and thus this network can be used for planning which node needs to be knocked out or altered by drugs to lead to a specific result. Due to its modular construction and its easy to handle freeware this network can be adapted according to new knowledge, modified for different purposes or just be extended.

Many of the nodes embedded in this SA QS network are also known to play an important role in other staphylococci.[44] The *agr* locus for example is not only existent in SA, but it is a virulence factor in the QS of *Staphylococcus epidermidis* (SE). In SE the gene structure and sequence of the *agr* system are very similar to that of SA and may therefore play a comparable role as in SA.[45] The *Sae* two component system for example also plays an important role in *Staphylococcus carnosus*[44] and SE.[46]

Conclusion

We created, to our knowledge, the first Boolean network around the *agr*-locus including many different two component systems such as *Arl* and *sae* and other different important nodes and signalling cascades such as *SigB*, *Rot* or *Sar*. This network has two different steady states: one representing an invasive, toxic phenotype and the other representing a biofilm producing phenotype. This network was validated by comparing it with northern blot and microarray data. Further predictions were made about the QS, the reaction of different nodes and the biofilm building ability of different mutant strains. These predictions were compared to *in vitro* experiments such as northern blots and biofilm adherence assays. These experiments confirmed our predictions. Beyond that, to our knowledge for the first time, we were able to show *in vitro* and *in silico* that *sae* has a strong influence on the biofilm building ability of SA. When *sae* was knocked out *in vitro* or *in silico* the biofilm building ability of SA was increased. By complementation experiments the influence of *sae* on the biofilm building ability was confirmed.

The established network and simulation model allows to study QS and biofilm formation in SA and is made publicly available. The *in silico* network shown in the results qualitatively agrees well with the transcriptome data and additional experiments such as different knockout mutants. The network is fast and easily simulated and compared to experimental data such as gene expression data and experiments. In addition we obtain predictions that are also well supported by the experimental data of this paper. An important and specific regulatory role of *agr*, *Arl* and *sae* is delineated in the model with good support from the experimental data.

As biofilms have a strong effect on survival and virulence of bacteria[8] the newly found impact of *sae* could be a first step for inventing new drugs helping to fight SA infections and preventing SA from colonising catheters where they often form biofilms[9] in which they are better protected from drugs such as antibiotics and thus harder to eradicate.[9]

Funding

We gratefully acknowledge funding by DFG (collaborative grant TR34-A8 and Tr34-B1).

Acknowledgements

Stylistic and language corrections by Dr Rapp-Galmiche are gratefully acknowledged.

References

1 A. L. Cheung, A. S. Bayer, G. Zhang, H. Gresham and Y. Xiong, *FEMS Immunol. Med. Microbiol.*, 2004, **40**, 1–9.
2 J. M. Yarwood, D. J. Bartels, E. M. Volper and E. P. Greenberg, *J. Bacteriol.*, 2004, **186**(6), 1838–1850.
3 S. A. West, K. Winzer, A. Gardner and S. P. Diggle, *Trends Microbiol.*, 2012, **20**(12), 586–594.
4 M. B. Miller and B. L. Bassler, *Annu. Rev. Microbiol.*, 2001, **55**, 165–199.
5 J. M. Yarwood and P. M. Schlievert, *J. Clin. Invest.*, 2003, **112**, 1620–1625.
6 F. Götz, *Mol. Microbiol.*, 2002, **43**(6), 1367–1378.
7 H. Akiyama, M. Ueda, H. Kanzaki, J. Tada and J. Arata, *J. Dermatol. Sci.*, 1997, **16**, 2–10.
8 R. M. Donlan and J. W. Costerton, *Clin. Microbiol. Rev.*, 2002, **15**(2), 167–193.
9 B. R. Boles and A. R. Horswill, *PLoS Pathog.*, 2008, **4**(4), e1000052.
10 S. Jabbari, J. R. King, A. J. Koerber and P. Williams, *J. Math. Biol.*, 2010, **61**(1), 17–54.
11 N. Philippi, D. Walter, R. Schlatter, K. Ferreira, M. Ederer, O. Sawodny, J. Timmer, C. Borner and T. Dandekar, *BMC Syst. Biol.*, 2009, **3**, 97.
12 A. Funahashi, Y. Matsuoka, A. Jouraku, M. Morohashi, N. Kikuchi and H. Kitano, *Proc. IEEE*, 2008, **96**, 1254–1265.
13 A. Di Cara, A. Garg, G. De Micheli, I. Xenarios and L. Mendoza, *BMC Bioinf.*, 2007, **8**, 462.
14 M. Naseem, N. Philippi, A. Hussain, G. Wangorsch, N. Ahmed and T. Dandekar, *Plant Cell*, 2012, **24**, 1793–1814.
15 R. Schlatter, N. Philippi, G. Wangorsch, R. Pick, O. Sawodny, C. Borner, J. Timmer, M. T. Ederer and T. Dandekar, *Briefings Bioinf.*, 2012, **13**(3), 365–376.
16 J. Cassat, P. M. Dunman, E. Murphy, S. J. Projan, K. E. Beenken, K. J. Palm, S. Yang, K. C. Rice, K. W. Bayles and M. S. Smeltzer, *Microbiology*, 2006, **152**, 3075–3090.
17 P. M. Dunman, E. Murphy, S. Haney, D. Palacios, G. Tucker-Kellogg, S. Wu, E. L. Brown, R. J. Zagursky, D. Shlaes and S. J. Projan, *J. Bacteriol.*, 2001, **183**(24), 7341.
18 R. A. Brady, J. G. Leid, A. K. Camper, J. W. Costerton and M. E. Shirtliff, *Infect. Immun.*, 2006, **74**(6), 3415.
19 A. Resch, R. Rosenstein, C. Nerz and F. Götz, *Appl. Environ. Microbiol.*, 2005, **71**(5), 2663.
20 K. E. Beenken, P. M. Dunman, F. McAleese, D. Macapagal, E. Murphy, S. J. Projan, J. S. Blevins and M. S. Smeltzer, *J. Bacteriol.*, 2004, **186**(14), 4665.
21 G. D. Christensen, W. A. Simpson, J. J. Younger, L. M. Baddour, F. F. Barrett, D. M. Melton and E. H. Beachey, *J. Clin. Microbiol.*, 1985, **22**(6), 996–1006.
22 S. Bronner, H. Monteil and G. Prévost, *FEMS Microbiol. Rev.*, 2004, **28**, 183–200.
23 R. P. Novick, *Mol. Microbiol.*, 2003, **48**(6), 1429–1449.
24 K. Rogasch, V. Rühmling, J. Pané-Farré, D. Höper, C. Weinberg, S. Fuchs, M. Schmudde, B. M. Bröker, C. Wolz, M. Hecker and S. Engelmann, *J. Bacteriol.*, 2006, **188**(22), 7742–7758.
25 A. H. Bartlett and K. G. Hulten, *Pediatr. Infect. Dis. J.*, 2010, **29**(9), 860–861.
26 A. T. Giraudo, A. L. Cheung and R. Nagel, *Arch. Microbiol.*, 1997, **168**, 53–58.
27 C. P. Montgomery, S. Boyle-Vavra and R. S. Daum, *Infect. PLoS One*, 2010, **5**(12), e15177.
28 C. Goerke, U. Fluckiger, A. Steinhuber, V. Bisanzio, M. Ulrich, M. Bischoff, J. M. Patti and C. Wolz, *Infect. Immun.*, 2005, **73**(6), 3415–3421.
29 K. Rogasch, V. Ruhmling, J. Pane-Farre, D. Hoper, C. Weinberg, S. Fuchs, M. Schmudde, B. M. Broker, C. Wolz, M. Hecker and S. Engelmann, *J. Bacteriol.*, 2006, **188**(22), 7742–7758.
30 H. Kuroda, M. Kuroda, L. Cui and K. Hiramatsu, *FEMS Microbiol. Lett.*, 2007, **268**(1), 98–105.
31 K. Rogasch, V. Rühmling, J. Pané-Farré, D. Höper, C. Weinberg, S. Fuchs, M. Schmudde, B. M. Bröker, C. Wolz, M. Hecker and S. Engelmann, *J. Bacteriol.*, 2006, **188**(22), 7742–7758.
32 B. Fournier, A. Klier and G. Rapoport, *Mol. Microbiol.*, 2001, **41**(1), 247–261.
33 M. P. Trotonda, A. C. Manna, A. L. Cheung, I. Lasa and J. R. Penadés, *J. Bacteriol.*, 2005, **187**(16), 5790–5798.
34 E. Gustafsson and J. Oscarsson, *FEMS Microbiol. Lett.*, 2008, **284**, 158–164.
35 M. Mainiero, C. Goerke, T. Geiger, C. Gonser, S. Herbert and C. Wolz, *J. Bacteriol.*, 2010, **192**(3), 613–623.
36 M. Palma and A. L. Cheung, *Infect. Immun.*, 2001, **69**(12), 7858–7865.
37 M. Bischoff, J. M. Entenza and P. Giachino, *J. Bacteriol.*, 2001, **183**(17), 5171–5179.
38 K. C. Rice, T. Patton, S. J. Yang, A. Dumoulin, M. Bischoff and K. W. Bayles, *J. Bacteriol.*, 2004, **186**(10), 3029–3037.
39 K. C. Rice, E. E. Mann, J. L. Endres, E. C. Weiss, J. E. Cassat, M. S. Smeltzer and K. W. Bayles, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**(19), 8113–8118.
40 P. Sahlin, A. Levchenko and H. Jönsson, *PLoS Comput. Biol.*, 2010, **6**(6), e1000819.
41 J. P. Ward, J. R. King, A. J. Koerber, P. Williams, J. F. Croft and R. E. Sockett, *IMA J. Math. Appl. Med.*, 2001, **18**, 263–292.
42 E. Bonabeau, M. Dorigo and G. Theraulaz, OUP USA, 1999.
43 J. Schneider, A. Yepes, J. C. Garcia-Betancur, I. Westedt, B. Mielich and D. López, *Appl. Environ. Microbiol.*, 2012, **78**(2), 599–603.
44 R. Rosenstein, C. Nerz, L. Biswas, A. Resch, G. Raddatz, S. C. Schuster and F. Götz, *Appl. Environ. Microbiol.*, 2009, **75**(3), 811.
45 C. Vuong, F. Götz and M. Otto, *Infect. Immun.*, 2000, **68**(3), 1048–1053.
46 S. R. Gill, D. E. Fouts, G. L. Archer, E. F. Mongodin, R. T. DeBoy, J. Ravel, I. T. Paulsen, J. F. Kolonay, L. Brinkac, M. Beanan, R. J. Dodson, S. C. Daugherty, R. Madupu, S. V. Angiuoli, A. S. Durkin, D. H. Haft, J. Vamathevan, H. Khouri, T. Utterback, C. Lee, G. Dimitrov, L. Jiang, H. Qin, J. Weidman, K. Tran, K. Kang, I. R. Hance, K. E. Nelson and C. M. Fraser, *J. Bacteriol.*, 2005, **187**(7), 2426–2438.
47 P. A. Pattee, *J. Bacteriol.*, 1981, **145**(1), 479–488.
48 C. Wolz, P. Pöhlmann-Dietze, A. Steinhuber, Y. T. Chien, A. Manna, W. van Wamel and A. L. Cheung, *Mol. Microbiol.*, 2000, **36**(1), 230–243.
49 A. Toledo-Arana, N. Merino, M. Vergara-Irigaray, M. Débarbouillé, J. R. Penadés and I. Lasa, *J. Bacteriol.*, 2005, **187**(15), 5318.

