Construction and verification of the transcriptional regulatory response network of *Streptococcus mutans* upon treatment with the biofilm inhibitor carolacton

Vom Promotionsausschuss der
Technischen Universität Hamburg-Harburg
zur Erlangung des akademischen Grades
Doktor der Naturwissenschaften (Dr. rer. nat.)
genehmigte Dissertation

von
Padhmanand Sudhakar

aus
Trivandrum, Indien

2014

Gutachter:

Prof. Dr. An-Ping Zeng

Prof. Dr. Irene Wagner-Döbler

Prüfungsausschussvorsitzender:

Prof. Dr. Rudolf Müller

Tag der mündlichen Prüfung

23. Oktober 2014

Publications

1. P. Sudhakar, M. Reck, W. Wang, F. He, I. Wagner-Döbler, AP. Zeng. Construction and verification of the transcriptional regulatory network of the human dental pathogen *Streptococcus mutans* upon treatment with the biofilm inhibitor Carolacton. BMC Genomics 2014, 15:362
http://www.biomedcentral.com/1471-2164/15/362

2. F. He, W. Wang, P. Zheng, P. Sudhakar, J. Sun and AP. Zeng. Essential O2-responsive genes of *Pseudomonas aeruginosa* and their network revealed by integrating dynamic data from inverted conditions. Integrative Biology 2014, 6: 215-223.
http://pubs.rsc.org/en/content/articlelanding/2014/ib/c3ib40180d

3. L. Song, P. Sudhakar, W. Wang, G. Conrads, A. Brock, J. Sun, I. Wagner-Döbler, A.-P. Zeng (2012). A genome-wide study of two-component signal transduction systems in eight newly sequenced mutans streptococci strains. BMC Genomics 2012, 13:128
http://www.biomedcentral.com/1471-2164/13/128

Acknowledgement

I would first like to acknowledge the strong support of my supervisor Prof. Dr. An-Ping Zeng. My sincere thanks go especially to him for the kind of patience and confidence he bestowed upon me not just during my PhD stint but also during my previous years under his supervision. I would also like to sinerely thank Dr.Wei Wang and Dr.Feng He for their continuous support and well wishes. I am very fortunate to have had truly devoted colleagues for guidance and I would like to thank them all for the stimulating discussions and intellectual contributions. In addition, I heartily extend my gratitude to Prof. Dr. Irene Wagner-Döbler for being the secondary evaluator/reviewer and Prof. Dr. Rüdolf Müller for being the Chair of the evaluation committee.

Besides, I would like to thank everyone from the Institute for Bioprocess and Biosystems (IBB) for their co-operation and support. Special mention needs to go to all the colleagues who have been of immense help during my tenure as a PhD fellow at IBB. In particular, I would like to extend my gratitude to Mr. Lifu Song, Dr. Jibin Sun, Dr. Benedikt Schöpke for their inputs and valuable discussions on various topics. I also take this occasion to thank the immeasurable support provided by the technicians Mr.Ralf Grajetzki and Mr.Jan Sens in giving me a helping hand with technical, computational and miscellaneous issues.

Apart from the colleagues from IBB, I owe my gratitude to Prof. Dr. Irene Wagner-Döbler and Dr. Michael Reck from the Research Group of Microbial Communications, Helmholtz-Center for Infection Research, Braunschweig for their multiple contributions and firstly and foremost their permission to use the experimental data generated from their laboratory. In addition, their feedback and patience has been of immense value in compiling my work. I very much look forward to further co-operations with them in the future.

Although I was far away from my home country of India, the moral support provided by my parents, Mr. Ramesh Sudhakar and Mrs. Selvi Sudhakar, my brother Mr. Swaroopanand, my

wife Mrs. Sowmya Prabhu and my whole extended family comprising my grandparents Mr. Nagendra Rao and late Mrs. Santha Bai, my uncles and aunts and my loving in-laws (Mr. and Mrs. Ratnakar Prabhu) was invaluable.

I also take this occasion to thank the financial support provided from the project “Development of biofilm inhibitors using a systems biology approach” (0315411) financed by the German Federal Ministry of Education and Research (BMBF) in the framework of the Research Program "Medical systems biology - MedSys". Last but not the least, I take my pleasure in being able to thank the people of Germany for it is their hard earned money in the form of taxes which has financed this project. Also, on a personal note, my greatest debt is towards my wife Sowmya, who supported me all through the ups and downs of a PhD student’s life.

Abstract

Streptococcus mutans is an oral pathogen primarily responsible for causing human dental caries. The pathogenic abilities and cariogenic nature of *S. mutans* are closely related to its ability to form biofilms. Carolacton is a newly identified secondary metabolite from the myxobacteria *Sorangium cellulosum* causing altered cell morphology and death of *Streptococcus mutans* biofilm cells. To unravel key regulators and to uncover the underlying set of components and interactions comprising the network which mediate the observed effects of carolacton, the transcriptional regulatory response network of *S. mutans* biofilms upon carolacton treatment was constructed and analyzed. To date, reconstruction of a contextual genome-scale regulatory network has not yet been carried out for *S. mutans*. This study presents an attempt towards inferring the transcriptional regulatory network for *S. mutans* biofilms in response to the biofilm inhibitor carolacton. A systems biology approach combining time-resolved transcriptomic data, reverse engineering, transcription factor binding sites, and experimental validation was carried out.

The co-expression response network constructed from transcriptomic data using the reverse engineering algorithm - the trend correlation method, was found to consist of 8284 gene-pairs. The regulatory response network inferred by incorporating transcription factor binding site information into the co-expression network comprised 329 putative transcription factor-target gene relationships and could be classified into 27 sub-networks each regulated by a transcription factor. The sub-networks were observed to be enriched with genes sharing common functions thus lending biological meaning to the reconstructed regulatory network. The regulatory response network also displayed a hierarchy, albeit a double layered one, with respect to its global topology. Hierarchical organization has been known to be an important characteristic of regulatory networks. In addition, local network motifs which have previously been observed in the networks of model organisms like *E. coli* and *B. subtilis* could be identified within the *S. mutans* regulatory response network.

Sub-networks regulated by the pyrimidine biosynthesis regulator PyrR, the glutamine synthetase repressor GlnR, the cysteine metabolism regulator CysR, the global regulators CcpA and CodY and the two component system response regulators VicR and MbrC were found to be most related to the physiological effect of carolacton. The predicted interactions from the regulatory network between MbrC, which is known to play a role in the cell wall

stress response mechanism in *S. mutans*, and the *murMN*-SMU_718c genes encoding peptidoglycan biosynthetic enzymes were experimentally confirmed using Electro Mobility Shift Assays in cooperation with collaboratoring partners. Furthermore, deletion mutants of five key regulators predicted from the response networks were constructed and their sensitivities towards carolacton were investigated. Deletion of *cysR*, the node having the highest connectivity among the regulators chosen from the regulatory network, resulted in a mutant which was insensitive to carolacton, thus demonstrating not only the essentiality of *cysR* for the response of *S. mutans* biofilms to carolacton but also the relevance of the predicted network.

The network approach used in this study revealed important regulators and interactions as part of the response mechanisms of *S. mutans* biofilm cells to carolacton. It also opens a door for further studies into novel streptococcal drug targets.

Keywords

Streptococcus mutans; Caries; Biofilm inhibitor; Carolacton; Transcriptome; Reverse engineering; Transcriptional regulatory network; Key regulators; Network verification.

Zusammenfassung

Streptococcus mutans ist ein oraler Krankheitserreger, der für die Entstehung des Zahnkaries bei Menschen verantwortlich ist. Die pathogenen Eigenschaften - unter anderen die Kariogenität - von *S. mutans* sind eng mit seiner Fähigkeit zur Biofilmbildung verbunden. Carolacton ist ein neu identifizierter Sekundärmetabolit von dem Myxobakterium *Sorangium cellulosum* und zeigt eine inhibierende Wirkung auf das Wachstum des Biofilms von *S. mutans*, die sich durch veränderte Zellmorphologie und geschädigte Membran offenbart. Um den Wirkungsmechanismus von Carolacton zu verstehen sowie die zugrunde liegende Regulation der Genexpression in *S. mutans* aufzudecken, wurde in dieser Arbeit das regulatorische Netzwerk der Genexpression in Biofilmzellen von *S. mutans* anhand der Daten aus einer „time-series" Transkriptomanalyse rekonstruiert und analysiert. Da eine genomweite Rekonstruktion von regulatorischen Netzwerken für *S. mutans* bisher noch nicht durchgeführt wurde, stellt diese Arbeit einen ersten Versuch dar, ein regulatorisches Transkriptionsnetzwerk des *S. mutans*-Biofilms in Bezug auf die inhibierende Wirkung von Carolacton zu konstruieren. Ein systembiologischer Ansatz wurde angewendet, um mittels Reverse Engineering aus den zeitaufgelösten Transkriptom-Daten in Kombination mit Kenntnissen/Daten über Transkriptionsfaktor-Bindungsmotifs sowie experimentelle Validierungen die angestrebte Konstruktion des regulatorischen Transkriptionsnetzwerks zu realisieren.

Zuerst wurde aus den experimentellen Transkriptom-Daten durch Reverse Engineering ein Koexpressionsnetzwerk, das 8284 ko-exprimierten Gen-Paare beinhaltet, erstellt. Durch die Integration der Transkriptionsfaktor-Bindungsmotifdaten wurde dann das regulatorische Transkriptionsnetzwerk konstruiert. Dieses regulatorische Netzwerk umfasst 329 putative Transkriptionsfaktor-Targetgen-Beziehungen und kann in 27 Teilnetzwerke eingeteilt werden. Die Teilnetzwerke wurden jeweils durch einen Transkriptionsfaktor reguliert und mit Genen angereichert, die biologisch funktionsrelevant sind. Das regulatorische Netzwerk zeigte auch eine hierarchische Struktur in Bezug auf seine globale Topologie. Hierarchische Organisation ist ein wichtiges Merkmal eines regulatorischen Netzwerkes. Lokale Netzwerkmotifs, die bereits in regulatorischen Netzwerken von Modellorganismen, wie z.B. *E. coli* und *Bacillus subtilis* beobachtet wurden, konnten auch in diesem regulatorische Netzwerk von *S. mutans* identifiziert werden.

Es wurde festgestellt, dass die Teilnetzwerke, die von dem Pyrimidinbiosynthese-Transkriptionsregulator PyrR, dem Glutaminsynthetase-Repressor GlnR, dem regulatorischen Protein des Cysteinstoffwechsels CysR sowie den globalen Transkriptionsregulatoren CcpA und CodY und den Regulatoren VicR und MbrC aus Zwei-Komponenten-Signalübertragungssystemen reguliert wurden, im engsten Zusammenhang mit der physiologischen Wirkung von Carolacton stehen. Basierend auf dem konstruierten regulatorischen Netzwerk wurden unten anderen Wechselwirkungen zwischen MbrC, ein Transkriptionsregulator bekannt für seine Rolle in dem Zellwand-Stress-Antwort-Mechanismus von *S. mutans* und den Peptidoglycan-Biosynthese-Enzyme codieren Genen *murMN*-SMU_718c vorhergesagt. Dies wurde dann experimentell durch elektrophoretische Mobilitätsanalyse in Zusammenarbeit mit dem Projektpartner bestätigt. Außerdem wurden Deletionsmutanten von fünf prädiktierten Schlüsselregulatoren konstruiert und ihre Sensitivität gegenüber Carolacton untersucht. Die Deletion des Transkriptionsregulators *cysR*, der die höchste Konnektivität unter den identifizierten regulatorischen Netzwerkknoten aufwies, führte zu einem Mutant, auf den Carolacton keine Wirkung mehr zeigte. Dieses Ergebnis demonstriert nicht nur die Wichtigkeit von *cysR* für die Reaktion von *S. mutans* Biofilmzellen auf Carolacton, sondern auch die Prädiktionsvermögen des rekonstruierten regulatorischen Transkriptionsnetzwerks.

Der Ansatz dieser Arbeit führte zu der Enthüllung wichtiger Transkriptionsregulatoren und regulatorischer Interaktionen, die zu einem besseren Verständnis des Reaktionsmechanismus von *S. mutans* Biofilmzellen auf die Wirkung von Carolacton beitrug. Zudem können die entwickelten Methoden in Zukunft für weitere Studien mit neuen Wirkstoffen gegen Streptokokken angewendet werden.

Keywords

Streptococcus mutans; Karies; Biofilm-Inhibitor; Carolacton; Transkriptom; Reverse Engineering; Transkriptionsregulationsnetzwerk; Schlüsselregulator; Experimentelle Verifikation

Table of Contents

List of Figures

List of Tables

Abbreviations

TRRN	Transcriptional Regulatory Response Network
EMSA	Electro Mobility Shift Assay
TCS	Two Component System
RR	Response Regulator
gtf	glucosyltransferase
gbpB	glucan binding protein B
TSNI	Time Series Network Identification
GTAA	Genes corresponding to Transcripts with Altered Abundances
KEGG	Kyoto Encyclopedia of Genes and Genomes
TC Method	Trend Correlation Method
ORF	Open Reading Frame
PATRIC	Pathosystems Resource Integratrion Center
WT	Wild Type
RT-PCR	Reverse Transcriptase Polymerase Chain Reaction
SDS-PAGE	Sodium Do-decyl Sulfate Poly-Acrylamide Gel Electrophoresis
PSSM	Position Specific Scoring Matrix
RSAT	Regulatory Sequence Analysis Tools
SIM	Single Input Module
MIM	Multiple Input Module
FFL	Feed Forward Loop
MOFFL	Multi-Output FFL
UDP	Uridine Di Phosphate
UMP	Uridine Mono Phosphate
UDP-N-AG	Uridine Di-Phosphate N-Acetyl Glucosamine
ATR	Acid Tolerance Response
CFU	Colony Forming Units
LTTR	LysR Type Transcriptional Regulators
XIP	*sigX* inducing peptide
OAS	O-Acetylserine

Chapter 1. Introduction and aims of the thesis

1.1 Background and Problem Definition

Streptococcus mutans is an oral pathogen which, along with other closely related streptococci called the mutans streptococci, plays an important role in the formation of caries and tooth decay in humans. *S. mutans* is highly efficient in eroding the dental enamel and this is attributed to its ability to form biofilms which are generally difficult or impossible to eradicate by antibiotic therapy, because biofilm cells are resistant to antibiotics [1, 2] even if their planktonic form remains susceptible. This is in part due to the barrier effect of the polysaccharide matrix, but more importantly due to the profound genetic and metabolic adaptations of the cells to the sessile mode of growth [2, 3]. In addition, the high cell densities within the biofilms enhance the effect of quorum sensing, which lends new pathogenic and survival capabilities [4]. Besides the oral niche, *S. mutans* can also form biofilms on numerous internal surfaces such as implants, heart valves etc in the human body [5]. The occurrence of a series of synergistic effects within biofilms creates favourable conditions for caries-related pathogenesis and implant-related infections.

Recently, it was shown that carolacton, a secondary metabolite from the myxobacterial species *Sorangium cellulosum* has a high inhibitory activity against actively growing *S. mutans* biofilm cells, resulting in changes in cell morphology, elongation of cell chains, membrane damage and death of a part of the population [6]. Carolacton was also found to induce a dose dependent damage of *S. mutans* biofilms over a wide concentration range resembling a sigmoid dose response curve [6]. The chemical structure of carolacton was elucidated [7] and a *de novo* chemical synthesis recently published [8]. Carolacton inhibits *S. mutans* biofilms even at nanomolecular concentrations [6] implying that it primarily targets molecular entities which are present only as a few copies per cell. In this regard, carolacton is

very similar to compounds which target cellular signaling networks [9] rather than directly targeting functional enzymes in specific pathways associated with vital processes such as protein, DNA/RNA synthesis, cell division etc.

To decipher the genes whose expression is affected by carolacton, a time resolved transcriptome analysis of *S. mutans* biofilms after carolacton treatment was carried out by Reck et al [10]. Results from the study indicate that carolacton affects expression of genes related to biofilm formation, autolysis, pyrimidine and histidine metabolism, cell shape and cell division in addition to two component systems (TCSs) [10]. Even though the physiological and genetic responses of carolacton-treated *S. mutans* biofilm cells are known, the underlying network which orchestrates the expression of affected genes in response to carolacton still remains a mystery. This calls for an effort to uncover the effect of carolacton at the network level.

Biological networks fundamentally represent interactions or relationshsips in biological systems and can be represented in a graphical form for interpretation. In such graphic representations, edges normally denote the relationship(s) between the nodes or components (e.g. genes, proteins or metabolites) they connect. In the field of network biology, reverse engineering stands for the inference of biological networks from experimental data, such as datasets containing the expression profiles of the components in response to certain stimuli or a certain defined or undefined environment. The most predominant application of reverse engineering currently has been the inference of genetic co-expression networks from gene expression data, which are measured most commonly using cDNA microarrays or by RNA sequencing. A plethora of reverse engineering algorithms and methods has been developed for the reconstruction and inference of genetic co-expression networks from gene expression data. Excellent reviews about genetic network reconstruction from expression data have been published [11, 12].

In order to capture the network level events of biological systems upon exposure to various stimuli, reverse engineering methods have been developed to infer underlying networks from time series expression data, although static data has also been utilized. For instance, stress response networks of *Mycobacterium tuberculosis* after treatment with various drugs such as cerulenin, chlorpromazine, ethionamide, ofloxacin, thiolactomycin and triclosan were constructed from single time-point condition-to-condition measurements to delineate the differences between generic stress and specific drug responses [13].

In contrast, methods such as Time Series Network Identification (TSNI) were used for inferring co-expression networks and compound mode of action from time series gene expression data following interventions [14]. Although the inferred co-expression network using TSNI involved only a small subset of nine genes of the *E.coli* DNA-damage response pathway, it is seen as one of the first attempts in network reconstruction for determining the compound mode of action from a time series transcriptome. Further tools and algorithms were also developed to deal with whole genome network inference [15, 16], but the edges (representing relationships between genes) in such inferred co-expression networks were either undirected (no assigned causality) or carry only a statistical and/or theoretical probability of causality. Despite lacking directionality, undirected co-expression networks have nevertheless been used to infer critical genes and components involved in specific biological processes [17].

Even though genetic networks have been inferred under drug treatment conditions for some organisms, most of them have either been limited to compounds and molecules whose targets and mode of action were already known and/or limited to model organisms and certain human cell lines. Reconstruction of a contextual genome-scale regulatory network for the human caries pathogen *S. mutans* has not yet been carried out, although studies focusing on genome-wide transcriptional profiling have been reported [10, 18-21]. In this thesis, a workflow based

on a combinatorial strategy was executed in an attempt to infer a genome wide co-expression network for *S. mutans* biofilms under conditions of treatment with the biofilm inhibitor carolacton. Although Reck et al [10] have already measured the temporal progression of the *S. mutans* transcriptome in response to carolacton, their dataset is characterized by a small number of sampling points (five) and large time intervals, which does not allow for a reliable network inference. Hence, an extended time-series transcriptome was carried out encompassing a higher number of sampling points with relatively short time intervals. The trend correlation method [22] which has been used to infer genetic networks of T-cells [17] was used for inferring the co-expression network under carolacton treatment conditions. This was followed by the construction of a transcriptional regulatory response network (TRRN) by incorporating transcription factor binding site information.

1.2 Aims of the thesis

In contrast to traditional biological research focused on a few singled-out components, systems biology offers a new model of understanding biological systems from a holistic point of view. While previous models of biological research were primarily based on isolated parts, modern approaches are pinned on the premises of integration, connectedness and viewing the system as more than the mere sum of its composite parts. In this regard, the role(s) of genes and regulators in biological systems is/are attributed to their purported activities within the context of a vast network of interactions and components across different levels of organization. Thus, a phenotypic or physiological response is the cumulative result of many genes, their products and interactions among them at a systemic level.

Hence, to capture the true state and/or the underlying compendium of interactions of a biological system in transition or responding to any kind of chemical, biological or mechanical stimuli, information about multiple components across different levels of organization is warranted. System level analyses of biological effects and responses include

the integration of both static and dynamic information from various levels of cellular organization: genome, transcriptome, proteome and beyond. There exist two overarching approaches for the inference of network level responses: the first a top-down data driven approach and the second a bottom-up knowledge driven approach. While the latter offers a strategy to carry out detailed modeling of small scale networks involving components already known to play a role in eliciting the phenotypic response, the former approach is usually taken up in cases where there is a general lack of information about the exact components involved in the phenotypic response under study.

Carolacton is a secondary metabolite and shown to be an inhibitor of biofilms of the human dental pathogen *S. mutans*. The key regulators and genes which modulate the processes and mechanisms involved in the response of *S. mutans* biofilms to carolacton are not clearly known. The main aim of the thesis is to identify from the network inferred using a top down approach the key genes and regulators which mediate the response of *S mutans* biofilms to carolacton. The corresponding results from this work have also been published [380].

This thesis is based on the core results from a work-package within the BMBF financed project BioInSys ("Development of biofilm inhibitors using a systems biology approach") and implemented under the Medical Systems Biology flagship programme of the BMBF – MedSys. Specifically, the work-package and the thesis deals with the reconstruction of the transcriptional regulatory response network of *S. mutans* biofilms in response to carolacton. The transcriptional regulatory response network was reconstructed by integrating gene-gene associations derived from dynamic gene expression data using a reverse engineering algorithm and binding motif information. This was followed by the identification and verification of key regulators and regulatory relationships (from the inferred network) which mediate the response to carolacton.

Chapter 2. Literature review

2.1 *Streptococcus mutans*: a human dental pathogen

2.1.1 Genome

The first sequenced genome of *S. mutans* is the strain *UA159*. Published in 2002, this genome consists of 2,030,936 base pairs comprising 1963 open reading frames (ORFs) with an average GC content of 38% [23]. 63% of the ORFs have been assigned putative functions based on homology predictions while the remaining has not been assigned any known functions [23]. While 21% of the predicted ORFs have homologs from different species, 16% were found to be exclusive to *S. mutans* [23]. Since then, more strains (for example, the serotype c strain NN2025, the serotype k strain LJ23) have been isolated and sequenced [24-27]. Comparison of the sequenced strains LJ23, UA159 and NN2025 by BLASTP alignment have revealed strain-specific ORFs, although most of the protein-coding genes were found to be common among the strains [26].

Genome rearrangements among the strains as well as the possibility of horizontal gene transfer from closely related species such as *Streptococcus pneumoniae* have also been suggested [26]. Large genomic inversions possibly leading to genome plasticity and creating new genetic pools was also observed in a comparative analysis study between the two serotype c strains UA159 and NN2025 [27]. A recent sequencing and comparitive genome analysis of six clinical isolates of *S. mutans* has revealed significant divergences with respect to metabolic pathways and virulence genes [25].

2.1.2 Pathogenicity and virulence characteristics

The ability of *S. mutans* as the primary causative agent of human dental caries [28] to inflict oral caries (cariogenicity) is attributed to its biofilm forming capacity as well as a host of other factors (**Figure 1**) such as metabolic adaptation to different nutrient sources, production of adhesion molecules, acidogenicity (the ability to produce acid which erodes the enamel), aciduricity (the ability to withstand the low pH/high acidic microenvironment), quorum sensing, and genetic competence [3, 29]. Besides, its perfect adaptation to stressful conditions such as sporadic changes in pH, oxygen tension as well as the osmolality are tightly linked to its virulence and pathogenicity [3, 29]. Furthermore, it can compete with other oral species for survival [30-32].

2.1.2.1 Biofilm formation and adhesion

Although the virulent nature leading to the cariogenicity of *S. mutans* is due to a variety of factors, the most important of these is thought to be its ability to form biofilms [33, 34]. Moreover, biofilms serve to enhance synergistic effects by creating favourable environments such as protection from the host immune response, allowing the cells to remain in an area with an abundant supply of nutrients, and providing nutritional niches for different syntrophic bacteria thus promoting symbiosis [33, 35-37]. It is also known that the induction of competence among the bacterial cells in biofilms could exploit the large amount of free DNA which has been found in biofilms and thus increasing the possibility of the occurrence of beneficial mutations [38]. In addition, the high cell densities within the biofilms enhance the effect of quorum sensing, which lends new pathogenic and survival capabilities [4, 33, 39, 40]. Metabolic differences among the biofilm cells contribute to the resistance to anti-microbials as biofilm cells exist in a low metabolic state [2, 41, 42].

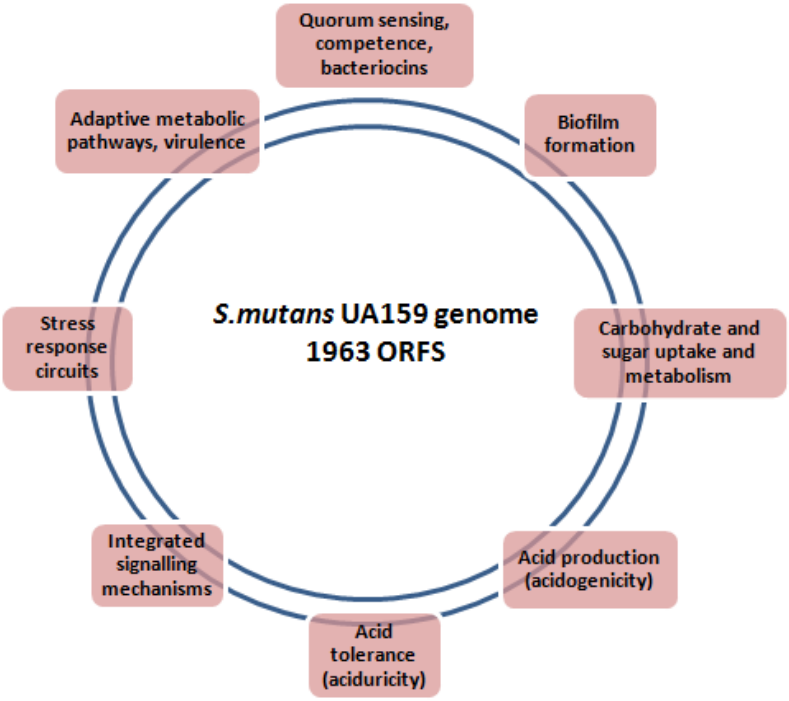


Figure 1. A graphical summary of the virulence attributes of *S. mutans*.

Besides the above mentioned advantageous properties of biofilms, other beneficial traits also include the occurrence of a population of dormant cells, transfer of antibiotic-resistance genes among strains and selection of resistance offering genes [43-45]. Biofilms, which form as a result of a series of complex interactions between proteins, glycoproteins, carbohydrates etc become difficult or impossible to eradicate by antibiotic therapy because cells within the biofilm become resistant to antibiotics even if their planktonic form remains susceptible [1, 2].

The formation of biofilms is preceeded by the anchoring of *S. mutans* cells to the human oral cavity. This process is mediated by many proteins, many of them with post-translational

modifications. These proteins help in the attachment of the cells or clusters of cells to any available surface as well as the salivary components in the human oral cavity [46-48]. *S. mutans* adheres to the tooth surface via both sucrose-dependent and sucrose-independent mechanisms, of which the former, upon consumption of sucrose, results in the rapid accumulation of *S. mutans* cells among the mixed oral population consisting of other colonizers as well [49, 50]. The binding of individual *S. mutans* cells and/or their colonies involve cell surface adhesion proteins such as surface antigen P1 or SpaP [51] which binds to salivary glycoproteins [52]. Yet another important adhesion protein is the cell wall associated protein WapA which is repressed in the presence of sucrose thus suggesting that it may equip *S. mutans* with the ability to bind to existing biofilms even without sucrose or glucan [53]. A critical process which enables *S. mutans* to form robust biofilms [54] and becoming cariogenic [55] is glucan production which occurs by the splitting and conversion of sucrose mediated by enzymes such as glucosyltransferases. The genome of *S. mutans* UA159 codes three glucosyltransferases namely GtfB, GtfC [56, 57] and GtfD [58] of which the first two are involved in the biosynthesis of water-insoluble glucan with alpha1-3 glycosidic linkages while GtfD produces glucan molecules with alpha1-6 glycosidic linkages [58].

Besides the glucan-producing glucosyltransferases, there has been an increasing body of evidence pointing to the occurrence of receptors responsible for glucan binding in *S. mutans*. These were called glucan binding protein A [59, 60], glucan binding protein B [61, 62], glucan binding protein C [63], and glucan binding protein D [64] which are known to play roles in adhesion [65], cell wall synthesis/cell division [62] and aggregation [64].

2.1.2.2 Acidogenicity, uptake and utilization of carbohydrates

The genome of *S. mutans UA159* codes for a series of interconnected biological processes and metabolic pathways which are dedicated to its virulence-causing acidogenicity [23]. An example of this is its capacity to utilize a variety of sugars to produce large amounts of acids

which erode the dental enamel thus initiating progressive tooth decay. The genome of *S. mutans UA159* possesses, in addition to dedicated specific ATP binding cassette transporter complexes specific to various moieties such as amino acids, alcohols and sugars, a series of phosphotransferase systems (PTSs) [23].

PTSs in *S. mutans* co-ordinate both the signaling and transportation activities of nutrient stress response mechanisms by importing sugars into the cell according to the nutritional state of the infection niche [66-70] since the oral cavity is subject to constant fluctuations of feast and famine. Among the *S. mutans UA159* PTSs, some are annotated either to be specific to a certain sugar or shown to have a broad range of sugar uptake specificities [23, 67]. Besides the transport of sugar molecules, PTSs have also been reported to modulate many other characteristics in *S. mutans* such as carbohydrate catabolite repression, expression of virulence factors such as glucosyltransferases, production of exopolysaccharides, fructan hydrolase expression, biofilm formation, ability to be transformed with exogenous DNA, and energy metabolism [71, 72]. In addition to PTS, *S. mutans* also transports sugars using non-PTS based uptake systems [73].

2.1.2.3 Acid tolerance

Various mechanisms attributed to multiple metabolic pathways and acid homeostasis are thought to be important for the aciduricity of *S. mutans.* Four different acid homeostatic mechanisms have so far been identified in *S. mutans*, all of which function to maintain a cytoplasmic pH that is more alkaline than the extracellular environment. While the membrane-bound F1F0-ATPase, which functions by extruding protons from the intracellular environment, is considered to the primary determinant of *S. mutans* acid tolerance [29, 74], other mechanisms such as the agmatine deiminase (AgDS) [75], glutamine-glutamate [76, 77] malolactic fermentation as well as the citrate pathway [78-80] related to the glutamine metabolism [76], also contribute to acid tolerance and pH homeostasis. It was observed that

the F1F0-ATPase and the agmatine deiminase (AgDS) acid tolerance systems also generated ATP molecules while extruding H+ ions out of the cytoplasm into the extracellular medium [75, 79, 80]. This mode of ATP generation could serve as a good energy source for cellular growth and maintenance when the cell is trying to battle acid production by activating acid tolerance eliciting pathways.

2.1.2.4 Signalling mechanisms

Signalling is considered to be an important element of survival and adaptation in bacteria as well as many other organisms. In addition to facilitating responses involved in house-keeping, they also integrate stress reception and response. However, *S. mutans* lacks prevalent bacterial sensing mechanisms such as alternative sigma factors for co-ordinating gene expression in response to various stress and stimuli [3, 23]. To overcome this, *S. mutans* has evolved by integrating signalling mechanisms into the general metabolic adaptation responses as well as its virulence modulation [3, 23] as evidenced by the documented functions of two component systems (TCSs) [39, 81-86]. In fact, TCSs comprise the major signal transduction components in bacteria in addition to PTSs.

While PTSs are involved in sensing and responding to sugars as well as their transportation, TCSs encompass a wide variety of functions. Moreover, the conspicuous absence of TCS proteins in mammalian genomes makes them interesting potential targets for the development of novel anti-bacterial drugs. Two component systems, as suggested by the name itself, are generally comprised of two classes of protein components namely the histidine kinases (HKs) and response regulators (RRs) whose encoding genes usually lie within the same operon. HKs are transmembrane sensors which upon sensing extrinsic or intrinsic signals are autophosphorylated at their conserved histidine residue (His) following which the phosphoryl group is then transferred to the aspartate (Asp) residue of the cognate response regulator.

Activated RRs then bind directly to DNA to modulate the expression of genes whose products function in the physiological response to the perceived signals [87-89].

14 TCSs were reported in the first sequenced strain *S. mutans UA159* although strain specific differences and similarities were also observed with respect to the TCS constituents [24]. In addition, an orphan response regulator without a cognate histidine kinase signal receptor was identified and found to play a role in pH homeostasis, glucan metabolism and adherence [90, 91, 92, 93]. TCSs in *S. mutans* have been reported to modulate the response to a wide variety of stresses including oxidative stress oxidative stress [94], acid stress [86], and cell envelope stress [95] in addition to influencing multiple virulence characteristics such as biofilm formation [96], competence [96, 97], autolysin production [98], glucan [96] and fructan metabolism [96], bacteriocin production and resistance [99], alarmone synthesis [83, 100] and quorum sensing [101].. Apart from TCSs, other mechanisms especially those involving the serine-threonine class of signal transduction components have also recently been identified in *S. mutans* and reported to modulate virulence characteristics such as competence development, bacteriocin production, cell wall metabolism, biofilm formation and acid resistance [102, 103].

2.1.2.5 Modulation of stress response

The *S. mutans* cells in the established and colonized biofilms are exposed to a variety of environmental and intrinsic stresses, all of which have led *S. mutans* to evolve primarily into a niche specific pathogen specialized in residing in the human oral cavity [104]. Besides acid stress, *S. mutans* is also exposed to other extrinsic and intrinsic stresses some of which include those which are imposed upon by host immune responses, nutritional deficiencies and fluctuations in the oral cavity, internal and inherent metabolic by-products, scarcity of micronutrients like calcium, cell wall damage, and damage to macromolecules such as proteins and DNA [3].

The immune response from the host is usually mediated via the production of oxidative radicals like hydrogen peroxide, and superoxide radicals [105]. Besides, a variety of other metabolic end products can result in oxidative stress and sporadic changes and deviations of the redox balance [105]. This can potentially inflict damage on components like enzymes, proteins, and affect the electron transport and energy metabolism leading to bacterial cell death [106]. *S. mutans* has evolved mechanisms for responding to such oxidative stresses by integrating stress signaling mechanisms with a variety of processes which neutralize and reduce the damaging effects of free oxidative radicals. One of the recently discovered players in oxidative stress response is the two component system VicKR [94]. The histidine kinase VicK of this TCS cluster harbors a domain belonging to the PAS domain family whose members are involved in the sensing oxygen tension, cellular redox state, or light intensity [107]. Interestingly, the VicKR system has also been implicated in the modulation of acid tolerance, competence, biofilm formation and cell envelope stress as well [96, 108]. This system regulates the expression of cell surface structure proteins such as glucosyltransferases, glucan binding proteins and fructosyltransferases [96]. Available evidence points to the fact that antibiotics such as vancomycin and polymixin which target the cell envelope also induced vicKR expression [109]. This gives further support to the notion that the VicKR system is involved in sensing bacterial cell-surface stress. Given the fact that the VicKR system controls the response to oxidative and cell envelope stress, it is interesting to know the relation between oxidative stress and formation of biofilms in *S. mutans*. Evidence from other bacterial species suggests that there is a definitive relation between oxidative stress and biofilm formation [110] since genes related to oxidative stress response were found to be differentially modulated during biofilm formation [111].

The modifications and alternations leading to the maintenance of the cell envelope and its components is thought to be an important contributor to biofilm formation since the cell

envelope harbors a variety of virulence factors such as glucosyltransferases and fructosyltransferases and also contributes to rendering protection against a variety of stresses [40]. In addition to the VicKR system, other TCSs such as the competence related ComDE [86], CiaRH [85, 112, 113], LiaRS [109, 114, 115], and BceRS [95] have also been implicated in either directly or indirectly modulating cell envelope stress. Reports point to the involvement of the LiaRS system not only in biofilm formation and responding to cell envelope stress but also in modulating the expression of genes which are involved in maintaining cell wall integrity [109, 115]. Recently, it was also demonstrated that the expression of the *vicKR* system is controlled by the LiaSR system [109] thus suggestive of possible hierarchies and cross-talk mechanisms between the TCSs. This could possibly be due to the co-ordinated responses which *S. mutans* might have to resort to for combating multiple stresses.

Nutritional scarcity and lack of energy sources is a commonly occurring type of stress to which *S. mutans* is exposed to in the human oral cavity due to the constant circulation of saliva and flushing. This places great importance on the adaptation of *S. mutans* to extreme conditions of nutritional famine during non meal times and excess during meal times [28, 116]. The requirement of *S. mutans* of carbohydrates for the formation of biofilms, the presence of an incomplete TCA cycle and the lack of an electron transport chain forces *S. mutans* to rely on and utilize dietary carbohydrates from the host. Furthermore, the sensing, uptake and metabolism of limiting sources such as sugars, carbohydrates etc have also to be co-ordinated so as to maintain constant supply of energy during famine as well as to ensure that there is a minimal fall-out/effect of harmful by-products as a result of rapid metabolism during times of feasting. Several mechanisms have been reported in *S. mutans* by which it manages to streamline the process of sugar and carbohydrate acquisition and metabolism. Energy metabolism is also modulated by the action of various enzymes such as lactate

oxidases [117, 118] which have been found to be non-ubiquitous and uniquely present in some genomes as observed in an in-silico analyses of six newly sequenced *S. mutans* strains [25]. This observation was also verified by PCR experiments [25].

The LevRS TCS forms part of a four-component system LevQRST which controls and sugar transport in *S. mutans* by modulating the expression of a PTS encoded by the *levDEFG* operon capable of transporting both fructose and mannose [119]. The LevQRST four component system along with the *LevDEFG* PTS it regulates, forms a potent and efficient mechanistic module which can not only sense (via LevQT) the extracellular concentration of fructose sugar but also activate a transport system (*levDEFG*) via the signal transfer mediated by the LevRS TCS. The LevQRST system was also reported to be involved as an activator of the gene encoding the fructan hydrolase enzyme [119] which cleaves fructose from inulin, and sucrose and raffinose from fructans [120]. In short, the fructan hydrolase enzyme helps in utilization of the extracellular polysaccharide stores. It was demonstrated that the *fruA* gene product plays a critical role in the cariogenicity of *S. mutans* indicating that the capacity *of S. mutans* to utilize exogenous fructans lengthens the time of exposure of host tissues to acids which are subsequently produced from the broken down and uptaken sugar monomers [121]. Thus via control of the *levDEFG* PTS as well as fructan hydrolase, the LevQRST system ensures that following the detection by the LevQT gene products, fructose sugar is taken in while the fructan hydrolase cleaves and releases the extracellular stores of sugar polymers for further consumption and acid production [119].

In addition to the fructan-specific *levQRST-levDEFG-fruA* response module, other global regulators such as CcpA (carbon catabolite repression protein) [71] and CodY [122] were also reported to be involved in managing nutritional stress. In *S. mutans*, CcpA has been shown to mediate global transcription of genes upon exposure to carbohydrates and upon being

knocked out, the mutant was observed to be acid resistant relative to the wild type [71]. Furthermore, the *ccpA* mutant was also found to be impaired with respect to its ability to transport and grow on sugars [71]. CcpA has been known to play major roles in carbohydrate uptake and metabolism in many other species of bacteria [123-127]. Experiments also support the possibility that there are redundancies in the systems responsible for carbon catabolite repression [71]. It was also shown that CcpA is required for the expression of fructan and glucan metabolizing enzymes fructosyltransferase (ftf) and glucosyltransferases (gtfB) [71]. Meanwhile CodY functions by helping *S. mutans* to adapt to nutritional stress by modulating amino acid metabolism namely the branched chain amino acids. CodY was also shown to be required for two key virulence properties of *S. mutans* namely acid tolerance and biofilm formation [122]. Taken together, the involvement of both global as well as specific regulators for sensing nutritional stress and co-ordinating nutritional stress responses is an indication of the importance of managing nutritional scarcity in the oral niche for *S.mutans*.

Different types of stress induce a lot of damage on the macromolecules such as DNA, proteins etc of *S.mutans.* Various mechanisms which respond to repair and restoration of damaged macromolecules were uncovered in response to different kinds of stress. Of note is the indispensable nature of the DnaK and GroEL proteins as evidenced by the study conducted by Lemos et al in which an imposed reduction of DnaK production resulted in a weakened ability to form biofilms as well as higher sensitivities to hydrogen peroxide, low pH and increased temperatures [128]. Other proteins such as the trigger factor RopA [129] and the surface associated protease HtrA [130] which are associated with post-translational modifications involving protein repair and alterations were also linked to virulence characteristics of *S. mutans*. Similarly genes and proteins belonging to DNA damage response pathways could be linked to stress response mechanisms [131].

2.1.2.6 Quorum sensing

The ability of *S. mutans* to form biofilms is considered as one of its most important virulence factors and is regulated by multiple systems. The most predominant among them is the density dependent signaling system termed as quorum sensing [132, 133], the core of which is the ComDE two component system [101, 134]. The phenomenon of quorum sensing enables bacteria to act collectively and this attribute of co-ordinated action modulates the expression of many of their virulence characteristics (such as biofilm formation, colonization, defense against possible competitor species, and adaptation to dynamic environments) when a certain threshold with respect to their local cell densities is attained [132, 133]. This is made possible via the production of signaling molecules (such as competence stimulating peptide (CSP) in the case of *S. mutans*) in the microenvironment and whose concentration is directly proportional to the cell density [40]. The detection of the signaling molecule CSP is mediated via the histidine kinase ComD of the ComDE two component system after which ComD is autophosphorylated. The phosphoryl group is then relayed to the response regulator ComE whose DNA binding capacity is subsequently altered upon phosphorylation. Thus CSP production and signaling triggers a series of downstream events which result in the modulation of genes responsible for various virulence phenotypes such as the well known acid tolerance response to biofilm formation [86].

Expression analysis using isogenic mutant derivatives deficient in the comD or comE or comC genes showed that the ComDE TCS had a positive regulatory effect on the expression of the genes coding for fructosyltransferase and glucosyltransferases [40]. It was also demonstrated that addition of CSP also upregulated the expression of the glucosyltransferase coding genes [40]. Some of the other virulence attributes which are regulated by the quorum sensing system in *S. mutans* include genetic transformation [101], and bacteriocin production [101, 134-136]. It was also observed that the QS system in *S. mutans* was found to be non-

responsive when exposed to foreign CSP (namely of *Streptococcus intermedius*) thus pointing to its plausible specificity [137].

Taken together, it can be stated that the stress response pathways and processes in *S. mutans* are integrated into the virulence mechanisms and hence the stress response regulon could be responsible for modulating wide-ranging biological functions when compared to other organisms and pathogens with larger and complex genomes [3, 23]. This could be due to the streamlined genome which *S. mutans* has evolved to have so as to develop into a specific pathogen with a specific niche – the human oral cavity.

2.2 Biological networks

2.2.1 Network definitions

Biological networks are graph based representations of biological systems to depict systemic components and the interactions among them. Generally, the interacting components are represented in biological networks by nodes and the interactions themselves by edges or linkages (**Figure 2**). Nodes usually depict components such as genes, proteins, metabolites, RNAs etc. The descriptions of the nodes vary according to the type of the biological network: in homologous networks such as co-expression and transcriptional regulatory networks (TRNs) [138], nodes denote genes; and nodes denote proteins in protein-protein interaction networks (PPIs) [139-141]. In heterogeneous networks such as metabolic networks [142], nodes could represent enzymatic components or small molecules such as metabolites and ligands. A linkage or an edge between two nodes meanwhile indicates a possible relationship between the components represented by the nodes. Edges may also carry attributes which indicate other lines of evidence such as mode of regulation (positive or negative), degree or strength of interaction etc.

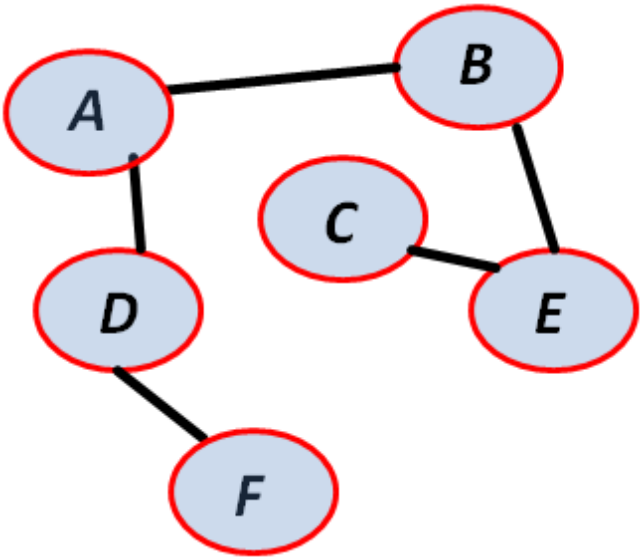


Figure 2. A simple graph-based representation of a biological network. Ellipsoids represent nodes which denote components (genes, proteins etc) of the network whereas the edges (connecting the nodes) denote linkages representing relationships between the nodes.

2.2.2 Classification of biological networks

2.2.2.1 Network classification according to linkage types

Biological networks can in general be classified based on types of linkages. A linkage can either be an experimentally determined physical relationship (termed herein as an interaction) or a putative/functional relationship (termed herein as an association).

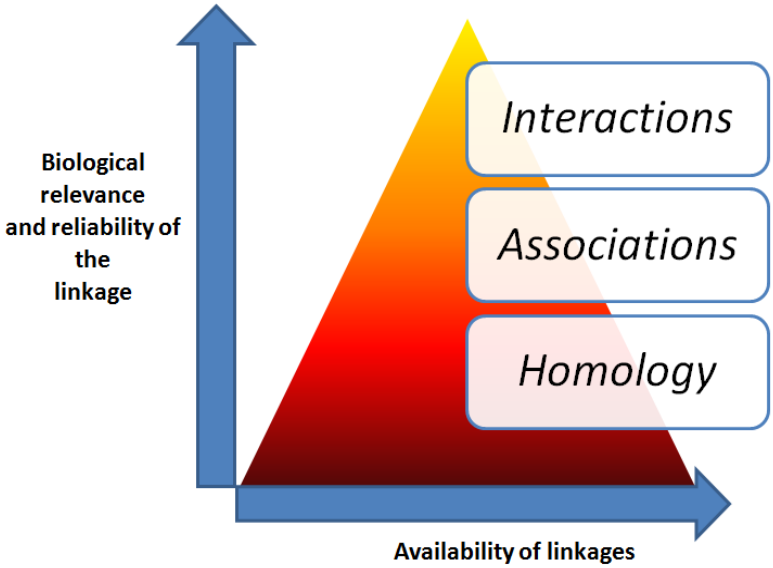


Figure 3. Biological relevance and reliability of linkages: Besides homology, functional linkages can also be derived from principles of phylogeny, gene fusion, gene neighborhood and co-expression. These comprise functional linkages or associations and have lesser relevance than physical interaction-based linkages which include protein-protein and protein-DNA interactions for example.

Interactions are implicit in terms of direct and experimental evidence whereas associations are putative even though they could possibly be due to predicted physical interactions which have not yet been revealed or discovered. Interactions refer to experimentally verified relationships such as protein-protein or protein-DNA/RNA interactions among others which also include protein-metabolite/small-molecule interactions and gene regulation mediated by small RNAs [143]. The type of linkage in turn determines its relevance (**Figure 3**): with physical interactions due to their experimental validation carrying more meaning and weightage than associations which are putative.

2.2.2.1.1 Interactions

Interactions refer to experimentally verified physical modes of contact between cellular components such as DNA, proteins, RNA, metabolites etc. As discussed before, interactions as compared to associations, comprise the most relevant and reliable form of linkages which can be assigned between any two components. The physiological and phenotypic characteristics of organisms are determined by the expression of and interplay among components from various levels of cellular organization: starting from the genome, transcriptome, proteome, metabolome etc.

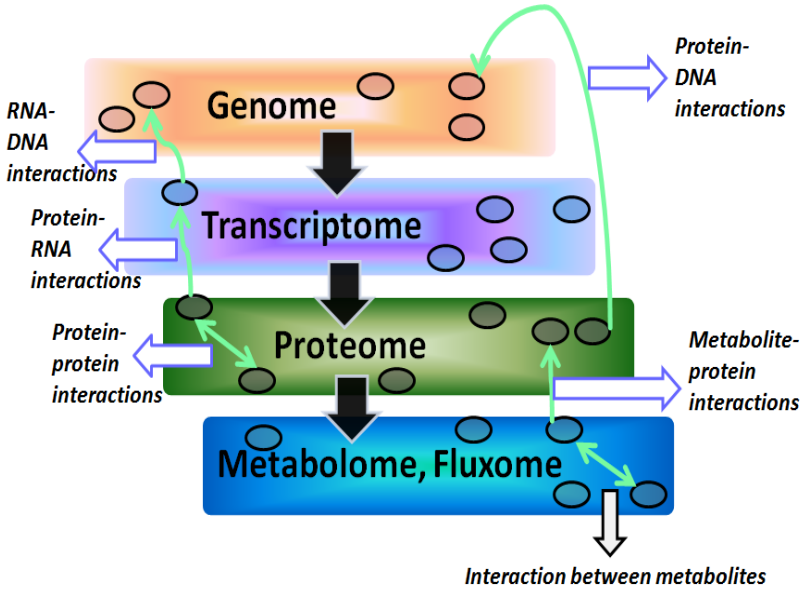


Figure 4. Various levels of interactions in biological systems. The green colored arrows denote the physical interactions which have been reported in literature. Uni-directional arrows represent interactions where control is modulated via the interactions whereas birectional arrows represent interactions which either mediate just structural binding and/or may or may not mediate control. Different types of RNAs (for example mRNA, rRNA, non-coding RNA factors, natural RNA aptamers etc) are pooled into the organizational level 'transcriptome'.

Further levels of organization such as the fluxome, secretome, surfactome etc have also been elucidated in biological system [144, 145]. Control mechanisms have also evolved to modulate the expression of the components involved in the various levels of cellular organization. It could be observed from **Figure 4** that control over the expression and activities of components (for examples genes) from a particular level (for example the genome), is mediated by components further downstream in the dogmatic lineage such as proteins and transcripts as evidenced by protein-DNA and RNA-DNA interactions (**Figure 4**). In addition, post-translational mechanisms such as protein-protein and protein-metabolite interactions also serve as agents of modulation and feedback [146].

Interactions mediating gene expression (control over transcript initiation)

Since genes are the fundamental information coding blocks which impart function at other levels via their protein products or RNA or small peptides, gene expression and its control is deemed to be critical [147]. In line with this, bacteria have evolved a variety of mechanisms to control gene expression: control where the transcription of information from DNA to RNA is modulated. Transcription results in the production of a transcript (messenger RNA or mRNA) from the coding sequence of the gene. The transcription of genes in bacteria is known to be controlled via two major mechanisms: protein-mediated and RNA-mediated DNA-interactions [148-151].

Protein-mediated modulation of gene expression (both positive and negative) occurs as a result of the binding/unbinding of a host of transcriptional regulators and other accessory factors such as RNA-polymerases, sigma factors etc to/from the upstream regulatory regions found in the promoter areas of the gene [153] (**Figure 5**). The interaction between the transcription factor and the upstream promoter sequence involves the recognition of a binding site comprised of 5-20 nucleotides (in the genetic upstream sequences) by the transcription factor [153].

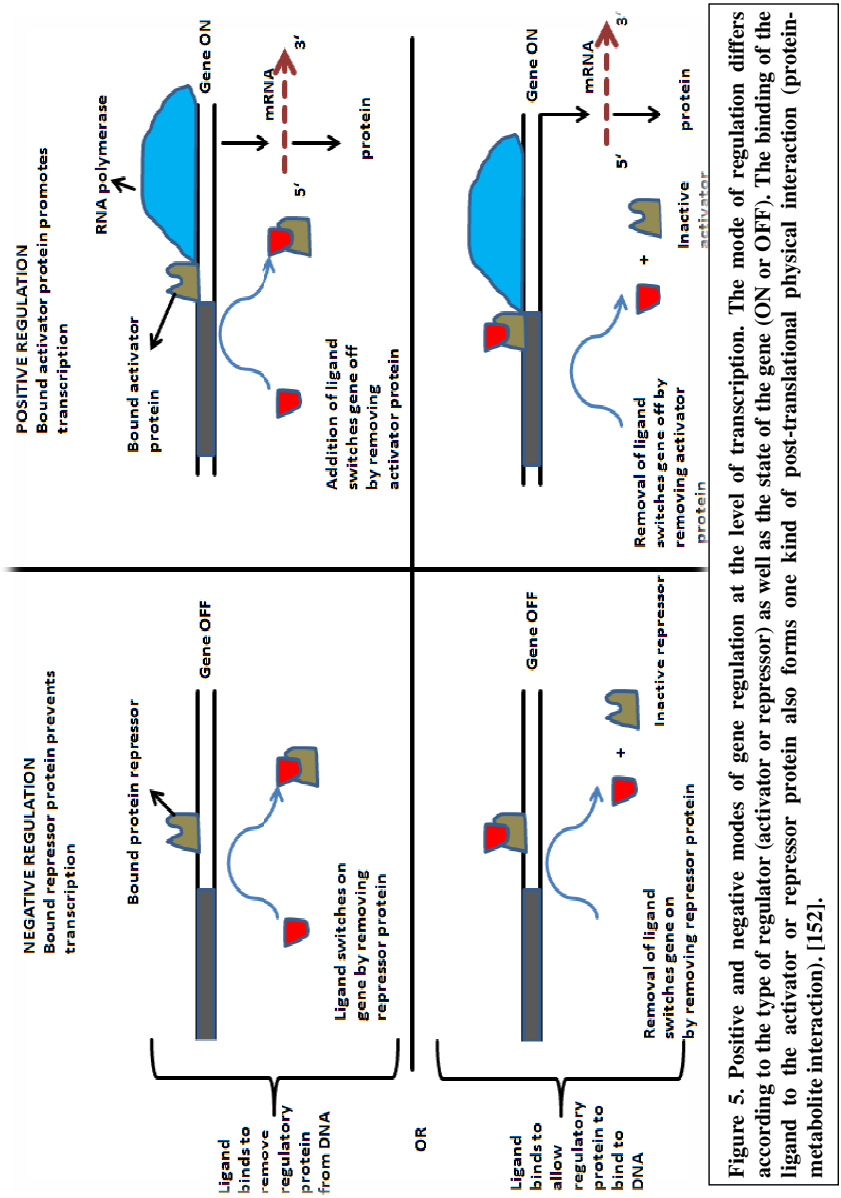


Figure 5. Positive and negative modes of gene regulation at the level of transcription. The mode of regulation differs according to the type of regulator (activator or repressor) as well as the state of the gene (ON or OFF). The binding of the ligand to the activator or repressor protein also forms one kind of post-translational physical interaction (protein-metabolite interaction). [152].

Various experimental methods such as electromobility shift assay (EMSA) [154-156], DNAase footprinting [157], ChIP-chip [158, 159] etc have been developed to verify such interactions. This mechanism of gene regulation is considered to be efficient since genes are transcribed and expressed only when and if a specific stimulus or signal triggers and activates the corresponding regulator or transcription factor. But on the other hand, it also brings with it the inherent disadvantage of not being able to respond quickly to rapidly changing conditions. Time-scales in the range of a few minutes could have elapsed between the perception of the signal and the elicitation of the transcriptional response [153].

The second type of gene regulation involving DNA interactions is related to riboswitches [151, 160-163] which are structured domains that usually reside within the non-coding regions of messenger RNAs. Metabolites bind to highly specific binding pockets in riboswitches causing allosteric structural changes following which the riboswitches bind to DNA to inhibit the transcription of the full-length mRNA [160-162].

Post transcriptional interactions (control over transcript fate)

To reduce the time lag between stimulus recognition and modulation of protein production via regulation of gene expression by transcript initiation, bacterial systems have developed other mechanisms to control gene expression and responses. One of the most prominent among them is the control established via transcripts or in other words conveniently defined as post-transcriptional mechanisms. There are a variety of sub-mechanisms within this broad category encompassing interactions exerting modulation via the transcripts. Riboswitches (**Figure 6**) not only regulate gene expression by inhibiting transcript initiation by DNA binding but also interact with other transcripts and thereby inhibiting translation initiation [151]. This happens when riboswitches induce formation of a helix that sequesters the ribosome binding site thereby reducing the efficiency of translation intiation [164]. There are also other models of

post-transcriptional control by riboswitches which induce the formation of an intrinsic transcriptional terminator upon ligand binding [165].

In addition, evidence also suggests that riboswitches play a role in splicing control as well [167]. Besides ligands and small molecules which also involve metabolites, proteins also bind to transcripts thereby preventing their translation to the protein form [168]. Post transcriptional interactions help in bypassing the transcriptional initiation regulatory circuit thereby adding to the flexibility and robustness of the biological system to respond to sporadic changes in the environment.

Protein-protein interactions

Next to interactions governing transcriptional initiation, protein-protein interactions (PPIs) (**Figure 7**) comprise the widely known post-translational interaction type due to the large scale techniques which were developed to screen out PPIs. The techniques for inferring PPIs include protein-fragment complementation assays [169, 170], affinity purification [171, 172, 173], fluorescence resonance energy transfer [174], mass spectrometry [175-177], protein microarrays [178], microscale thermophoresis [179], immunoprecipitation [180] and yeast two-hybrid system [181]. In biological terms, PPIs are associated with many functions [182]: for example, protein components from signal transduction pathways interact with each other via specific domains meant for facilitating such binding [183, 184].

Some of the PPIs which happen in signal transduction pathways like two component systems are associated with the transfer of phosphoryl groups [183, 186]. PPIs are also observed in enzymatic complexes where the different subunits orient themselves in sterical positions with each other so as to become functional [187].

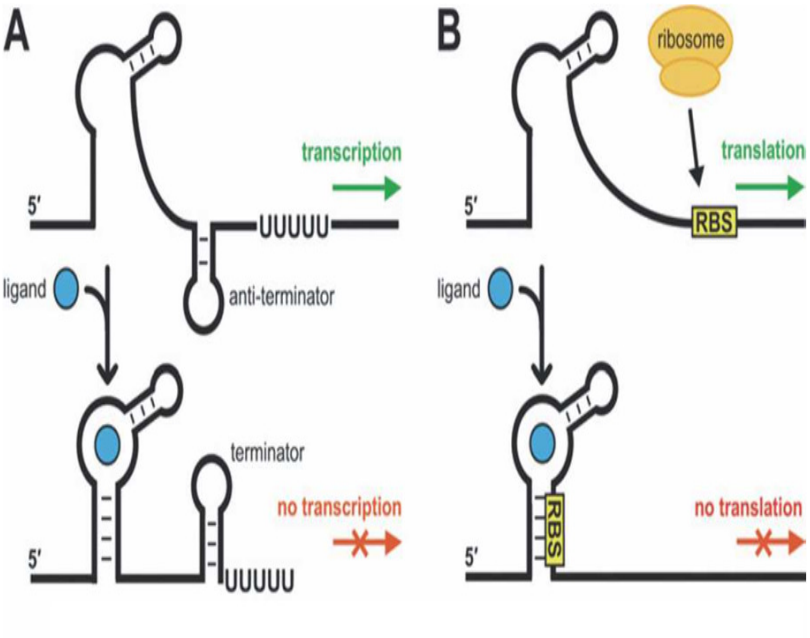


Figure 6. Riboswitches as a classical example of gene expression control established during the transcriptional as well the translational stages. Even though many mechanisms exist [151], riboswitches function to either facilitate a pre-mature transcription termination (A) or the inhibition of intiation of the translation (B) as a result of the binding action of a ligand or small molecule leading to formation/dissolution of RNA structures such as terminator loops on the transcript. Figure adapted and modified from [166].

Other examples in which PPIs play major roles in cellular control include dimer/multimer formation of structural and functional proteins. Transcription factors in their dimerized form for example have been shown to have significantly altered activities in relation to their monomeric forms [188]. Such activated forms of transcription factors after dimerization then regulate gene expression via transcriptional initiation [188] as explained before in section 2.2.1.1.1. In addition to functional, regulatory or enzymatic roles, PPIs are central to the formation of cellular structural complexes [140].

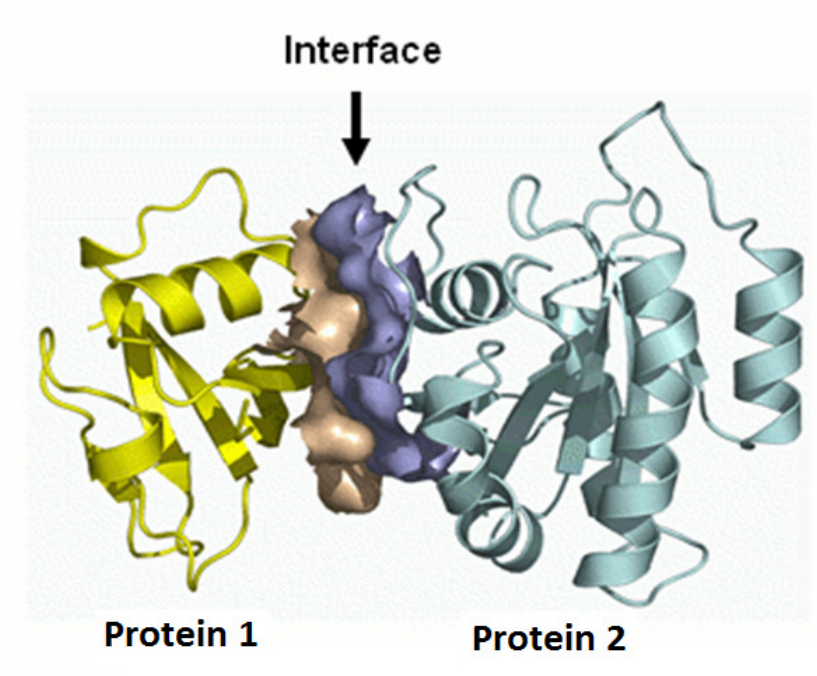


Figure 7. Protein-protein interaction depicted pictorially. Adapted and modified from [185]. Proteins interact with each other to form enzymatic and/or structural complexes and some protein-protein interactions are also associated with transfer of phosphoryl groups between each other [183, 186].

Protein-metabolite interactions

Small molecules, metabolites and ligands such as amino-acids and their derivatives for example play significant roles in the metabolic control of enzymes and their activities. Metabolites of an enzymatic pathway function by binding to enzymes which catalyze the product's formation/degradation or the synthesis of an intermediate [189, 190]. The most commonly observed form of this kind of enzymatic feedback via ligand binding to enzymes is the feedback inhibition [191].

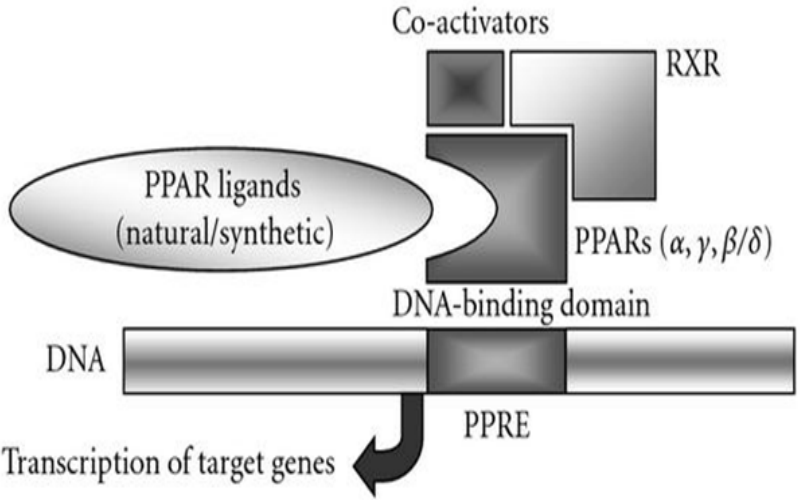


Figure 8. An example of a gene-regulation mechanism working on the principle of protein-ligand as well as protein-protein interactions. In this case, the example is derived from the case of peroxisome proliferator-activated receptors (PPARs) which in the presence of co-activator proteins and PPAR ligands bind to peroxisome proliferator response elements or PPREs on DNA thereby inducing the transcription of target genes. Adapted from [192].

Binding of small molecules and ligands to enzymatic proteins results in changes in the conformation and stability of the enzymes thereby affecting their activities [191]. Metabolic networks pool these kinds of interactions where the enzymes and metabolites are the key components. Not only enzymes, but also other proteins which function as transcription factors have also known to have altered activities and affinities upon binding of ligands [193](**Figure 8**).

Other post-translational interactions

Most proteins after being produced from the ribosomal assembly are processed in a variety of different ways before ending up as a functional protein product in the bacterial cell. Most of these modifications deal with the enzyme-catalyzed covalent addition of accessory functional

groups such as glycosyl groups, methyl groups, hydroxyl groups, mannosyl groups, N-ADP-ribosyl groups, acyl groups etc to side chain residues of the protein [194]. Alternatively, modifications of proteins where their peptide backbones are cleaved by proteases or auto-catalysis have also been reported [194]. Thirdly, the altered proteins could acquire new functions as a result of the primary modifications and could catalyze further reactions leading to post-translational changes of other proteins [146, 194].

The above described classification of interactions is not exhaustive in itself and has been explained only to give the reader an overview of physical interactions in biological systems.

2.2.2.1.2 Associations

The underlying cause of a predicted functional linkage such as an association may or may not be a physical interaction. In simple terms, associations are putative. Associations between two components such as genes are predicted from their functional contexts. Traditionally, the functional context of genes, or for that matter even those of other components such as proteins, were determined from the similarities of their sequences to previously characterized ones followed by the annotation of the query gene if the sequence similarity satisfied a specified cutoff. Non-homology based methods have recently been applied to understand functional relationships between proteins. Functional relationships between two genes or proteins are assigned based on their phylogenetic profiles in multiple organisms, gene fusion events (products of genes fused into a single protein), gene neighborhood (both genes in the same genomic vicinity), gene order conservation and co-expression at the transcript or protein levels [195-199]. **Table 1** provides an overview of the advantages and disadvantages of the various approaches used to predict associations.

Table 1. Advantages and disadvantages of the various approaches for predicting associations between genes/proteins.

Approach for association inference	Advantage	Disadvantage
Phylogenetic profiles	→Availability of large number of sequenced genomes for the determination of phylogenetic profiles across multiple genomes from various classes, families etc. →Conserved absence or presence in multiple species is representative of the evolutionary pressure of the proteins/genes to have co-evolved →Profiles compared across multiple genomes increases confidence on predictions	→Small deviations in sequence similarity could impact homolog prediction →The biological meaning behind predicted associations cannot be explained → Inferences from the profiles based on paralogs rather than orthologs can lead to erroneous predictions →The closeness of the taxonomic relationships between the genomes of the compared organisms are not weighted for the prediction of associations
Gene fusion events	→Ideal for assigning functional linkages or associations to proteins involved in consecutive metabolic steps and/or components of molecular complexes	→Promiscuous domains can cause uncertainties →Possibility of false positive predictions due to functionally interacting proteins →Deletion of intergenic sequences can also lead to false positive predictions → Inferences based on paralogs rather than orthologs can lead to erroneous predictions
Approach based on gene-neighborhood	→Based on intergenic distance for assigning associations and does not rely on homology →Captures basic biology : the closer the genes on a genome, the higher is their probability of being related	→Fails to capture the relationships between genes located far apart on the genome
Gene order conservation	→Based on conservation of gene orders such as operons for assigning associations and does not rely on homology →Straight forward in prokaryotes due to the presence of operons →Various tools available to predict operon structure in microbial genomes	→Eukaryotic gene organization is more complicated than in prokaryotes →Could get complicated in prokaryotes in cases where intraoperonic regulation is prominent
Co-expression	→Wide availability of gene expression profiles →Numerous tools and algorithms for gene expression analysis and reverse engineering for predicting co-expression based relationships → Associations based on co-expression derived from time-course expression profiles increase confidence in the predicted associations in that they represent relationships based on dynamic data	→The biological meaning behind predicted associations cannot be explained as such →Need for additional biological information for assigning causality

Association of genes based on gene order conservation

Conservation of gene order in terms of pairing has been used routinely for predicting physical interactions between proteins [200]. It has also been used for protein function assignment if a protein is part of a conserved pair with one among the two proteins not annotated. At the same time, if the products of both the genes/proteins have been assigned only tentative functions, then gene order conservation can be used to predict physical interactions and assign functional associations [201, 202]. In bacteria genes encoding for proteins with similar functions are organized into operons.

Products of genes from within an operon have been shown to be functionally associated [203, 204] and even known to be members of functionally interacting protein pairs [200]. For example, genes found to occur as members of conserved gene ordering or gene pairs encoded interacting proteins belonging to various functional categories. Some of these included ribosomal proteins, ATP synthase complex proteins, transporters, enzymes, chaperones, cell division, complexes corresponding to proteasomes and nodulins [200]. In a specific study (based on the *Mycobacterium tuberculosis* genome) aimed at assigning functional associations based on conservation of gene order, it was found that proteins encoded by genes which are clustered have a higher rate of functional linkages than genes which are not clustered or paired [203].

Association of genes based on conserved phylogenetic profiles

Two genes or proteins can be said to have functional relatedness if they are found to occur simultaneously across multiple genomes. The fundamental principle (**Figure 9**) of inferring linkages based on the conservation of phylogenetic profiles is that pairs of homologous/non-homologous proteins or genes that are often present or absent together in genomes are prone to have co-evolved [205-207].

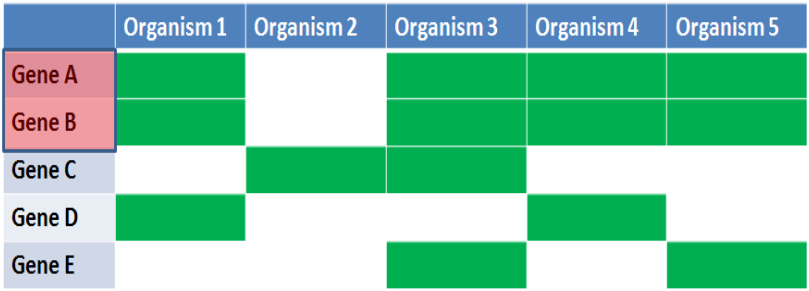


Figure 9. A pictorial depiction of an example showing the differential phylogenetic profiles of genes. Green cells indicate that the ortholog of the gene is present in the genome of the corresponding organism and white cells denote the absence. Genes such as A and B which share similar phylogenetic profiles are more likely to be involved in the same functional process or metabolic pathway and hence functionally linked. This paves the way for assigning a putative functional linkage (i.e and association) between the genes A and B.

This driving force for co-evolution could be a functional aspect which calls for the products of both the genes. The protein products may be required for imparting complete functionality to a metabolic pathway or a structural or enzymatic complex which in turn might be critical for the expression of a physiological phenotype, adaptation mechanism etc [205]. In other words, the co-evolutionary driving force behind conserved genes could be to enhance the fitness of the organism. Studies have found that genes within the same metabolic pathway have higher chances of sharing similar phylogenetic profiles than genes belonging to different pathways suggesting that the former are under the influence of a greater selective pressure than the latter

[205]. Hence, functional association-based linkages can be assigned to pairs of proteins or genes which have co-evolved in multiple genomes.

Association of genes based on gene fusion events

When a protein in a certain species is made up of fused subunits which usually correspond to a single full-length protein in another species, it is known as a composite protein. The individual proteins harboring the domains are termed as component proteins. Using comparative genomics, it can be assessed if a composite protein in a species is similar to two component proteins in another species [208] (**Figure 10**).

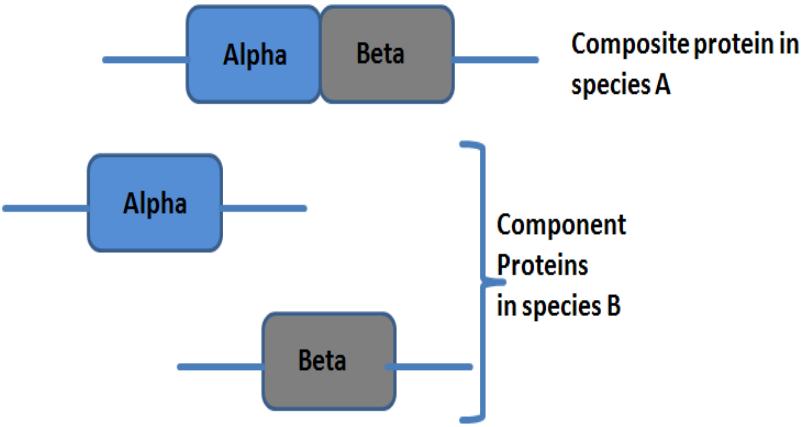


Figure 10. Graphical illustration of the concept of gene fusion events. The two domains (Alpha and Beta) of the composite protein (corresponding to a single gene) in species A is found to be encoded by two separate genes in species B. The evidence regarding the composite protein from species A supports the possibility of a functional gene fusion linkage between the two individual proteins in species B.

The genes encoding the composite proteins need not necessarily be neighbors. If a composite protein is satisfactorily identified, then it can be said that the component proteins are likely to be having a functional linkage [208, 209]. This predicted functional linkage could either be a plausible physical interaction or involvement in the same metabolic pathway or biological process [208, 209]. Since the notion of a physical interaction between the component proteins

is also a computationally predicted one, it is still considered as a functional linkage or an association.

Association of genes based on gene neighborhood

If two genes are found in close proximity to each other in a certain species of interest and at the same time, if their orthologs also exhibit the same kind of behavior in the genomes of multiple species, then the two genes can be defined as neighborhood genes (**Figure 11**). The neighborhood could be attributed to putative functional associations in which the gene products are involved [210, 211].

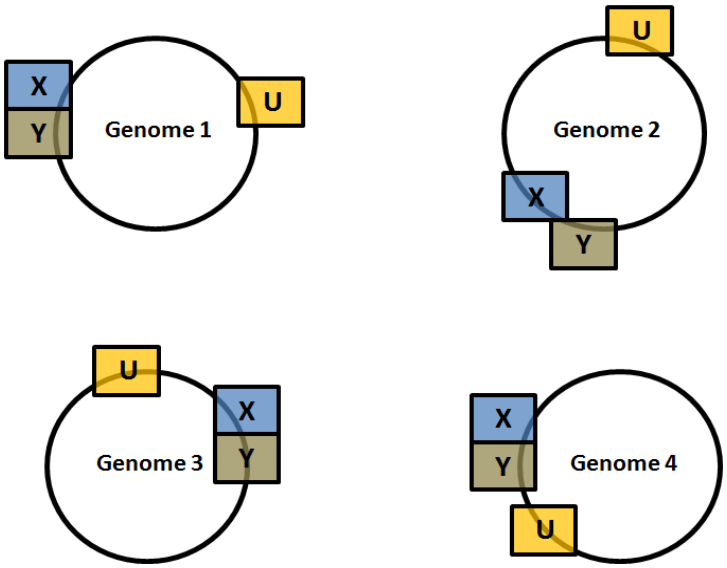


Figure 11. The genomic proximity profiles of three genes X, Y and U and their orthologs in the genomes of four different species as an example. It can be inferred that the genes X and Y are obvious gene neighbors as against the pair of X and U since U is positioned randomly compared to the X-Y pair. Genes corresponding to such conserved neighborhood across species could predictably be functionally linked.

Association of genes based on co-expression of their transcripts and/or protein products

Static as well as time-series –omic datasets at the level of the transcriptome and proteome are widely available and have been used to infer the causal mechanisms which could generate such profiles. One example in this regard is the inference of transcriptional regulatory networks (TRNs) from co-expression patterns and binding motif data. Besides, genes within an operon could also be co-expressed due to the common control exerted by the binding of a set of regulators or transcription factors to the upstream regulatory regions of the operon. It has been shown that co-expression patterns between two genes are not just attributed to common regulatory control or operonic membership, but could also be implicated with the functional aspect of the gene products. In other words, genes/proteins observed to be co-expressed over many conditions have a higher probability of being functionally associated than non co-expressed genes [22, 212].

Association of genes based on prediction of transcriptional regulatory and protein-protein interactions

Regulation of gene expression at the transcriptional initiation (see section 2.2.1.1.1) level is one of the most predominant types of well-studied and documented physical interactions in bacterial systems. In addition, they also constitute the largest group of experimentally verified physical interactions in model organisms such as *E.coli* [213], *B.subtilis* [214], *P.aeruginosa* [215] among others. Despite this, verified transcriptional regulatory interactions in most other bacterial species can at best be described to be a handful in number when compared to model or standard organisms.

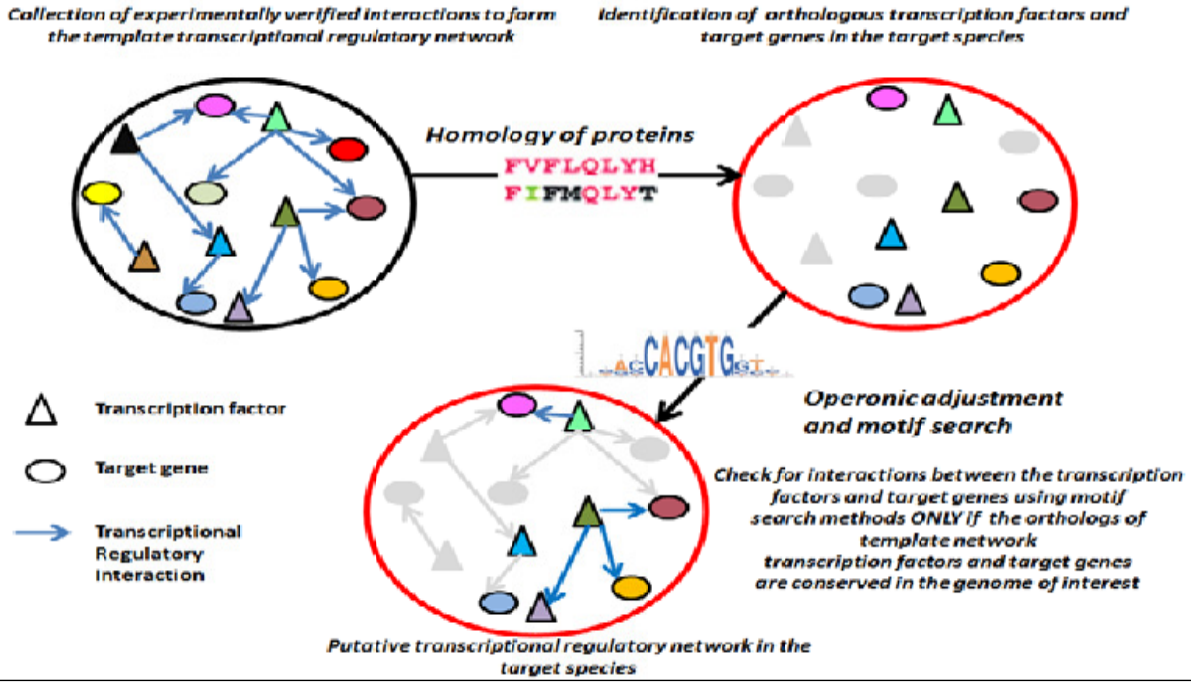


Figure 12. Comparitive genomic-based reconstruction of transcriptional regulatory networks. To start with, experimentally verified transcriptional regulatory interactions are compiled from different organisms of the same kingdom (to form the template network). Thereafter, it is checked whether the transcription factors and target genes in the template network have orthologs in the genome of interest. The next step involves the computational verification of the interactions in the genome of interest. This is performed by motif search analysis commonly called as pattern matching. Pattern matching methods help to identify if the consensus motif of the conserved transcription factor is identified in the upstream regulatory regions of the conserved target gene(s) in the genome of interest. Grey triangles and grey ellipsoids refer to the "non-present" orthologs of template network transcription factors and target genes.

Since such interactions form the basis for the construction of gene regulatory networks derived from gene expression data, it is imperative that this gap in knowledge be bridged so as to enable the construction and analysis of transcriptional regulatory networks (TRNs) for organisms which are not well studied. The determination of transcriptional regulatory interactions using a bottom-up strategy requires a lot of time and labour and large scale estimation warrants additional capital investments.

Hence, given the current limitations, and the fact that genome sequences of multiple organisms are available, bioinformatics-based techniques are employed to predict transcriptional regulatory interactions from genome data. The most widely used method of functional-genomic based prediction of bacterial transcriptional regulatory interactions involves the extrapolation of experimentally verified relationships from model organisms to the organism of interest based on the best-hit trilogy principle (**Figure 12**) as described elsewhere [216]. Briefly, this principle searches for orthologic counterparts of the transcription factor and the target gene as well as the conservation of the binding motif (from the experimentally verified interaction in either a model species or any other bacterial species) in the organism of interest.

2.2.2.2 Network classification according to directionality of linkages

Networks can also be either classified into directed or undirected depending on whether the linkages in the network are assigned directionality (**Figure 13**). By directionality, the causal nature of the linkage is implied. Directed networks consist of experimentally verified and predicted transcriptional regulatory networks. Metabolic networks can also be termed as directed when the direction of conversion of enzymatic reactions (both reversible and irreversible) is known. On the other hand, the remaining types of networks based on physical interactions as well as associations are generally undirected due to lack of information

pertaining to verified or predicted mechanisms which give rise to causality. Hybrid networks encompassing both directed and undirected linkages can also be compiled.

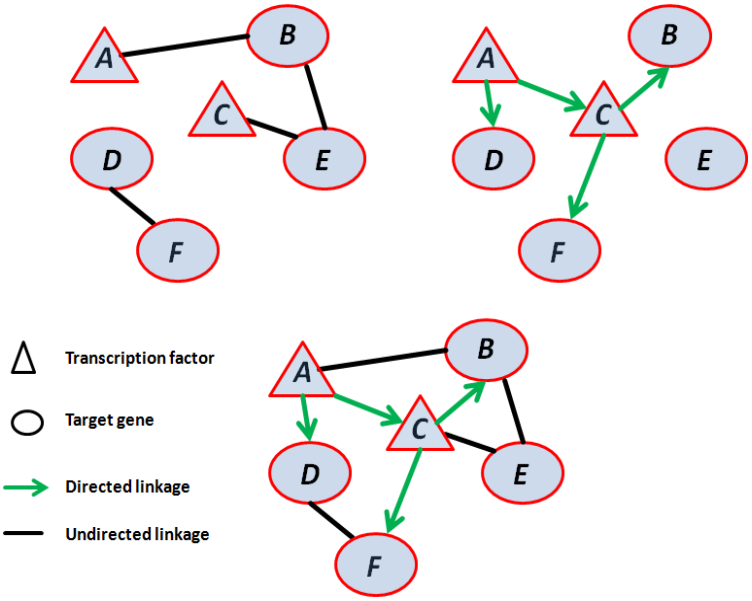


Figure 13. Different types of biological networks based on the availability of linkage directionality. The top left and right networks are purely undirected and directed networks respectively whereas the bottom network is a hybrid one with both directed and undirected linkages.

2.3 Inferring genome-wide transcriptional regulatory response networks

The most basic level of control of gene expression is exerted during the regulation of transcription. Even though other levels of gene regulation exist (as explained and outlined in Chapter 2), extensive progress on genome-wide network mapping has not yet been achieved with respect to these regulatory modes. Even as new regulatory modes and mechanisms are

being discovered, the techniques of analysis being established for evaluating such regulatory relationships are still at their infancy. This has thrown major limitations at scaling-up the scope of new regulatory modes and relationships to whole-genome levels. On the other hand, the inference of transcriptional regulatory networks on whole-genome scales, although not still complete, has been performed for various organisms [217-224] with available information. This has been partly due to reduced costs for carrying out whole genome gene expression profiling using microarray [225, 226] and RNA-seq technologies [227, 228], the availability of transcriptional regulatory information such as binding sites [229], new flourishing technologies like ChiP-Seq and ChiP-Chip [159, 230], the numerous reverse engineering algorithms [11] and the application of comparative genomics in extrapolating transcriptional regulatory relationships across sequences and annotated genomes [231, 232] . In this section, a brief but informative overview is provided with an outline of the current state of development with regard to whole genome transcriptional regulatory network inference as well as the various challenges and drawbacks involved in various steps of the network inference process.

2.3.1 Steps involved in the inference of transcriptional regulatory networks

The process of inferring transcriptional regulatory networks can be classified into three primary steps: generation of gene expression data, application of reverse engineering algorithms to deduce gene-gene correlation and the integration of biological causality into the gene-gene co-expression network inferred in the previous step. Although it could be argued that other miscellaneous procedures could be involved, the transcriptional regulatory network inference procedure can be broken down into the main steps as stated above.

2.3.2 Generation of gene expression data

Since transcriptional regulatory networks deal with the transcriptional egulation of gene expression, gene expression data at the transcript level is of central importance and serves as a starting point from which further analysis can be carried out. The measurement of gene expression at the transcript level has traditionally been measured using reverse transcription PCR (RT-PCR) [233], differential display RT-PCR [234], FISH (flouroscent in-situ hybridization) [235], Northern blotting [236], and reporter gene assays [237]. Although these techniques were quite accurate and efficient, they had practical limitations on scaling up and hence their applications were limited to a small scale involving a few dozen genes. However, these techniques can be put to use to carry out a second confirmation of the expression of a few selected genes as well. But with the advent of improved technologies such as DNA microarray technology [238], it became practically possible to monitor the expression of thousands of genes from a single genome basically on a single chip. Other novel high-end developments in measuring gene expression include the Serial Analysis of Gene Expression (SAGE) [239], RNA-seq [240], comparative EST (Expressed Sequence Tags) [241], macroarays [242] and Massively Parallel Signature Sequencing (MPSS) [243]. The developments and improvements in robotics and computational informatics have made it possible for the establishment and advancement of these high-throughput methods as general practice for measuring gene expression in modern biological research. These innovative technologies though have brought in new challenges as well as opportunities into the horizon of scientific research. Nevertheless, these developments bode well in that they contribute positively to basic scientific themes in systems biological research such as inferring and understanding gene regulatory networks.

The expression of genes can either be monitored over time (time-resolved/time-series), between conditions (static/sample-to-sample) or a combination of both of the above (time

resolved conditional). Here again, the advent of high-throughput technologies has had an overwhelming impact on the experimental design itself. Since gene expression data is so very essential for the network inference procedure, the higher the quality and greater the volume of data, the higher is the reliability and accuracy of the inferred network. In this context, it is bound to be obvious that gene expression monitored over time with intervals (preferably short time intervals) in accordance with the time-scale of the biological process under study would be more informative than a static-approach based gene expression measurement. Whereas cost and practical difficulties limited gene expression measurement during former times to a static/sample-to-sample design, the positive impacts of high-throughput technologies have now made it possible for researchers to monitor gene expression under multiple conditions over narrow time-intervals.

2.3.3 Inference of gene-gene expression correlation

As explained before, gene expression measurements at the transcriptional level are the basis for inferring genetic networks including transcriptional regulatory networks. Before arriving at transcriptional regulatory relationships between a pair of genes, statistically significant co-expression relationships between the genes need to be established. While there exist variants of many simplistic concepts such as the [244] which have been widely used to infer the co-expression relationships between genes, reverse engineering concepts and algorithms have taken the center-stage in the last decade or so in contributing to the inference of genetic networks. As a concept, reverse engineering with respect to biological networks refers to "the process of revealing the network structure of a biological system by reasoning backward from observed data" [11]. Many review articles on reverse engineering methods and their role in inferring genetic networks have been published [11, 12, 245, 246].

Reverse engineering tools and algorithms can be grouped together [11] into major methods based on their commonalities with respect to their theoretical bases and concepts such as

linear regression [14], graphical gaussian models [247, 248], neural networks [246, 249], bayesian networks [250-253], partial Pearson or Spearman correlation [254], mutual information [255], pair-wise statistical association linkage [22, 256], ordinary differential equations [14], S-systems [257, 258], genetic algorithms [259], linear state-space model [260] etc. Tools based on each of these concepts have their own advantages and drawbacks thus prompting the suggestion [11, 12, 22] from many researchers in this field of study that multiple methods need to be used in order to minimize the fraction of false positives and maximize accuracy in predictions.

While a plethora of methods are available for the inference of genetic networks from static gene expression data, the number of methods for treating time-series data is relatively low in number. Furthermore, as of date, no method has been specifically designed for dealing with time resolved gene expression data with non-periodic sampling intervals (with dissimilar intervals between samples). This translates into a major bottleneck for researchers since the time-scale of biological processes vary from one process to another. In this context, experiments need to be designed with sampling intervals which are in line with the time-scale of the biological process or system under study so as to capture the true dynamics representative of the system.

As part of the work included in this thesis, the reverse engineering method termed the Trend Correlation Method [22] was used to infer the co-expression based genetic network consisting of correlated connections between genes. The TC method allows the inference of gene-to-gene time-lagged positive or negative “associations” or pairs of genes based on extracting the main features of the change trend and the correlation of gene expression changes between consecutive time points [22]. Furthermore, the TC method was adapted to the non-periodic nature of the gene-expression dataset by calculating the correlation coefficient from the change rates rather than the change levels between consecutive time points since the time

interval of the measurements was not constant (personal communication, Feng He). In comparison to partial correlation, the TC method has further advantages in that it enables the identification of time-lagged relationships, estimates the co-expression from both the change levels and change trends and calculates the correlation co-efficient and P-value from the expression levels corresponding to the time intervals with similar match trends.

2.3.4 Assigning causality to co-expression networks and recent developments in genome-scale transcriptional regulatory network inference

As discussed previously, networks can either be fully directed, fully undirected or contain edges some of which are directed and others undirected. In networks such as genetic co-expression networks, linkages between genes are undirected whereas the edges of other networks such as the transcriptional regulatory networks for example are characterized by information specifying directionality and causality. The mere presence of an edge (as inferred from significant co-expression and correlation) between two genes does not necessarily represent a directed relationship between them. Thus, even though genetic co-expression networks have been used to infer critical components involved in biological processes [17, 261-266], their very nature of being undirected points out the lack of biological information for uncovering basic mechanistic aspects underlying the observed expression patterns of genes. In simple terms, further evidence is required for at least assigning putative biological causality and subsequently for pursuing precise experimental validation.

Regulatory networks on a genome-scale have been inferred by using combinatorial methodologies [217] incorporating not only gene-gene interactions (predicted from co-expression) from reverse engineering algorithms but also biological information such as data on binding motifs and promoter elements [218], functional genomics [219], genome annotation [220] and transcription factor activities [221]. Although non-exhaustive due to the lack of biological information (such as incomplete knowledge of transcription factor binding

sites), genome wide transcriptional regulatory networks have nevertheless been constructed from heterogeneous data for a few organisms such as *Escherichia coli* [221], *Saccharomyces cerevisiae* [222], *Bacillus subtilis* [218], *Candida albicans* [223], *Streptomyces coelicolor* [219], *Halobacterium NRC-1* [220] and others. Despite the advances made in the field of genome wide regulatory network reconstruction, most of these attempts have been carried out for model or standard organisms such as *E.coli*, *S.cerevisiae*, *B.subtilis* [218, 221, 224] for which accumulated biological information exists. Some progress has also been made in the case of human cells as shown by Basso et al. where human B cell gene regulatory interactions were inferred [267]. Hence, one of the major current challenges in the field of network biology centers around the whole genome directed regulatory network inference strategies for less-studied organisms which are either pathogens or are industrially important.

Even though the last decade has witnessed an explosion in terms of the generation of component interaction data such as protein-protein interactions [268-275], transcriptional regulatory relationships [214, 215, 219], metabolite measurements and fluxes [276-278], protein-metabolite binding [279, 280], pathogen-host interactions [281], interaction compendiums [282] etc to name a few, the scenario is far from satisfactory or ideal for most species with the exception of a few model standard organisms.

Although real biological networks are much more complicated in nature due to their heterogeneous nature of interactions among components from different levels of organization, the main focus of this thesis is the inference of a genome-wide transcriptional regulatory response network of the human caries pathogen *S. mutans* under treatment with the biofilm inhibitor carolacton. The inherent disadvantage of non-directionality in co-expression networks can also be overcome by incorporating binding site information of transcription factors as has been demonstrated for both prokaryotic and eukaryotic organisms [218, 283, 284]. Regulation of gene expression at the transcript intiation stage comprises one of the most

fundamental and well-studied modulatory mechanisms in bacteria and is mediated by protein-DNA interactions involving regulator proteins. The availability of annotated genomes of a large number of prokaryotic organisms has also provided biologists with the possibility of applying functional genomic and comparative genomic approaches for assessing regulatory elements. As described previously, regulatory proteins bind/unbind to/from the promoter elements in the upstream regions of genes and/or their corresponding operons to activate or repress their expression depending on the nature of the regulator protein itself (activator or repressor). Regulator proteins bind to the genetic upstream elements via a specific sequence or pattern of nucleotides anywhere between 5-25 base pairs in length on the DNA. Such specific sequence elements which mediate the binding between regulator proteins and the promoter elements of genes are called transcription factor binding sites and all the genes under the control of a specific regulator comprise its regulon. A collection of binding sites retrieved from various target promoter elements for a specific transcription factor gives rise to a binding motif which in other words represents a binding site consensus. Transcription factor binding site information and inferences from genetic co-expression under specific conditions can then be integrated to yield contextual transcriptional regulatory response networks (TRRNs) as is the case presented in this thesis.

2.4 Inferring the transcription factor – target gene map using putative binding sites

2.4.1 Sources of binding site data

2.4.1.1 Literature information on *S. mutans* transcription factors

Various experimental methods and techniques such as EMSA, nuclease protection assays, DNAase I footprinting assays among others have been used to validate numerous transcriptional regulatory relationships. Of the more than 100 known and predicted

transcriptional regulators in the *S. mutans* UA159 genome, only a few have been studied experimentally and an even smaller number characterized with respect to their binding site(s). The sparse characterization and elucidation of transcription factor binding sites is observed for many other non-model organisms in addition to *S. mutans* as well due to limitations with respect to the scale and feasibility of experimentation [216]. This poses a critical problem for inferring a transcriptional regulatory network on a genome-wide level. The non-availability of experimentally verified transcriptional regulatory binding sites for the *S. mutans* transcriptional factors was compensated by applying comparative genomic approaches [285-287].

2.4.1.2 Comparitive-genomics based prediction of transcription factor binding sites

The availability of a growing number of complete prokaryotic genomes throws open the door for the bioinformatic analysis and prediction of cis-acting regulatory elements across species and genomes. The basic tenets of such an approach involve the description of known regulons in non-standard organisms (pattern matching [288])(**Figure 14**) and the *ab-initio* prediction of novel regulons (pattern discovery [288]).

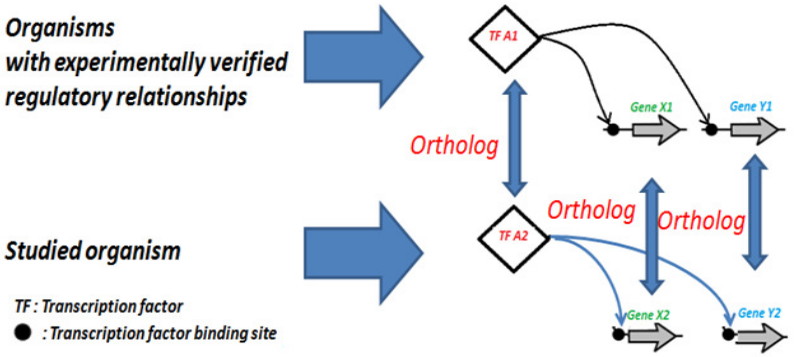


Figure 14. Abstract principle followed to arrive at regulons in non-characterized organisms.

In addition to the above mentioned paradigms in the prediction and analysis of cis-acting regulatory elements, their reliability and accuracy can be improved by checking for

consistency (of orthologous sites) across multiple genomes in addition to the functional annotation of the genes harboring the sites in their promoter elements [216]. Whereas pattern matching and extrapolation reveals new putative regulons in not-so-well characterized species, pattern discovery concerns itself with the identification of novel motifs conserved from sets of genes sharing common functions, genes which are co-regulated, orthologous genes and phylogenetic footprinting [289-293]. Put together, these approaches open up new avenues for the computational reconstruction of bacterial regulons based on transcription-factor binding sites.

2.4.1.3 Regulatory databases

In addition to literature and publications, binding sites can also be retrieved from regulatory repositories and databases such as RegulonDB [219], Prodoric [215], DBTBS [214], RegTransBase [229] to name a few. These databases contain a collection of regulatory interactions from literature for multiple bacterial species. Besides, there are also databases such as RegPrecise which are specifically dedicated to regulons that were reconstructed using comparative genomic approaches across a wide collection of prokaryotic genomes [231].

2.4.2 Methodologies, axioms and statistics of pattern matching

Pattern matching

Proteins which regulate gene expression bind to specific locations on DNA and these are called transcription factor binding sites. Many bioinformatics based methods have been formulated for the identification of TF binding sites in DNA sequences [294, 295]. Pattern matching is defined as the process of searching for sites recognized by a known TF, and requires prior knowledge of sites that describe the binding specificity of the TF. Pattern discovery is defined as the prediction of novel motifs in a set of genes that are putatively

regulated by some TF, without any prior information about its binding specificity**.** In this work, the focus is on Pattern Matching (PM) and it is applied to the detection of binding sites in the regulatory regions of genes from the constructed co-expression network of *S. mutans* biofilms in response to carolacton.

Pattern Matching methods can be divided into two classes namely matrix- and string-based PM [296]. In string-based PM methods, a string (represented by nucleotides in various combinations according to the composition of the binding site) is used to denote a binding site. A simple string-based pattern matching generally gives poor predictions for transcription binding sites since a single string-based representation fails to capture the binding site variability and degeneracy. Strings are also not reliable when searching for new sites since they do not account for possible degeneracies which could occur in sites which have not yet been characterized, unless allowances for mismatches are made [297-299]. The matrix form measures the number of occurrences of each of the four nucleotides at every position of the binding motif resulting in a 4 –column matrix. Such matrices are called Position Specific Scoring Matrices or PSSMs and can be compiled in different formats [288, 296]. PSSMs would also account for the variants arising from the degeneracies and typically accommodates position-specific variability. PSSMs also enable a more quantitative and objective description of the binding specificity, taking into accounts the frequency of each nucleotide at each position of the motif. Matrix-based pattern matching has the advantage of returning a probabilistic description of motif degeneracy [296]. PSSMs are constructed by first aligning the experimentally verified binding (known) sites of the TF. The alignment is then used to derive a count matrix, which in other words, is defined as the numerical count in raw numbers, of every base pair in each position. The count matrix, in turn yields the frequency matrix: which as the name suggests, is an extension of the count matrix, but differs in that the numerical count is replaced by frequency, i.e, the proportion of each residue occurring at a

specific position in relative terms [297, 300-303]**.** The main difficulty of matrix-based pattern matching lies in the choice of appropriate parameters: most programs stipulate the user to specify the matrix width, and the expected number of site occurrences. Since this information is typically not provided, the user has to make guesses, or try various possibilities and select the most convincing result.

Information content of PSSMs

The "goodness" of a matrix is generally estimated by a score (typically dependent on the the information content of the matrix). Various strategies have been developed to optimize the information content of a matrix extracted from a sequence set. Each PSSM has a distinct characteristic which is representative of its ability to classify or distinguish between true positive and false positive binding sites. This characteristic is called the information content of the matrix and it allows one to estimate an upper limit for the expected frequency of the binding sites in random sequences [300, 304]. The information content of the matrix was defined in 1986 by Schneider [304] and was based on a concept called Shannon' entropy. The information content signifies how information is scattered across different sites and to compare one site with another. The information content takes into account inconsistencies or variabilities in individual positions within a series of sites. It is a measure that promotes direct comparisons between sites recognized by different TFs. Basically, the information content of a PSSM accurately illustrates the distinctness of the sites used for the matrix construction from all the possible sequences in the entire genome of the organism in concern. In other words, the information content of a PSSM specifically tells us about the binding interaction between the TF and its binding sites. The information content of a PSSM is calculated from the data on the sequences to which the TF binds and also from the nucleotide content of the organism's genome [304]. There are plenty of algorithms which have been developed for determining the alignment to yield the PSSM with the best information content using various

methods such as expectation maximization [305], Gibbs algorithm [306] and the greedy approach [307].

Weight Score of Predictions

The weight score is an important parameter used for evaluating the validity of the predicted sites. It gives a first-hand idea of the quality of predictions, although, it does not give an indication of the randomness of the result. It gives quantitative and qualitative descriptions of the predictions. Qualitatively, if the weight score is positive, it is highly probable that the prediction is true and if the weight score is negative, the prediction is false. Quantitatively, the magnitude of the score determines the extent to which it could be true. Nevertheless, the quantitative interpretation of the weight score is not so informative and reliable since the score itself depends on the information content of the matrix. This is a major drawback of this assessment measure [300]. In Pattern Matching, the weight score is calculated by taking the log of the ratio between the probability that a sub-sequence of the same length as the number of positions in the PSSM occurs within the context of the PSSM and the background model [296]. Initially the calculation of weight scores was limited to Bernoulli background models only, but was then further extended to accommodate Markov-chain based background models [306, 308].

Background Models

As described in Section 4.1.4.3, the weight score of a predicted site is dependent on the background model. The background model, as the word itself suggests, is a reference and it functions as a base-line for which to compare the PSSM-based predictions. It serves to answer the following question: What shall the probability of occurrence of the TF binding motif in a query gene be compared to? If this question has to be answered, one has to consider a base-line model, which exposes the probability that such a motif occurs randomly in the given context, for example, in the genome of the organism under study [294, 296, 300]. The

background model, in other words, provides significance to the predicted sites, in the light of a given context. The background model as such can be chosen according to the degree of constraints to be imposed. Background models of different markov chain orders can be used to execute this constraint, which takes into account whether there is independence between successive residues or not. In addition to this, there is a constraint based on sequence type. This constitutes our freedom to choose what kind of sequences make up the background model: whole genome or all the regulatory sequences or just the upstream or downstream regulatory sequences, or dual regulatory regions within the coding sequences etc. According to our choice, the content of the background model can be adjusted [294, 296, 309, 300].

Site P-value

In Section 4.1.4.3, it was explained that the weight scores as such do not give a clear insight into the randomness and 'occurrence by chance' of the predictions. This drawback can be rectified only if a measure is available which gives a quantitative appraisal of the randomness. Since the weight scores are determined by the information content of the PSSMs, there is a tendency for some PSSMs to score certain weights more frequently than others. The site-P value, on the other hand, is a unique measure, since it quantifies the randomness from the expected distribution of all the scores.

Pattern matching tool *matrix-scan*

matrix scan, the tool used here for pattern matching scans sequences with Position Specific Scoring Matrices (PSSMs) and scores each position or (a sequence segment) according to a weight score [300, 307]. *matrix scan* was developed by Turatsinze and co-workers in 2008 for searching for patterns in the cis-regulatory regions of genes by using PSSMs which represent the binding specificities of TFs [296]. As explained in Ref.62, the tool used in this work has lots of advantages which outweigh its drawbacks. The most prominent feature of the tool is that it supports the search using one PSSM as well as many PSSMs (which are supported in

various formats), i.e, a single query sequence can be searched for the presence of binding motifs of more than one TF and that too in a single run of the program. In addition, it also supports the prediction of cis-regulatory modules, which are segments of sequences with a high scatter of binding sites. This is very helpful when working with query sequences and promoters whose activation or repression is mediated by an array of TFs rather than just one. The next advantage is that the tool supports a variety of methods to train the background models, which makes it user-friendly. Another plus point for the tool is that the choice of the background models is also flexible: both Markov and Bernoulli models are supported. Moreover, the order of the Markov chains can be changed from an order of 0 to 7 depending on the stringency level desired. One of the most important, if not the most important attribute of the tool is its ability to interpret the reliability of the predictions quantitatively in the form of a P-value. This gives us an indication of the false-positive rate which can occur by chance. Other features which standout include its in-built function to impose thresholds on various other parameters for screening the prediction results. The most notable limitation of the tool is that it takes considerable computing time for performing a whole genome scan [296].

Statistical procedures in Pattern Matching

PM uses PSSMs to scan the regulatory sequences of potential target genes for the presence of binding sites. As a sequence is scanned for the presence of a potential binding sites using a definite PSSM, a certain score is assigned to each sub-sequence (which is a string starting from its respective position of the parent query sequence) The magnitude of the weight score depends on the similarity between the motif represented by the PSSM and the sub-sequence itself. The program *matrix-scan* used in this work to predict binding sites works on the basis of an algorithm which functions by scanning the query sequences with a PSSM by selecting, at each position, a sequence segment of the same length (sub-sequence) and by assigning a score to the sub-sequence. This is performed for every sub-sequence in the query sequence.

Then, all sub-sequences which have scored above a certain threshold score are considered as matches (predicted sites). The score which is used in this algorithm is called the weight score and is defined as the log ratio between the probability for the sub-sequence to occur according to the motif model represented by the PSSM and the probability for the same sub-sequence to occur under the background model. Positive weight scores indicate that a sub-sequence is more likely to be an instance of the motif than an instance of the background [295, 296, 300].

Since the set of sites given as the output depend on the weight score cut off, it can be seen that the threshold imposed on the score has such a big influence on the outcome. But this threshold is very often chosen as an ad-hoc measure or as a very educated guess. Besides, the expected distribution of the weight scores is a direct function of the information content of the PSSMs used and this would mean that certain weight scores would be meaningful to some PSSMs and not for others. Hence, to derive a randomness measure, the expected frequencies of the weight scores are calculated to obtain the expected distribution of all the scores, from which the site P-values are determined. The P-value is defined as the probability to get a given score randomly, and can be used as a good estimate of the rate of false-positives and hence is more informative than the usual weight scores [295, 296, 300]. As an example, a P-value threshold of 0.01 would mean that one false positive prediction is expected for every 100 bps, whereas a P-value of 0.000001 would mean that one false positive prediction is expected for every million bps. Setting the threshold on the P-value rather than the weight score is even more crucial when sequences are scanned with multiple PSSMs. Indeed, each motif has its own size and information content, which critically influences the expected distribution of weights [295, 296, 300].

The *matrix scan* tool supports background models of many types: Bernoulli models that assume independence between successive residues, as well as higher order Markov chains, where the probability to find a residue at a given position depends on the residues found at m

preceding positions (m is the order of the Markov chain). Markov models of order m determine the frequencies of words of length k = m + 1. A Markov chain of order 0 corresponds to a Bernoulli model [295, 296, 300]. The purpose of the background model is to estimate the probability of occurrence of a sequence motif outside the potential target motif [295] and this will have a significant bearing on the final outcome and hence has to be chosen in accordance with the type of query sequences.

Reliability adding measures

As explained before, the predictions from the binding site analysis always have matches which are true binding sites for the TF in addition to mismatches which are disguised as true sites due to the statistical nature of the tool. Since it is an inevitable fact due to the very nature and statistical approach of Pattern Matching, certain considerations have to be made to make the predictions more reliable. Reliability to a certain degree is provided by the randomness measure: the P values of the predicted sites. Apart from this, the actual significance of the predictions can be increased if as much biological information can be incorporated as possible. Hence, some measures are proposed for improving the significance of the predictions.

PSSM-specific optimized P-values

As a critical reliability indicator, the P-value gives a measure of the randomness of any prediction. Since the predictive capacity of each PSSM (due to its varying information content) differs from case to case, a generic or common P-value could not be used to screen out "putative-target hits" (i.e genes or operons harboring binding sites predicted by the corresponding PSSMs). Hence, it is imperative to determine a strategy by which customized and optimized P-values specific for every PSSM are objectively calculated.

Presence of site repeats

Transcriptional regulation, be it activation or repression, relies also on homo-co-operativity, which is basically, the modulation of regulation mediated by the cumulative action of many copies of the same TF. This requires the presence of TF binding site repeats some of which may or may not overlap. The presence of multiple binding sites for the same TF can indicate possible homo-co-operative combinatorial regulation [310, 311]. The detection of multiple predicted TF binding sites in the promoter elements of a gene or operon increases the chance of the latter being a true positive target of the TF.

Chapter 3. Materials and Methods

3.1 Generation of time series microarray data

The transcriptomic measurements were obtained from the BioInSys project partner at HZI (working group of Prof. Dr. Irene Wagner-Döbler), Braunschweig. Static biofilm cultures of *S. mutans* UA159 were grown in routine 24 well plates (Greiner Bio One, Frickenhausen, Germany) using 800µl Todd Hewitt Broth (Becton Dickinson, Heidelberg, Germany) supplemented with 0.5% sucrose (Sigma, Taufkirchen, Germany) per well as media. Overnight cultures of planktonic cells in Todd Hewitt Broth were diluted in the biofilm medium to a starting OD of 0.05 and biofilms were grown for 3 hours under anaerobic conditions (80% N_2, 10% CO_2, 10% H_2) in a chamber (Don Whitley Scientific, Shipley, England). Thereafter, the supernatants were completely removed and exchanged with fresh medium (control samples) or either fresh medium supplemented with 2.5 µg/ml carolacton (treated samples). For the control, an equivalent volume of methanol was added. Samples were taken after 5, 20, 40, 60, 80, 100, 120, 160, 200, 240 and 300 minutes by removing the supernatant and adding 200 µl fresh medium and 400µl of RNA protect (Qiagen, Hilden Germany). Biofilms representing one condition were scraped from at least 3 individual wells and pooled together. Cells were pelleted by centrifugation, washed with Molecular biology grade water (Sigma, Taufkirchen, Germany) and lysed with a combination of Lysozyme/mutanolysin treatment and mechanical disruption of cells with glass beads [10].

Total RNA was extracted using the RNeasy Kit (Qiagen, Hilden Germany) as recommended by the manufacturer. The overall experiment was repeated once. 2 µg of total RNA from each sample was labeled with Cy3 and Cy5-ULS (Kreatech, Netherlands) according to the protocol of the company. Fragmentation of RNA and further processing was performed according to

This chapter has been adapted from the publication P. Sudhakar, M. Reck, W. Wang, F. He, I. Wagner-Döbler, AP. Zeng. Construction and verification of the transcriptional regulatory network of the human dental pathogen *Streptococcus mutans* upon treatment with the biofilm inhibitor Carolacton. BMC Genomics 2014, 15:362.

the Agilent hybridization kit procedure. Samples were hybridized in a dye swap design on 8x15K *S. mutans* custom arrays (Agilent, Böblingen, Germany) for 14h at 65°C as previously described [10]. The array design is described in detail in [10]. Slides were washed and scanned as recommended by Agilent. For feature extraction, the Agilent extraction software (V. 10.7) was used. The further data analysis was performed using the R package LIMMA and MATLAB. For dye specific within-array normalization, the LOWESS algorithm was applied. Quantile normalization was used for between array-normalization. Genes corresponding to rows with blanks were omitted. From the log transformed and background corrected median signal intensities, the fold change between treated and untreated samples was calculated for every gene at each time point. The log2 (treated/untreated) expression ratios were then subjected to a within-gene normalization. Genes with a log-fold ratio greater than 0.8 were regarded as genes corresponding to transcripts with altered abundances (GTAAs). The time-series microarray data set generated as part of this study is available in the Gene Expression Omnibus (GEO) repository [accession id GSE53264].

3.2 Functional enrichment analysis of GTAAs

Enrichment analysis was performed separately for both the up- and downregulated GTAAs (at every measured time point) with respect to known biological functional classes and KEGG metabolic pathways in order to gain an overview of the various functional categories affected by carolacton treatment. 15 main and 97 sub functional classes of *S. mutans* UA159 were compiled using the gene annotation information furnished in the Oralgen database (version 2011 and now re-named as The Bioinformatic Resource for Oral Pathogens). The gene membership information of *S. mutans* UA159 corresponding to 84 different pathways was retrieved from the KEGG database (version 2011). The hypergeometric test with Benjamini Hochberg correction was used to quantify the significance of the overlap between the GTAAs and the functional categories. A corresponding significance score was also calculated as a

negative logarithmic function of the corrected P value. The significance score is a user friendly measure of the relevance of the overlap between functional categories and the GTAAs. The higher the significance score, the more relevant is the overlap. Categories with significance score greater than 0 represented significant enrichment. Functional categories were considered to be most significantly over represented if the significance score was higher than 1.

3.3 Inferring gene-gene correlation relationships from transcriptomic data using the Trend Correlation (TC) method

The TC method [22] allows the inference of gene-to-gene co-expression "associations" or gene-pairs based on extracting the main features of the change trend and the correlation of gene expression between consecutive time points [22]. Time lagged co-expression relationships are those in which the expression patterns of two genes are better correlated when a time shift is introduced [22]. In other words, the expression pattern of one gene in a gene-pair might be better correlated (determined by the TC method algorithm upon providing a desired time shift requirement) to that of the second gene when an introduced time-shift classified one of the genes as leading and the other as trailing. In this work, the TC method was adapted by calculating the correlation coefficient from the change rates between consecutive time points since the time interval of the measurements is not constant (personal communication Feng He). Potential co-expression relationships were checked for every possible regulator-gene pair given the minimum number of expression measurements in the chosen time window was four. The genetic response network consisted of optimal expression relationships which started either at 0, 5 or 20 min post carolacton treatment in order to capture the correlations mediating the initial response to carolacton. The significance of correlation of a gene-pair for every time-window is measured by the co-expression P-value by procedures described previously in the TC method by He et al. [22]. The individual P-values

of correlation for a gene-pair within every considered time-window are compared among each other. Optimal correlation for a gene-pair refers to the time window within which the correlation between the genes is most significant statistically. Parameters such as the correlation co-efficient, the trend score, and the P-value were calculated as described previously [22] to characterize the degree and randomness of correlation between the temporal expression patterns of genes. The association between two genes was considered to be statistically significant if the corresponding P-value of their co-expression was determined to be lower than the optimal P-value cutoff calculated using standard performance metrics as explained below.

3.4 Performance metrics for the determination of the optimal P-value threshold

The optimality of the P-value threshold was determined based on the F-score (eq.1) which in turn is a function of the equilibrium between the Recall (eq.2) and Precision (eq.3) functions which are defined below.

- F-score: this is predictive of the balancing property or equilibrium between the Precision and Recall parameters. The F-score usually reaches a maximum peak at the P-value optimum.

$$F - score = \frac{2 * Precision * Recall}{Precision + Recall} \quad (1)$$

- Recall (also termed as True Positive Rate or Sensitivity) : defined as the fraction of true "associations" which are inferred as significant by the TC method

$$Recall = \frac{Number\ of\ true\ positives}{Number\ of\ true\ positives + Number\ of\ false\ negatives} \quad (2)$$

- Precision (also termed as Positive Predictive Value) : defined as the fraction of inferred significant "associations" which are true

$$\boldsymbol{Precision = \frac{Number\ of\ true\ positives}{Number\ of\ true\ positives + Number\ of\ false\ positives}} \quad (3)$$

Gene-pairs within the same operon because of their virtue of coherent co-expression were used as the "gold standard" or true positives. False positives constitute the inferred significant "associations" between genes from different operons whereas false negatives represent the intraoperonic gene-pairs which were considered as insignificant and which were seemingly missed out at the particular P-value threshold. The operon information was predicted by Pathway Tools [312]. Differential edges [313, 314] in response to carolacton treatment were pooled together to form the final co-expression network.

3.5 Construction of Position Specific Scoring Matrices (PSSMs) representing transcription factor binding motifs

Experimentally verified DNA binding sites of *S. mutans* transcription factors were retrieved from literature when available. The comparitive genomics based approach using the concept of regulogic extrapolation as explained in [216] was carried out to retrieve orthologous regulatory relationships in *S. mutans* from experimentally verified relationships in *B.subtilis*. Validated binding sites of *B.subtilis* transcription factors were retrieved from the DBTBS database [214]. OrthoMCL [315], which is a genome-scale algorithm for grouping matching protein sequences, was used to identify orthologous genes shared between *S. mutans* UA159 and other gram positive bacterial species including some representative species from the Streptococcus genus. The OrthoMCL results were verified by manually searching for the presence of commonly occurring domains in the protein sequences of the orthologous pairs of transcription factors.

Table 2. The list of *S. mutans* UA159 transcription factors and their binding motif information compiled in this study. DNA binding motifs for 44 transcription factors were mined out based either on the experimentally verified binding sites in *S. mutans* or using comparative genomics approaches employing the concept of regulogs [216]. *When binding sites were extrapolated regulogically from *B.subtilis*, the information such as sequence conservation, E-value, and commonly present domains (according to SMART database) between the *B.subtilis* transcription factor and its ortholog in *S. mutans* UA159 are shown. Sites from the regulogs in other species within the Streptococcus genus were used to complement the regulogic sites in *S. mutans* so as to improve the predictive power and information content of the resulting PSSMs. PSSMs from the REGPRECISE database [231] were derived from already enriched sets of regulogic sites. °indicates spaced motifs for which motif width is not a suitable indicative parameter. ^The optimal P-value and optimal weight score for PSSMs was determined from the *matrix-quality* tool [316] analysis and subsequently used for pattern matching. For PSSMs for which the optimal P-values could not be determined as a result of the low complexity of the motifs, a string-search based tool called *genome-scale dna pattern* was used to determine putative sites. These transcription factors are denoted by 'DP' under the optimal P-value and optimal-weight score columns. TF: Transcription factor. BS: *Bacillus subtilis*. AAA (SMART database accession number SM00382): Domain found in ATPases associated with a variety of cellular processes; Bac_DnaA_C (SMU00760): Domain representing the C-terminal regions of bacterial DnaA proteins; PDZ (SM00228): Domain representing distinct regions found in diverse signalling proteins; REC (SM00448): Type of response regulator domain; HTH_LUXR (SM00421): DNA-bending helix-turn-helix domain present in transcriptional regulators of the LuxR/FixJ family of response regulators. HTH_MERR (SM00422): Domain found in transcriptional regulators which bind DNA via a helix-turn-helix structure and which mediate mercury-dependent induction of the mercury-resistance operon. CoA_binding (SM00881): Domain found in a number of proteins including succinyl CoA synthases, malate and ATP-citrate ligases. HTH_ARSR (SM00418): DNA-binding, winged helix-turn-helix (wHTH) domain present in transcription regulators of the arsR/smtB family and involved in stress-response to heavy metal ions; Trans_Reg_C (SM00862): Domain found in the C-terminal regions of response regulator proteins.

The list of *S. mutans* UA159 transcription factors and their binding motif information compiled in this study

TF name	Locus tag of *S. mutans* UA 159 TF	TF description	Source of motifs*	Motif width	IUPAC TF binding motif consensus	Total information content of PSSM	Information content per column	Optimal P-value^	Optimal weight score^
AdcR	SMU_1995c	Putative transcriptional regulator of zinc homeostasis	Regprecise	16	amtTaACYRG TtaAkt	12.0655	0.75409375	1.80E-05	8.9
ArgR	SMU_2097	Transcriptional regulator of arginine biosynthesis and degradation	Regprecise	16	ttKtatawwwAta Maa	7.78467	0.486541875	2.40E-05	9
BirA	SMU_1578	Biotin biosynthesis regulator	Regprecise	NA°	ATaGTTAAC-(16)-GTTAACtAa	--	--	DP	DP
CcpA	SMU_1591	Catabolite control protein A.	Regprecise	16	wtGaaArCGyTt tCaw	9.59376	0.59961	2.60E-04	8
CiaR	SMU_1129	Two component response	Regprecise	NA°	TTTAAG-(5)-aTTAAG	--	--	DP	DP

		regulator							
CodY	SMU_1824c	Transcriptional repressor of amino acid metabolism	Regprecise	15	aattTTCwGAawaTT	10.6889	0.712593333	4.93E-04	5.1
ComE	SMU_1917	Response regulator of the competence regulon	Ref [97]	11	rGGaTTkrCCa	9.49081	0.862800909	9.96E-05	5.8
CopY	SMU_424	Negative copper transcriptional regulator	Regprecise	20	tmRtytACAaatGTArayww	12.9104	0.64552	4.08E-06	11
CtsR	SMU_2030	Putative transcriptional regulator of heat shock response	Regprecise	24	twwkkYttGaCYaktyytGaCCaa	11.2163	0.467345833	1.70E-05	10.6
CysR	SMU_852	Transcriptional regulator of cysteine metabolism	Ref [352]	13	TATCASmGyGaTa	10.8424	0.834030769	4.12E-05	8
DnaA	SMU_01	Chromosomal replication initiator protein	DBTBS (BS, DnaA, 44%, 5e-128, AAA, Bac_DnaA_C)	9	ttwtCcACA	--	--	DP	DP
FruR	SMU_870	Transcriptional repressor of the fructose operon	Regprecise	10	tTGmtwGamw	6.17237	0.617237	4.7E-05	8.4
FurR	SMU_593	Putative ferric uptake regulator protein	Regprecise	15	TtakAatsatTmTaa	9.98815	0.665876667	5.76E-05	8.9
GalR	SMU_885	Galactose operon repressor	Regprecise	19	aaaatTTTASTAAAatttt	11.0203	0.580015789	1.62E-05	12.5
GlnR	SMU_363	Glutamine synthetase repressor	Regprecise	19	aTGTyAkrwaaymTrACAt	13.4055	0.705552632	9.95E-05	5.9
HomR	SMU_930c	Transcriptional regulator of homocysteine metabolism	Ref [352]	15	TaTaGCyaaCyaTCa	13.2191	0.881273333	7.94E-06	10.1
HrcA	SMU_80	Heat inducible transcription repressor	Regprecise	27	TTAGCaSTCkwkkdamaaGAGTGCTAA	22.2571	0.824338	1.91E-05	5.9
HtrA	SMU_2165c	HtrA transcription factor	DBTBS (BS, HtrA,38%, 3E-75, PDZ)	8	TTTYCaCa	--	--	DP	DP
LacR	SMU_1498	Lactose repressor	Regprecise	9	aaaCaaaaa	--	--	DP	DP
LevR	SMU_1964c	Two component response regulator	DBTBS (BS, Des,31%,6E-28,REC,HTH_LUXR)	17	aTTTTtymkwwmwkyrt	--	--	DP	DP
MalR	SMU_1566	Putative maltose operon transcriptional repressor	Regprecise	20	ttayGCAArCGyTTGCrywa	15.3079	0.765395	3.27E-05	10
MbrC	SMU_1008	Cell envelope stress response regulator	Ref [95]	14	TTaCamtttTGTaa	11.074	0.791	1.38E-05	12.1
MleR	SMU_135c	Regulator of malolactic fermentation	DBTBS (BS, HexR, 22%, 3E-10)	15	aTCTaRyawaaKRtG	--	--	DP	DP
Mta	SMU_1790c	MerR family transcriptional	DBTBS (BS, Mta, 37%, 7E-44,HTH_MERR)	23	GaCYCtMmCSymSSGtyatrGs	14.3604	0.624364	2.08E-06	9.4

		regulator			C				
MtaR	SMU_1225	Putative transcriptional regulator of methionine metabolism	Regprecise	17	tATAGtttmaarCTATA	12.2283	0.719311765	6.03E-05	7.5
NagR	SMU_1065c	GntR family transcriptional regulator of N acetylglucosamine utilization	Regprecise	20	maatwGGwMTAtACCawttw	13.5623	0.678115	1.84E-05	9.1
NmlR	SMU_677	MerR family transcriptional regulator involved in carbonyl stress response	Regprecise	19	vCTTGaCTTGRaGTYvaCT	16.8334	0.885968421	8.12E-06	8.3
NrdR	SMU_1923c	Transcriptional regulator of ribonucleotide metabolism	Regprecise	16	aCaCaAtATmTtGtGt	11.4352	0.7147	1.60E-05	10.9
PdxR	SMU_953c	Putative transcriptional regulator of pyridoxine biosynthesis	Regprecise	NA°	TATTGTA-(18)-TACAATA	--	--	DP	DP
VicR	SMU_1517c	Two component response regulator	DBTBS (BS, YycF, 71%, 4E-114, REC, Trans_reg_C)	17	TGTamTrryGmYGkmak	9.17903	0.539943	1.41E-05	9.7
PflR	SMU_491	Putative transcriptional regulator of formate metabolism	Regprecise	19	MGwAAywrawtyaaTTwCG	11.7093	0.616278947	2.03E-05	8.4
PipR	SMU_2134	Putative transcriptional regulator	Regprecise	19	TaTraTmtAtTwtAtYatA	12.1798	0.641042105	9.34E-05	7.5
PurR	SMU_356	Purine operon repressor	DBTBS (BS, PurR, 51%, 2E-96)	8	aaarCGaa	--	--	DP	DP
PyrR	SMU_856	Pyrimidine operon regulatory protein	DBTBS (BS, PyrR, 54%, 1E-63)	31	CCTTTaaCmCTGTCCCGTGaGGCaGGCaaGG	35.6892	1.15127	2.17E-06	16.7
Rex	SMU_1053	Redox-sensing transcriptional repressor	DBTBS (BS, Rex, 45%, 3e-62, CoA_binding)	22	TTaGYSwMTTTYTYwwCwaaRR	15.6870	0.713044	7.84E-05	6.3
RpoD	SMU_822	RNA polymerase sigma-70 factor	DBTBS (BS, SigB, 25%, 1E-14)	NA°	ttGACa-N17-tatRat	--	--	DP	DP
ScrR	SMU_1844	Sucrose operon repressor	Regprecise	20	tmtGyCAarCGytTGrCaka	15.5795	0.778975	1.23E-05	9.5
SdpR	SMU_1647c	Predicted transcriptional regulator	DBTBS (BS, SdpR, 51%, 7E-27, HTH_ARSR)	7	TCTAAAT	--	--	DP	DP
SloR	SMU_186	Metal dependent transcriptional regulator	Ref [18, 378]	22	maaagamwsytksSttatTtkS	8.47192	0.385087	8.46E-05	8.2
SMU_1193	SMU_1193	Putative transcriptional regulator of transporter genes	Regprecise	22	tTGTayyAtataawtaRtACAa	13.9978	0.636263636	3.10E-05	8.9
SMU_1349	SMU_1349	TetR family trancriptinal regulator	Ref [346]	16	TatTSTtTwwTRwTwa	--	--	DP	DP

SMU_2125	SMU_2125	Predicted transcriptional regulator	DBTBS (BS, HxlR, 36%, 8E-17)	14	CaCCyayvrTTTTt	--	--	DP	DP
SMU_289	SMU_289	Putative transcriptional regulator of ascorbate utilization genes	Regprecise	22	tTGaaCaCadrMSTGTGtKwYa	15.8219	0.719178	1.03E-05	8.6
TreR	SMU_2040	Repressor of the trehalose operon	Regprecise	20	MMaaStTGYmkrCaaStTKK	13.2721	0.663605	4.08E-05	8

To increase the reliability of a particular predicted regulog in *S. mutans*, a check was performed to ascertain if the regulog was conserved in at least some species within the Streptococcus genus. Consensus motifs constructed from too few binding sites have a low predictive value [316]. Hence the available binding sites were enriched with corresponding putative sites from regulogs identified from different species within the Streptococcus genus. In addition to the PSSMs constructed from manually identified regulogs, PSSMs were also compiled from already assembled regulogic sites retrieved from the RegPrecise database [231] and from phylogenetic footprinting studies (see **Table 2** for more information). Each resulting set of binding sites was submitted to *consensus*, which is a tool for mining out conserved motifs. The tab-formatted PSSMs were then converted into TRANSFAC formatted PSSMs which enable a user-friendly organization of pattern matching results.

3.6 Determination of optimal P-values for every PSSM

An online Regulatory Sequence Analysis Tool (RSAT) termed *matrix-quality* [316] which objectively determines the optimal P-value for every PSSM was used. Medina-Rivera et al [316] showed that the optimal P-values for each PSSM could be determined from the divergence between the theoretical and empirical score distributions which in turn indicated the ability of the PSSM to recognize high-scoring putative binding sites. These divergences were captured by plotting the expected and empirical score distributions of predicted sites in the whole set of upstream non-coding sequences for every PSSM, and for the randomly permuted version of the PSSM [316].

3.7 Prediction of putative binding sites by Pattern Matching

The TRANSFAC formatted PSSMs were then employed to search for putative sites using the tool *matrix-scan* (http://rsat.ulb.ac.be/) [296, 317] with the determined optimal P-values set as the cut-offs. Putative sites were searched for in the upstream regions (1000 bps upstream of the start codon with no overlap with the coding regions of the preceeding ORF) of all genes in the *S. mutans* UA159 genome. A markov order of 0 and a background model comprising *S. mutans* specific upstream regions (no overlap with ORF coding regions) was used during the pattern matching scan. The resulting sites with a P-value smaller than the determined optimal P-value were considered as significant. In cases where the divergence between the theoretical and empirical score distributions was not observed due to the inherent low complexity of the motifs [316], a whole genome string-based site identification was performed using the non-matrix based tool genome-scale *dna pattern* (http://rsat.ulb.ac.be/) [317] with an allowance for a maximum of 2 mismatches. Both the forward and reverse strands were subject to the pattern matching procedure. Genes with predicted binding sites were considered to be putative target genes if the predicted sites were detected on any one of the two strands.

3.8 Biological context within the TRRN sub-networks

Even if the expression profiles of not all the genes in the same operon are well correlated with the expression of a certain regulator gene, we assume that all the genes in the same operon will be regulated by the regulator provided that the corresponding binding sites are found in the upstream regulatory regions of the operon. This is what we refer to as operon-oriented adjustment. To arrive at the TRRN, the co-expression network was superimposed onto the regulator-target gene binding site map. The genes in each sub-network of the TRRN were analysed to detect any significant over-representation of categories such as biological functional classes, KEGG metabolic pathways and gene ontology terms. The hypergeometric

distribution was employed to determine over-represented categories with the Network Analysis tool *compare-classes*. Self and reciprocal comparisons were avoided. Enrichment events with significance scores >=0 were considered as statistically significant. Gene ontology information was retrieved from the Pathosystems Resource Integration Center (PATRIC) [318]. All the network visualizations were performed using Cytoscape [319]. Network motifs were analysed using the tool MFINDER [320]. The significance of the discovered motifs in the real network was determined in comparison to their occurrences after 10000 randomizations (at a P-value <=0.01). Heat map representation of metabolic pathways was generated using the Mayday visualization tool version 2.12 [321].

3.9 Heterologous expression and purification of response regulator proteins

Coding sequences corresponding to the *S. mutans* response regulators MbrC (SMU_1008) and VicR (SMU_1517c) were PCR-amplified with Phusion-Polymerase (NEB) using primer pairs MbrCFor/MbrCRev and VicRFor/VicRRev respectively (**Table 3**). Homologous flanks to the vector sequence of pET28c were implemented at the 5` ends of the primer. The purified DNA was subsequently cloned in expression vector pET28c (Novagen) cut with NcoI/XhoI (Fermentas) using the CloneEZKit (Genescript). Resulting plasmids bearing the coding sequence for a N-terminal His-Tag were verified by sequencing and transformed in BL21Star. Cells were grown in 4l shaking flaks, induced with 1mM IPTG at an OD_{600} of 0.5 and harvested 4 hours after induction. The resulting cell pellet was re-suspended in lysis buffer (50mM Tris, 150 mM NaCl, pH 7.4) supplemented with 4 mg/ml lysozyme (Sigma) and frozen at –20°C. After de-freezing the pellet, 100µg DNAse I (Roche) and 5mM $MgSO_4$ (Sigma, final concentration) were added and 5 cycles of sonification for 60 seconds were applied to lyse cells (duty cycle 0.5 Sec, pause 1 sec, amplitude 60%) [380] (courtesy Dr. Reck, HZI, Braunschweig).

Table 3. Primers used for the PCR-amplification of the coding sequences of the response regulators MbrC and VicR of *S. mutans* as well as for the amplification of the potential MbrC binding sites in the promoter regions of known and predicted targets

Primer	Nucleotide sequence (5`- 3`)	Purpose
MbrCFor	taagaaggagatataccatggctatgctaaagcaagaaaaaatttac	Expression of MbrC
MbrCRev	gtggtggtggtggtgctcgagtttaattaaataccctactcctctt	Expression of MbrC
VicRFor	taagaaggagatataccatggctatgaagaaaattctaatcgttgacg	Expression of VicR
VicRRev	gtggtggtggtggtgctcgaggtcatatgatttcatgtaataacca	Expression of VicR
EM1006F	gagctgtaaatttctcaggcttct	Forward primer for amplification of MbrC binding site in the SMU_1006 promoter (positive control)
EM1006R	ttgatttctaaaaacatagcttctcc	Reverse primer for amplification of MbrC binding site in the SMU_1006 promoter (positive control)
EM610F	tttgtgctttagaattaatgttgga	Forward primer for amplification of potential MbrC binding site in the SMU_610 promoter
EM610R	ttttacttttacgaaaaccgtaagtt	Reverse primer for amplification of potential MbrC binding site in the SMU_610 promoter
EM718F	ttgcaagcaagcaataaatatga	Forward primer for amplification of potential MbrC binding site in the SMU_718 promoter
EM718R	ttagcatcttggatagcagtgta	Reverse primer for amplification of potential MbrC binding site in the SMU_718 promoter

Finally 0.1% NP40 (Sigma, final concentration) and 10 mM imidazol (Sigma, final concentration) were added and cell debris was removed by centrifugation and filtration (0.45 µm). His-tagged proteins were purified on a HIS-Talon (Clontech) column using a Duo Flow FPLC (Biorad) at a flow rate of 0.5 ml/h. Resulting purified proteins were desalted using PD10 columns (Amersham), eluted in 1xPBS buffer and measured using a spectrophotometer (Nanodrop ND1000, Peqlab) to determine the protein concentration. The response regulator was over 99% pure as confirmed by SDS-PAGE (data not shown). Promoter regions

containing the potential binding sites of MbrC were PCR amplified using Taq-polymerase. The primers used are listed in **Table 3**. PCR-products were purified using the PCR-Purification Kit (Qiagen) and used as targets for EMSA [380] (courtesy Dr. Reck, HZI, Braunschweig).

3.10 Deletion of potential binding sites

Potential binding sites of the response regulator MbrC were deleted using a PCR-driven overlap extension method. Motifs in the promoter regions of SMU_610, SMU_718, and SMU_1006 were deleted by two initial PCRs generating overlapping DNA sequences. Taq polymerase (Qiagen) was used for the amplification of regions flanking the potential binding sites. PCR-products were purified from an agarose gel using the Gel Extraction Kit (Qiagen) and used as template for a third PCR with Phusion polymerase using primers (**Table 4**) spanning the whole region. Purified PCR-products were cloned blunt-end in the EcoRV restriction site of vector pGEM5Zf (+) (Promega). Resulting plasmids were verified by sequencing and used as template for a PCR to finally generate the DNA-targets for EMSA [380] (courtesy Dr. Reck, HZI, Braunschweig).

Table 4. Primers used for the deletion of MbrC binding sites

Primer	Nucleotide sequence (5`- 3`)	Purpose
610Del1	tttgtaatataaattgatttgatttgtcaaatttttgcttg	Deletion of potential MbrC site in SMU_610 promoter
610Del2	tcaaatcaatttatattacaaaaatatatttggaagattta	Deletion of potential MbrC site in SMU_610 promoter
718Del1	ttcctaatttctctatcttactaaaaattagcttatttgtc	Deletion of potential MbrC site in SMU_718 promoter
718Del2	tagtaagatagagaaattaggaaccagtatgtcaatta	Deletion of potential MbrC site in SMU_718 promoter
1006Del1	ttaatgtcaattacgattctttaagtg	Deletion of MbrC site in SMU_1006 promoter
1006Del2	aaagaatcgtaattgacattaatttaat	Deletion of MbrC site in SMU_1006 promoter

3.11 EMSA procedure

The response regulator (MbrC and VicR) proteins were activated by acetylphosphate (Sigma). An aliqot of the protein was incubated for 2h at RT (room temperature) in reaction buffer (25mM acetylphosphate, 50mM Tris-Hcl, 50mM KCl, 10mM $MgCl_2$, 4mM dithiothreitol) as described previously. Excess of acetylphosphate was removed by filtration and the protein was serially diluted in binding buffer (10 mM Tris, 1mM EDTA, 100mM KCl, 100µM DTT, 5% vol/vol glycerol, 10µg/ml BSA, pH 7.5). 0.5 pmol of target and competitor DNA was added to each reaction and incubated for 1h at RT. The unrelated response regulator VicR which has a similar MW and pI as the response regulator MbrC under study was used as a negative control to rule out unspecific DNA binding. 4µl of the reaction mixture was applied on a 5% acrylamide gel run in Tris-borate-EDTA (TBE) buffer at pH 7.4. Gels were stained using SybrGold and visualized in a transilluminator (Alpha DigiDoc, Biorad) at 254 nm [380] (courtesy Dr. Reck, HZI, Braunschweig).

3.12 Construction of gene deletion mutants and a *cysR* complementation strain

Upstream and downstream flanking regions of *cysR* were PCR-amplified using primers CysR P1/2 and CysR P3/4 (**see Table 5** for the primers used to generate deletion mutants) and genomic DNA of *S. mutans* UA159 as template. The erythromycin resistance cassette was amplified from genomic DNA of a previously constructed mutant [10]. Restriction sites of AscI and FseI were introduced via the 5`-termini of the primers. After restriction digestion of purified PCR-products with the appropriate restriction enzymes, the up and downstream flanks were ligated to the ERM cassette using T4 DNA Ligase and directly transformed in the *S. mutans* UA159 WT strain according to the procedure of Li et al [134]. Gene deletion

strains were selected on THBY agar plates containing 10µg/ml erythromycin. Genomic DNA of isolated clones was analysed with PCR and primers P1 and P4 spanning the entire gene deletion construct. Sequencing was used to further confirm the correct insertion of the gene deletion construct via double homologous recombination. For the generation of four other gene deletion strains (*glnR*, *fabT*, *rgg* and *spxA*), a similar equivalent procedure was carried out [380] (courtesy Dr. Reck, HZI, Braunschweig).

Table 5. Primers used for constructing the gene deletion mutants

Primer	Sequence	Function
CysR-P1	TGTCAAACGCGAGGTGTTAG	*cysR* deletion
CysR-P2	GGCGCGCCCGAAAGGTACCGCTGTTAGC	*cysR* deletion
CysR-P3	GGCCGGCCCTTATCCACGTGTGCTCTCAAA	*cysR* deletion
CysR-P4	GCCTGATGTGCTTGATCATTT	*cysR* deletion
FabT-P1	TTCGCATTGCAGAACTTGTC	*fabT* deletion
FabT-P2	GGCGCGCCGTTTTCAAGCTCATCTCTTCGAT	*fabT* deletion
FabT-P3	GGCCGGCCTGGTGCTTTGGAGAAGGGGTTA	*fabT* deletion
FabT-P4	TTCCATTTTCGACACATTCG	*fabT* deletion
Rgg-P1	CACGTAAGAGCAACAATCTAGCC	*rgg* deletion
Rgg-P2	GGCGCGCCTTCATTTTTAGACCGCGTGA	*rgg* deletion
Rgg-P3	GGCCGGCCAGCAATGAAATTGGCAGGGAGT	*rgg* deletion
Rgg-P4	GACATCACTTTCATTTGGAGGA	*rgg* deletion
SpxA-P1	TGGGCTTCATAGAGGGCATA	*spxA* deletion
SpxA-P2	GGCGCGCCCTCGCTTTACGGCAGCTTGTA	*spxA* deletion
SpxA-P3	GGCCGGCCTCGCAAACAAGAACTGCGTCAA	*spxA* deletion
SpxA-P4	CAGCAGCATAGTCCCAAGGT	*spxA* deletion
GlnR-P1	TAGCAAAGGGGTGGATTTGT	*glnR* deletion
GlnR-P2	GGCGCGCCATAGGAAAAACCGCCATTGA	*glnR* deletion
GlnR-P3	GGCCGGCCACGCAACATCTTGGTGGTTTA	*glnR* deletion
GlnR-P4	GGAGCTTCATAACCAGGAACC	*glnR* deletion

C-CysR1-F	GTGGCAGGTTATGTTAATAAAGGCGAA	Complementation *cysR*
C-CysR1-R	CCTAGACGAATATATTTGACAACAAAAAACTTG	Complementation *cysR*
Q-CysR1-F	CCTTTGCTGACTGCCTTTTC	Quantitative RT-PCR
Q-CysR1-R	GTGTCCTTGGGCAACTTCAT	Quantitative RT-PCR
Q-CysR2-F	CTTTGCTGACTGCCTTTTCC	Quantitative RT-PCR
Q-CysR2-R	TCGGTCGGTCACATTGTAAA	Quantitative RT-PCR
Q-CysR3-F	TTGTCAGTCAGCCCAGTTTG	Quantitative RT-PCR
Q-CysR3-R	TCGGTCGGTCACATTGTAAA	Quantitative RT-PCR
ComE-F	GCGATGGTCGACAATTATATAATCAATTGACAACGGC	RT-PCR positive control
ComE-R	GCGATGGTCGACTCATTTTGCTCTCCTTTGATCAG	RT-PCR positive control
ComX-F	ATTCCGGCATAGCTCAGTTG	RT-PCR positive control
ComX-R	GCGATGCATATGTTTTAGCCGGAGCTTTTTCA	RT-PCR positive control

RNA was isolated from the exponentially growing *cysR* gene deletion strain and was verified for the absence of *cysR* mRNA with RT-PCR using the 3 primer pairs Q-CYSR1-F/R; Q-CYSR2-F/R; Q-CYSR3-F/R. Primers specific for the ComX (ComX F/R) and ComE (ComE F/R) encoding genes were used as positive controls for the RT-PCR reaction. Reactions in which the reverse transcriptase was replaced with water functioned as negative controls. For the complementation of the *cysR* gene deletion strain, the entire *cysR* gene including its promoter region were PCR-amplified with primers C-CYSR1-F and C-CYSR1-R and cloned blunt end via the SmaI restriction site in the replicative plasmid pDL278 [322]. Resulting plasmids were verified by sequencing and transformed in the *cysR* gene deletion strain as described elsewhere [134] (courtesy Dr. Reck, HZI, Braunschweig).

3.13 Viability measurements of the carolacton treated *cysR* deletion strain using Live/Dead viability staining and by Cfu determination

Viability measurements using live/dead viability staining of carolacton treated and untreated biofilms of the *S. mutans* UA159 wildtype, gene deletion and complementation strains were essentially performed as previously described [10]. Biofilms of *S. mutans* UA159 WT were used as positive control. The experiment was repeated in two biological replicates. For the cfu determination, biofilms representing one condition were harvested and pooled from 5 individual wells. Chains and biofilm clumps were dispersed and destroyed using mild ultrasonification conditions as described before [323] and plated on THBY-agar plates (containing 10µg/ml erythromycin for the mutant) in 3 technical replicates. Live dead staining before and after sonification verified that the sonification had no significant influence on cell viability. The overall experiment was repeated in 2 biological replicates [380] (courtesy Dr. Reck, HZI, Braunschweig).

Chapter 4. Results: Network construction and analysis

4.1 Dynamic genomewide transcriptome profiling

A high-resolution transcriptomic time-course experiment with *S. mutans* biofilms (with 11 samples taken from 5 to 300 min post carolacton treatment) was performed to capture the temporal response over an extended period of time. Our time series dataset included a sample taken at 5 min post carolacton treatment and is expected to identify genes whose expression patterns were immediately affected by carolacton. A total of 772 GTAAs (gene transcripts with altered abundances) were identified at a log-fold threshold >=0.8. It can be inferred from **Figure 15** that the number of GTAAs increased with time post carolacton treatment. Except at 5 mins (where no gene was classified as a downregulated GTAA at the cutoff employed), the number of downregulated GTAAs was always greater than that of upregulated ones

4.2 Over-represented functional categories in the up- and downregulated GTAAs

Enrichment analysis with respect to gene functions according to the common functional classification and KEGG metabolic pathways was performed separately with the set of up-regulated GTAAs and the set of down-regulated GTAAs at every time point in order to gain an overview of the various functional categories of genes affected by carolacton. The results from the enrichment analysis of upregulated GTAAs indicated that transcripts from the pyrimidine ribonucleotide biosynthesis (P-value <= 4.09E-18), nucleosides, purines and pyrimidines' transport (P-value <= 2.45E-2), and metabolism of alanine, aspartate and glutamate (P-value <= 7.2E-4) were most significantly over-represented at both 5 min and 20 min but not at 40 min.

This chapter has been adapted from the publication P. Sudhakar, M. Reck, W. Wang, F. He, I. Wagner-Döbler, AP. Zeng. Construction and verification of the transcriptional regulatory network of the human dental pathogen *Streptococcus mutans* upon treatment with the biofilm inhibitor Carolacton. BMC Genomics 2014, 15:362.

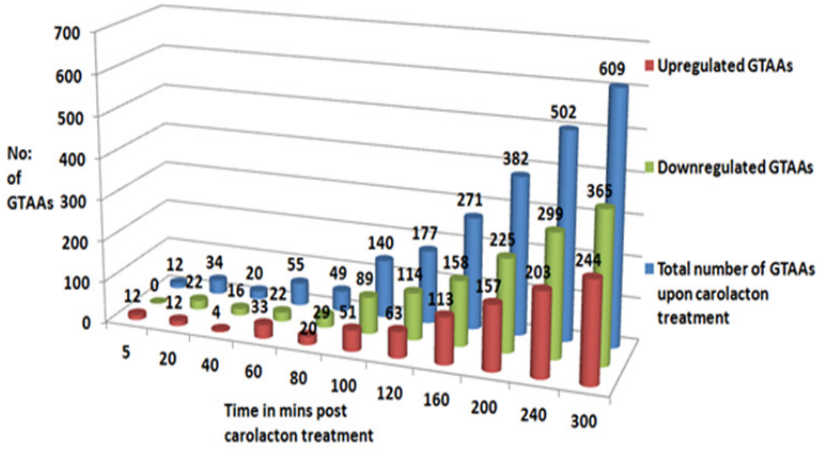


Figure 15. Temporal progression of the number of up- and downregulated GTAAs (Gene Transcripts with Altered Abundances) in *S. mutans* biofilms upon treatment with the biofilm inhibitor carolacton. Carolacton treatment resulted in the changes in the expression of 772 genes (absolute fold change >= 0.8 at any one of the sampling times) when compared to the non treated condition.

Especially, pyrimidine ribonucleotide biosynthesis and metabolism were among the most significantly over-represented functional classes within the set of upregulated GTAAs at 5 min post carolacton treatment, suggesting an essential role of pyrimidine biosynthesis and metabolism in the initial response of *S. mutans* to carolacton. At later time points, more gene functional categories were found to be over-represented among the set of upregulated GTAAs due to probable "spill-over" effects of secondary responses and/or downstream events.

Enrichment analysis of the downregulated GTAAs at 20 min revealed categories such as glutamate biosynthesis, TCA cycle, anaerobic energy metabolism, glyoxylate and dicarboxylate metabolism, nitrogen metabolism as well the signal transduction related two component systems (TCSs). TCS impart various virulence attributes to *S. mutans* by sensing and responding to various stress signals, enabling it to survive and tolerate unfavourable conditions [85, 324]. Among the TCS genes with altered transcript abundances is the *vicRK*

TCS which is responsible for many of the pathogenic and virulence attributes of *S. mutans* [19, 82, 84, 94, 96, 109, 325] as well as harboring the essential response regulator *vicR*. The set of downregulated GTAAs in the later phases starting from 20 min was also found to be enriched with more categories over the course of the experiment. In general, downregulated GTAAs in the early part of the response were enriched with categories corresponding to the processes of signaling, energy metabolism and amino acid metabolism. Interestingly, genes of the alanine, aspartate and glutamate metabolic pathways were found to be over-represented in both the up- and downregulated GTAA sets at the time point of 20 min after carolacton treatment, suggesting enzymes from these pathways are modulated differently in response to carolacton. The information from this time-series transcriptome data largely agree with those from a previous study by Reck et al [10] where only five sampling points were used; they show the reproducibility of the transcriptome analysis and also provide a fine temporal resolution which is important for deciphering the network level responses to carolacton.

4.3 Contextual co-expression network

In order to identify co-expression relationships among genes immediately affected by carolacton, the co-expression network was confined to statistically significant optimal correlations which started either from 0, 5 or 20 min after carolacton treatment. The contextual co-expression network inferred according to the workflow shown in **Figure 16** consisted of 8284 edges (gene-gene co-expression relationships). 5430 (65.5%) of the 8284 edges were characterized by time lagged co-expression relationships. 3959 (47.7%) of the co-expression relationships could be described as being inverted (opposite change trend in expression) whereas the remaining were described as being positive (same change trend in expression).

4.4 Regulator-target gene binding site map of *S. mutans* UA159

Of the more than 100 known and predicted transcriptional regulators in the *S. mutans* UA159 genome, only a few have been studied experimentally and an even smaller number characterized with respect to their binding site(s). The non-availability of experimentally verified transcriptional regulatory binding sites for the *S. mutans* transcriptional factors was compensated by applying comparative genomic approaches [216, 286, 287]. Binding site data for 44 transcriptional regulators were recovered from different sources such as experimental data from *S. mutans* itself as well as information from comparative genomic analysis (see **Table 2** for more information on the motifs corresponding to the 44 transcriptional regulators).

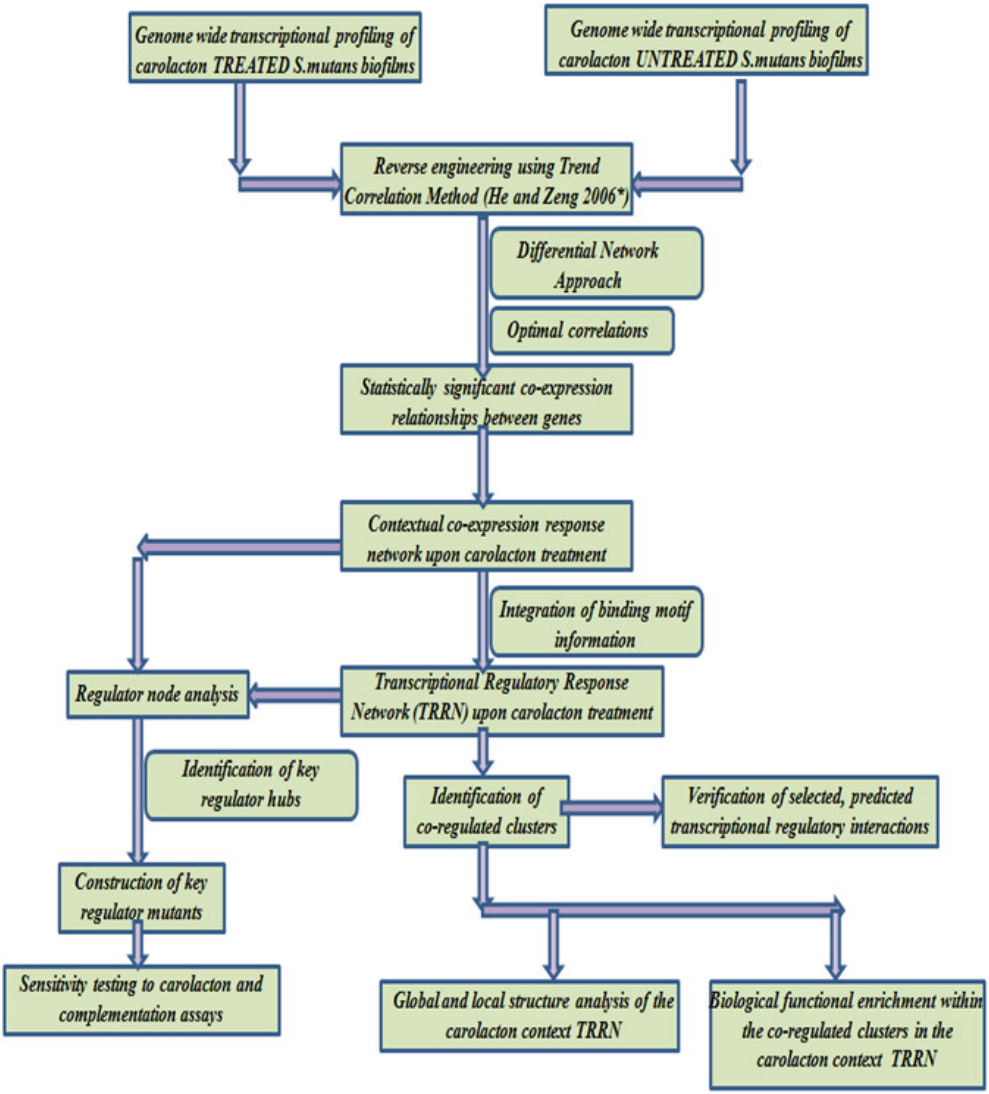


Figure 16. Workflow to capture the network level effects of the biofilm inhibitor carolacton on *S. mutans* biofilms. The directions of the arrow marks denote the flow of data processing and sequential steps. Shapes of boxes have no particular significance while the descriptions within the boxes represent the steps corresponding to data generation, algorithms, data processing, network and experimental analyses. * indicates the reference [22].

In cases where the collective number of binding sites inferred for a transcriptional regulator were fewer or equal to three, respective regulogic sites [231] from other species within the Streptococccus genus were retrieved. These were then used to construct PSSMs (Position Specific Scoring Matrices representing the binding motif consensus of transcription factors) with improved predictive capacities.

PSSMs were also constructed from regulogic sites available in the RegPrecise database [231]. Since the predictive capacity of each PSSM differs from case to case (due to its varying information content), a generic or common P-value could not be used to screen out "putative-target hits" (i.e genes or operons harboring binding sites predicted by the corresponding PSSMs). An online Regulatory Sequence Analysis Tool (RSAT) termed *matrix-quality* [316] was used to objectively determine the optimal P-value for every PSSM. The PSSMs were then used to predict binding sites (at the corresponding optimal P-values) in the upstream regulatory regions of all the genes in the *S. mutans* UA159 genome. A total of 1397 unique regulator-target gene relationships (corresponding to 2056 predicted binding sites) based on binding site data alone corresponding to 44 *S. mutans* UA159 transcriptional regulators were thus compiled. After integrating operon structure information, 2445 regulator-target gene relationships were found to comprise the final regulator-target gene binding site map. The global regulators CcpA (catabolite repressor protein) and CodY (global regulator of amino acid metabolism) together accounted for ~29.22% of the 2445 regulator-target gene relationships based on binding sites.

4.5 Transcriptional Regulatory Response Network (TRRN) of *S. mutans* under carolacton treatment

Combining inferred correlations in the co-expression network and the regulator-binding site map resulted in 227 transcriptional response regulatory relationships with predicted direct biological causality. After operon-oriented adjustment, the transcriptional response regulatory network (TRRN) (**Figure 17**) comprised 329 relationships among 307 genes, of which 37 were found to be transcription factors. Based on the predicted transcriptional regulatory connections, the TRRN genes could be organized into 27 co-regulated groups or sub-networks. A co-regulated group or sub-network is defined as a set of genes which are putatively regulated by a common transcription factor. 10 of the 37 regulators were associated with predicted incoming connections only and were contained within the 27 sub-networks with different size distributions**.** The largest sub-networks was that putatively modulated by CodY (which is a known global regulator [122]) and comprised 84 genes. The second largest sub-network consisted of 26 genes under the putative control of CysR (the cysteine metabolism regulatory protein encoded by SMU_852) followed by two sub-networks each containing 22 genes and predicted to be modulated by the essential response regulator VicR and the global regulator CcpA respectively.

4.6 Structural analysis of the TRRN

The TRRN (**Figure 17)** represents a "waterfall-like" model with respect to its global structure and hierarchy [326, 327] with some regulators also exerting control over other sub-networks by putatively modulating the expression of their corresponding transcription factors. For example, it can be seen from the inferred TRRN (**Figure 17)** that the global regulator CodY putatively regulates the expression of two distinct genes encoding the TCS response

regulators VicR and ComE, thus potentially exerting an indirect influence on their sub-networks as well. Other examples of this kind of hierarchical relationship were also observed in the TRRN. Although the concept of global hierarchical organization was observed in the regulatory response network studied herein, it was nevertheless confined only to a double-layered hierarchy as opposed to multi-layered hierarchies reported in well-studied organisms [327, 328]. This could primarily be due to the lack of biological connectivity data; firstly as a result of the limited experimentally verified binding site information of *S. mutans* transcription factors. And secondly, the functional or evolutionary divergence of *S. mutans* transcription factors from their corresponding orthologs (with experimentally verified sites) in well-known model organisms may have a limiting influence on the comparative genomic based extrapolation of binding sites [216]. Thirdly, the double-layered hierarchy could also be attributed to the absence of modulation of actual regulatory relationships in the studied biological context although the incompleteness of the network owing to lack of motif information is more plausible.

Overlap between sub-networks previously observed in bacteria [329, 330] is also a characteristic of the inferred TRRN. 40 of the 283 TRRN target genes with predicted incoming regulatory connections were found to be under the putative control of more than one transcriptional regulator (implying an incoming degree > 1). From a local network structure point of view, putative interactions involving genes belonging to just one sub-network resemble single input motifs (SIM) whereas genes belonging to multiple sub-networks form multiple input motifs (MIMs) [331]. Besides MIMs, overlap between sub-networks could be caused by more complicated motifs such as the feed forward loop (FFL) in which a regulator encoding gene is modulated by another regulator, both of which control the expression of a common gene [332]. If multiple genes are controlled by a single FFL, then it is termed as a multi-output FFL (MOFFL) [333].

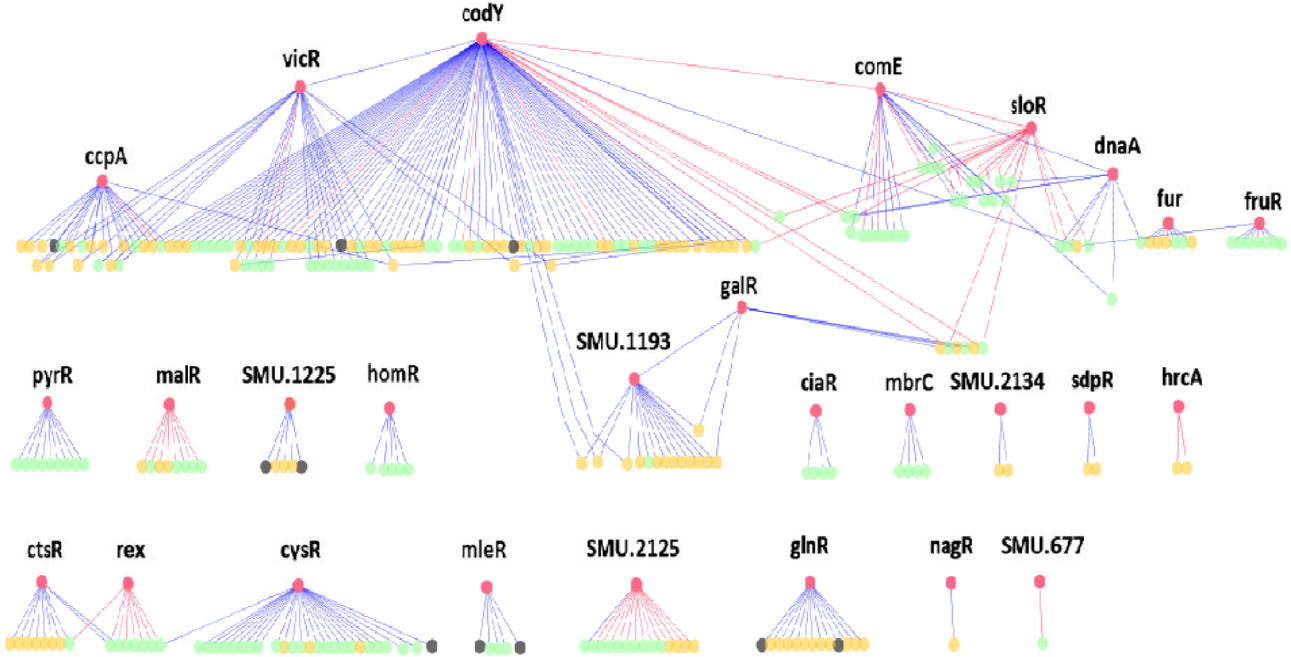


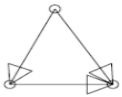
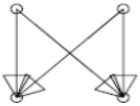
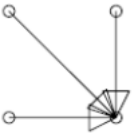
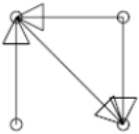
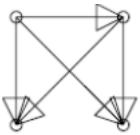
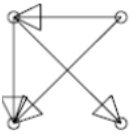
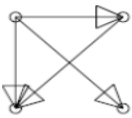
Figure 17. Topological view of the transcriptional regulatory response network (TRRN) of *S. mutans* biofilms upon carolacton treatment. The TRRN was inferred by overlaying the regulator-target gene binding site map onto the co-expression network. It consisted of 27 co-regulated gene groups or subnetworks each under the control of a transcription factor and comprised 329 regulatory interactions involving 307 genes. Some of the sub-networks overlap with each other as a result of genes modulated by more than one transcription factor. The 27 regulators with outgoing connections (marked with red nodes) along with 10 other regulators only with incoming connections (marked with black nodes) found to be among target genes within the sub-networks are indicated as well as non-regulator target genes (marked with green nodes). If the upstream regulatory regions of the target genes or their corresponding operons harbored multiple putative binding sites of its predicted regulator(s), then they are indicated by (marked with orange nodes). In addition, the regulators with outgoing connections are also marked with their gene names above their respective nodes. Blue arrows indicate a positive (activation) relationship whereas red arrows stand for a negative (repression) effect. Connections always flow from top to bottom. Spatial positions of transcription factor nodes are manipulated so as to pictorially depict the possible hierarchies. The carolacton context TRRN shown herein is observed to be organized as a double layered hierarchy. The network was visualized in Cytoscape.

Table 6 lists the various motif types identified (P-value <= 0.01 after 10000 randomizations) in the carolacton context dependent TRRN among which were found 4 FFLs including other motifs such as regulatory bi-fans [331] and their derivatives [334]. Local network topologies such as the motifs found here are consistent with previous observations in other organisms [331-333, 335] and could be important for eliciting quick and flexible responses to stimuli [332]. Although the exact biological mechanisms which give rise to such motifs remain to be understood, these observations can be used for detailed quantitative modeling of smaller sub-systems involved in the response of *S. mutans* biofilms to carolacton.

4.7 Biological context within the TRRN sub-networks

Put together, 22 enrichment events involving 6 KEGG metabolic pathways, 7 biological functional classes and 6 gene ontology terms were collectively identified within the 27 sub-networks. At least 7 cases of functional enrichment events were observed in which the transcription factor or regulator belonged to the same functional category as that found to be over-represented in its corresponding sub-network, confirming that the constructed regulatory network is biologically meaningful. For instance, the carolacton context TRRN genes co-regulated by the glutamine synthetase repressor GlnR known to control glutamine metabolism [76] in *S. mutans* were enriched with genes belonging to the functional categories “Amino acid biosynthesis” and “Nitrogen metabolism”.

Table 6. The different types of local network motifs identified in the *S. mutans* TRRN upon carolacton treatment. The motifs were identified and their corresponding statistical scores determined using the MFINDER tool [320, 334]. °after 10000 randomizations. *SD denotes the standard deviation of the number of occurrences of the motif in randomized networks. ^The "Z score" is a measure of the statistical significance of the motifs and is determined as (M_{real} - M_{rand}) / SD. Motifs with P-values <= 0.01 were considered as significant.

Network Motif	Common name	Number of occurrences in the real network (M_{real})	Number of occurrences in the randomized° network (M_{rand}) +/- SD*	Z score of motif occurrence^
	Feed forward loop (FFL)	11	2.5 +- 1.8	4.72
	Bi-fan motif	55	9.5 +- 4.3	10.58
	Multiple input module (MIM)	3	2.6 +- 0.6	0.67
	--	12	2.8 +- 2.6	3.54
	Possible bi-fan derivative	17	1.3 +- 2.1	7.48
	--	162	39.6 +- 27	4.53
	--	401	153 +- 133	1.86

On closer observation, most of the genes within the GlnR sub-network were related to glutamine metabolism. Genes belonging to the biological functional class of cell envelope metabolism were over-represented in the sub-network regulated by the response regulator MbrC associated with cell envelope stress response [95]. Similarly, the genes putatively modulated by the pyrimidine regulatory protein PyrR contained an over-representation of pyrimidine metabolism genes. Several instances were also observed in which the functional category to which the regulator belonged to was different from that found to be over represented within the sub-network (**Figure 18, Table 7**). Besides the above described intra- and inter categorical relationships between regulators and target genes, the TRRN also contained already verified regulatory relationships.

Table 7. List of enriched categories in the co-regulated groups (sub-networks) of the *S. mutans*TRRN upon carolacton treatment. Over-represented categories include KEGG metabolic pathways (KEGG MP), Biological functional classes (BFC) and gene ontology terms (GO). The details pertaining to the calculation of the significance scores were provided in the material and methods chapter.

Enriched category term	Category type	Co-regulated group(CG)	Number of genes in the category [A]	Number of genes in the co-regulated groups [B]	Intersection (A ΙΙ B)	P-value of Intersection	E- value of Intersection	Significance score
Oxidative phosphorylation	KEGG MP	ctsR CG	10	11	8	1.70E-15	3.84E-12	11.42
Ribosome	KEGG MP	SMU.2125 CG	49	14	9	8.40E-09	1.90E-05	4.72
Pyrimidine metabolism	KEGG MP	pyrR CG	40	9	6	7.60E-07	0.00172	2.76
Valine leucine and isoleucine biosynthesis	KEGG MP	codY CG	19	84	9	3.60E-05	0.08243	1.08
Galactose metabolism	KEGG MP	galR CG	18	7	3	0.00031	0.69901	0.16
Nitrogen metabolism	KEGG MP	glnR CG	10	13	3	0.00037	0.83221	0.08
Purines pyrimidines nucleosides and nucleotides	BFC	pyrR CG	61	9	8	5.20E-12	2.60E-09	8.58
Protein synthesis	BFC	SMU.2125 CG	123	14	9	1.90E-08	9.60E-06	5.02
Amino acid biosynthesis	BFC	codY CG	86	84	17	4.60E-08	2.40E-05	4.63
Energy metabolism	BFC	ctsR CG	149	11	8	1.30E-07	6.90E-05	4.16

Cell envelope	BFC	mbrC CG	81	4	3	0.00027	0.13868	0.86
Transport and binding proteins	BFC	SMU.1193 CG	210	14	7	0.00027	0.14075	0.85
Signal transduction	BFC	ccpA CG	61	22	5	0.00045	0.23044	0.64
Transport and binding proteins	BFC	homR CG	210	5	4	0.00061	0.3109	0.51
Transport and binding proteins	BFC	ccpA CG	210	22	8	0.00135	0.6938	0.16
Amino acid biosynthesis	BFC	glnR CG	86	13	4	0.00188	0.96285	0.02
Hydrogen ion transporting ATPase activity, rotational mechanism	GO	ctsR CG	8	11	8	3.50E-17	4.50E-13	12.35
Hydrogen ion transporting ATP synthase activity, rotational mechanism	GO	ctsR CG	8	11	8	3.50E-17	4.50E-13	12.35
Ribosome	GO	SMU.2125 CG	48	14	9	6.40E-09	8.10E-05	4.09
Structural constituent of ribosome	GO	SMU.2125 CG	49	14	9	7.80E-09	9.90E-05	4
Translation	GO	SMU.2125 CG	55	14	9	2.30E-08	0.0003	3.53
Intracellular	GO	SMU.2125 CG	69	14	8	3.50E-06	0.0449	1.35

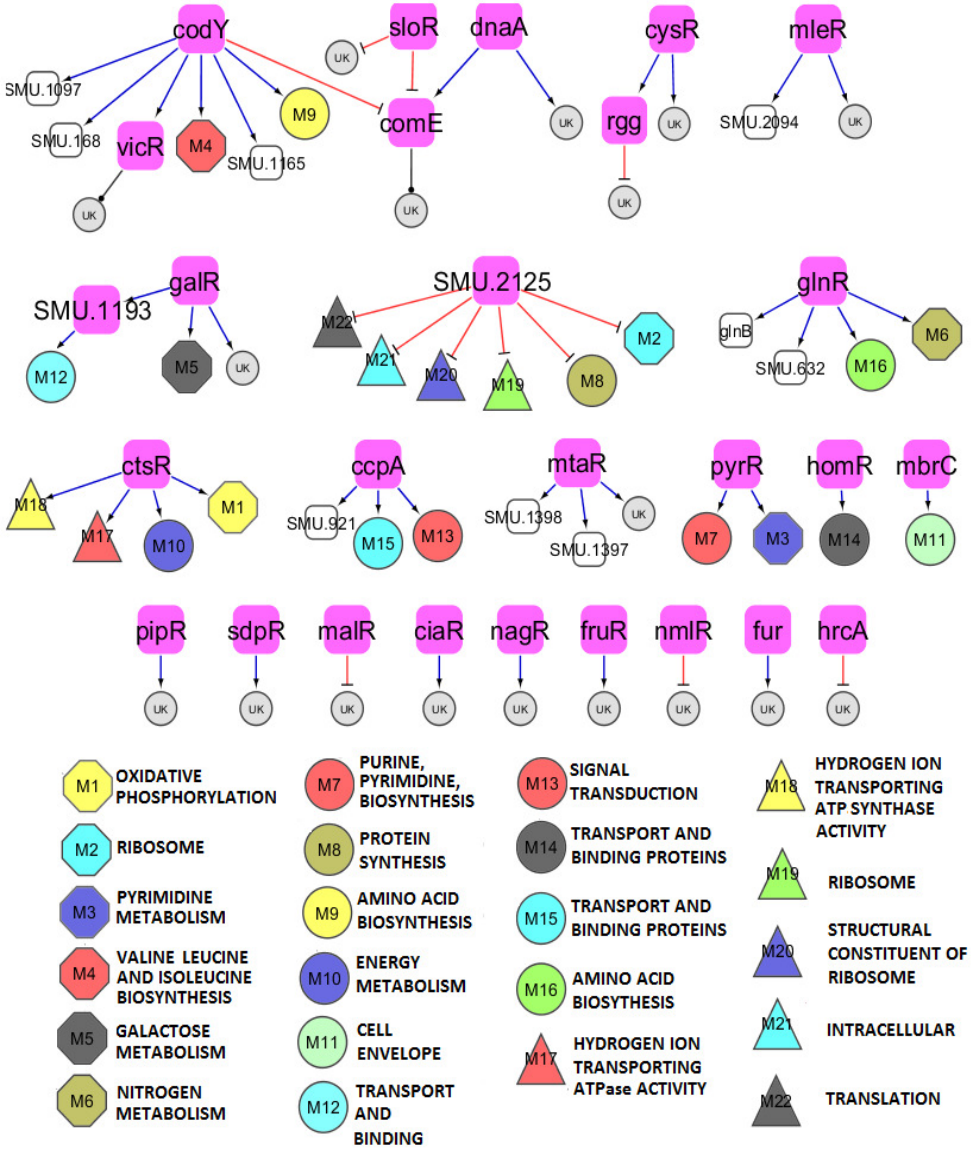


Figure 18. Categorical enrichment within the sub-networks comprising the *S. mutans* biofilms TRRN upon carolacton treatment. The co-regulated groups or sub-networks within the TRRN were found to be enriched with functional categories such as KEGG metabolic pathways, biological functional classes and gene ontology terms thus lending biological meaning to the inferred regulatory response network. Co-regulated groups denote sets of genes predicted to be commonly modulated by a transcription factor

(indicated by). The regulatory response network also includes other transcriptional factors with only incoming connections ( ) and found to be among the co-regulated groups. Some of the sub-networks were found to be enriched with functional categories such as KEGG metabolic pathways ( ), biological functional classes ( ) and gene ontology terms ( ) whereas others did not display any enrichment ( ). Black edges with ellipsoid target edges ( ) represent relationships with dual regulation between the regulator and the co-regulated genes. Dual regulation indicates that within sub-networks (consisting of genes commonly predicted to be modulated by a particular transcription factor), there are transcriptional regulatory relationships some of which are characterized by positive (activational) expression patterns and others by inverted (repressive) expression patterns.The meanings of all the other arrow symbols are as described in the legend of Figure 17.

4.8 TRRN sub-networks co-regulated by global transcription factors

The sub-networks of two global transcriptional regulators namely CodY and CcpA [71, 122] were among the top three sub-networks in terms of the number of genes regulated. CcpA is reported to modulate sugar uptake and metabolism, carbon catabolite repression and expression of virulence-related genes in *S. mutans* [71]. The treatment of *S. mutans* biofilms with carolacton caused an immediate downregulation of CcpA modulated genes (**Figure 19**). The CcpA sub-network consists of genes encoding components of two mannose-specific phosphotransferase systems (*ptnAC (manLM)*-SMU_1879 and SMU_1960c-SMU_1961c) [72, 119, 336], as well as three ATP dependent transporter complexes (*msmEFG*, SMU_921-SMU_922 and SMU_1315c-SMU_1317c) among others. While *msmEFG* is reportedly involved in the energy-dependent transport of multiple sugars such as stachyose and raffinose [337], the substrate specificities of the other two gene clusters are unknown.

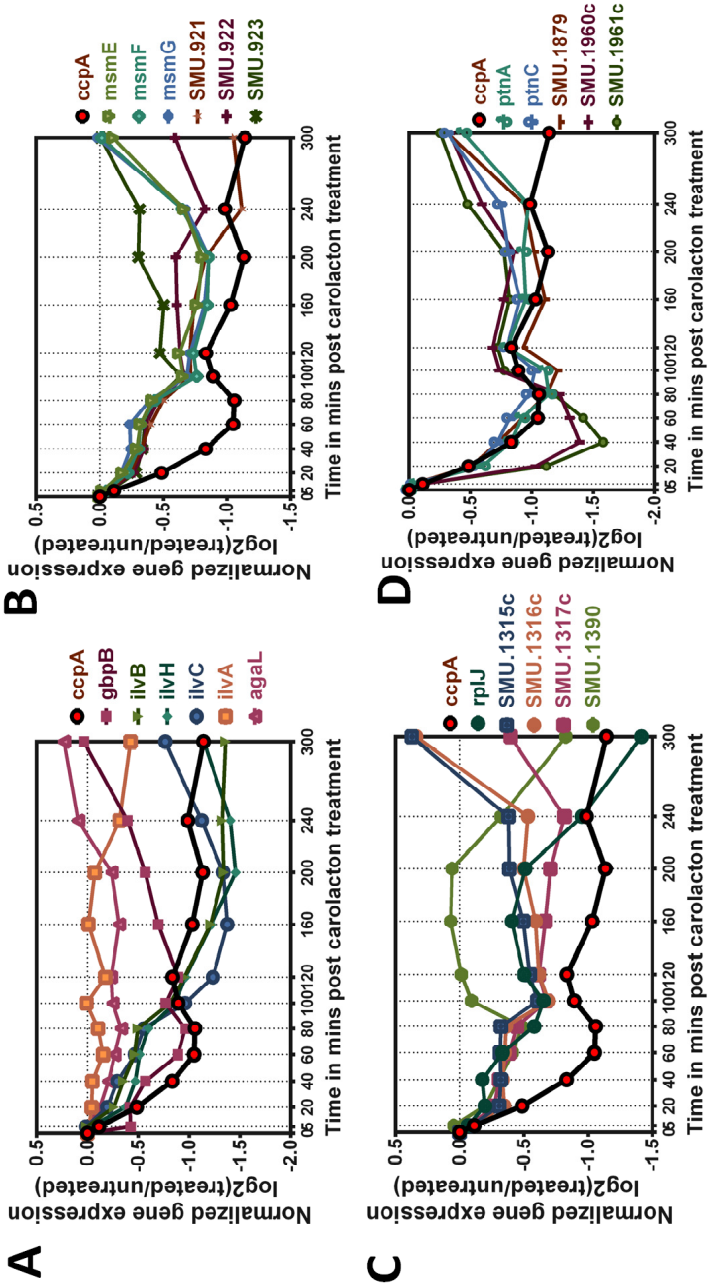


Figure 19 (A-D). Normalized expression profiles of genes co-regulated by the global transcription factor CcpA.

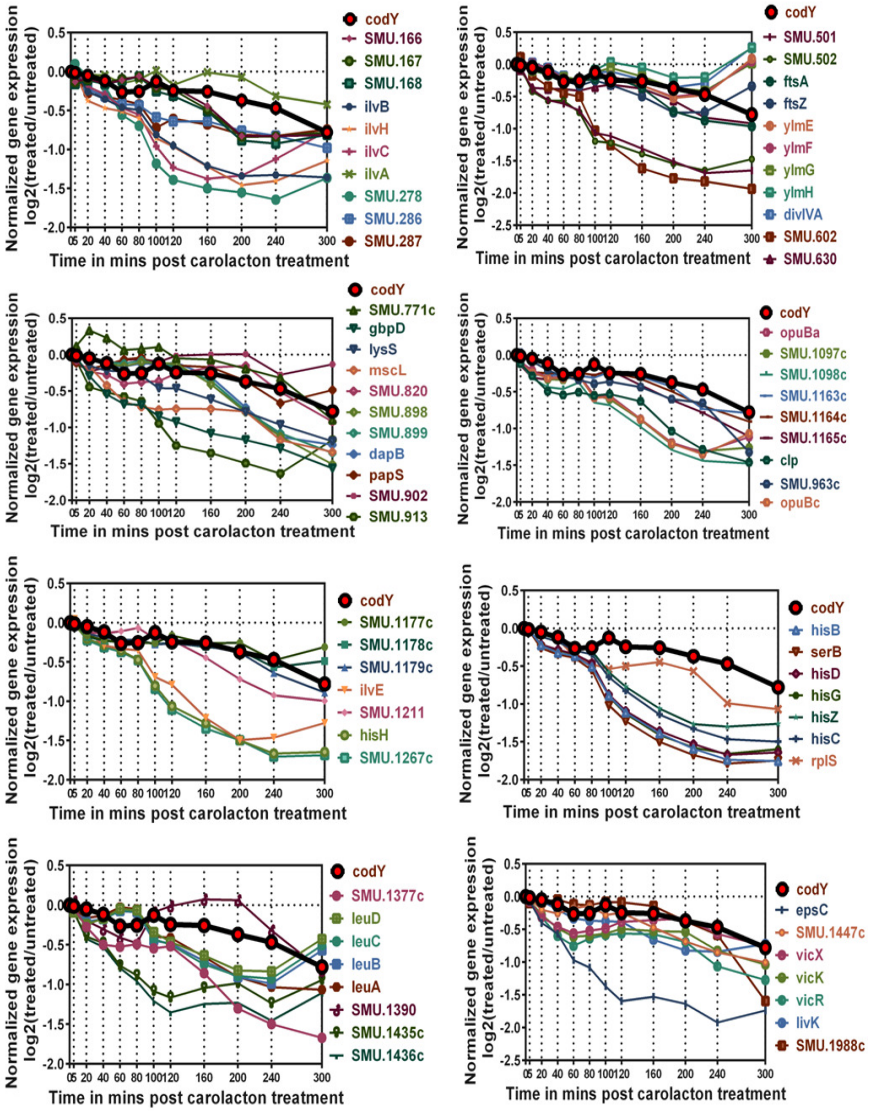


Figure 20. Normalized expression profiles of genes co-regulated by the global transcription factor CodY. Only genes which showed expression patterns that are non-inverted with that of *codY* expression are shown here.

Genes in the sub-networks of other sugar-specific transcription factors such as the repressors for galactose (GalR) and fructose (FruR) were also found to be downregulated. Sugars serve among others as substrates for the synthesis of glucans and associated by-products required for the adherence of cells to the tooth surface and the formation of biofilms, as well as for the formation of organic acids which aid cariogenic disease progression [28, 338-341].

The gene cluster co-regulated by CodY (**Figure 20**) is enriched with genes belonging to the biosynthesis and metabolism of amino acids, particularly the branched chain amino acids such as leucine, valine and isoleucine. The control of the metabolism of these amino acids by CodY has already been reported [122]. Hence, from the analysis of the regulatory response network, it can be stated that carolacton has an inhibitory effect on the central metabolism of sugars and amino acids by downregulating sub-networks modulated by the global transcriptional factors CcpA and CodY.

4.9 Immediate induction of the pyrimidine metabolism

Peptidoglycan is an important constitutent of the gram positive cell-wall. It is expected that cell membrane damage and biofilm inhibition would have a substantial effect on pathways and genes related to its synthesis and metabolism. UDP-N-acetylglucosamine, a key intermediate in the biosynthetic process of the cell wall component peptidoglycan, is produced by glycolysis, sugar metabolism as well as the pyrimidine metabolic pathway [342]. Expression data indicate the absence of immediate modulation and the downregulation at later time points of the glycolytic pathway as well as of the pathways related to the metabolism of various sugars such as fructose, mannose and galactose.

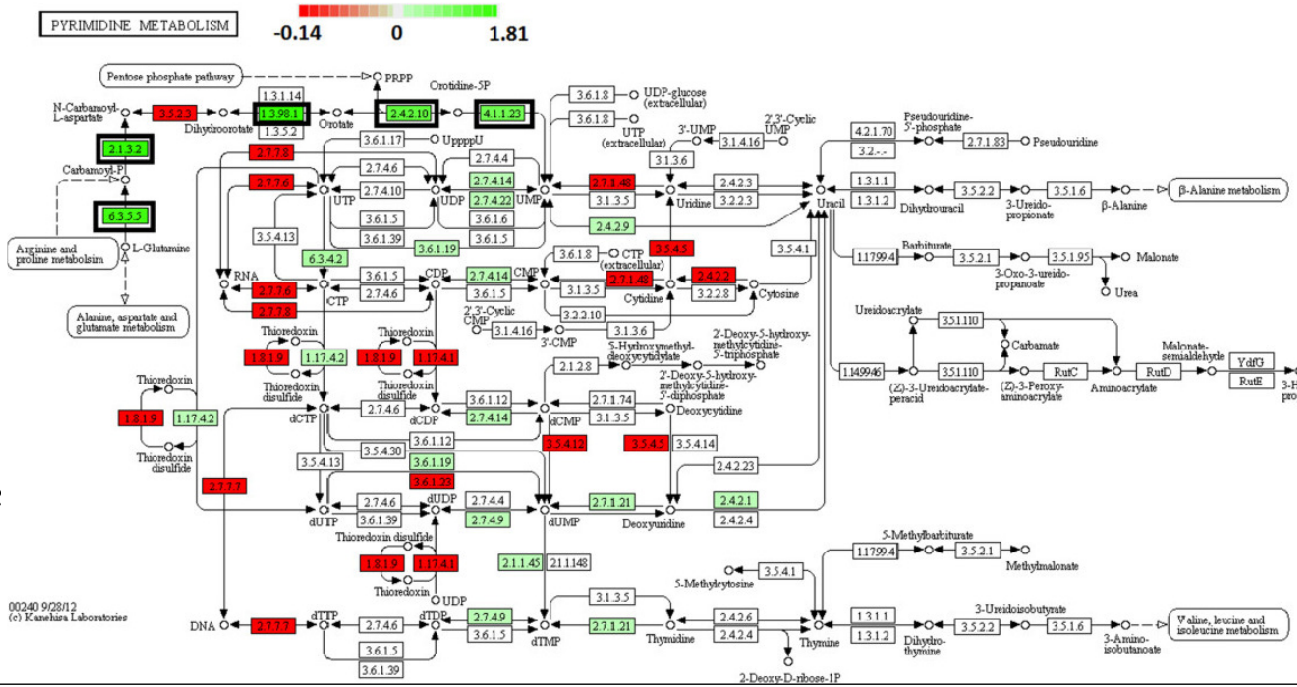


Figure 21. Heat map representation of the transcriptional response of the *S. mutans* pyrimidine metabolic pathway upon carolacton treatment. Genes from the pyrimidine metabolism pathway were among the first to be modulated upon carolacton treatment. The log2-fold expression change of pathway genes at 5 min post treatment were used for the heat-map representation. Green indicates upregulation and red downregulation. The scale is indicative of the corresponding changes in normalized gene expression. Pathway genes encoding enzymes catalyzing reactions leading up to UMP were strongly upregulated while most of the other pathway genes exhibited relatively weak modulation. Enzymes marked in black boxes indicate the corresponding strongly upregulated transcripts of the pathway. White cells correspond to pathway enzymes not found in the genome of *S. mutans* UA159. If a particular enzyme corresponds to multiple transcripts (as a result of multiple protein subunits constituting an enzyme), then the transcript with the highest amplitude of log2-fold change was used. Graph generated using the Mayday visualization tool version 2.12.

On the contrary, two operons encoding pyrimidine biosynthesis genes (namely the *pyrEFDZ* and *pyrRPBA-carB* operons) belonging to the PyrR sub-network and coding for the enzymes of the pyrimidine metabolism pathway (**Figure 21**) were upregulated by about 1 to 1.8 fold at 5 min post treatment (**Figure 22**). It is of note that most of the genes in the pyrimidine metabolism pathway were not transcriptionally altered with the exception of the two strongly upregulated *pyrEFDZ* and *pyrRPBA-carB* operons. These operons encode enzymes catalyzing the biochemical steps leading to the production of UMP and UDP (see **Figure 21**) suggesting that this part of the pathway is specifically activated. An upregulation of the pyrimidine metabolism pathway would produce pools of UDP-N-acetlyglucosamine (UDP-N-AG) for peptidoglycan synthesis in response to carolacton-induced membrane and cell wall damage. The upregulation of pyrimidine biosynthetic steps was also observed in an *S.aureus* strain harboring a mutation of a two component system essential for cell wall metabolism [343].

4.10 Carolacton affects glutamine metabolism

Genes in the TRRN predicted to be co-regulated by the glutamine synthetase repressor GlnR along with *glnR* itself were associated with a sharp downregulation initially (**Figure 23**). It is known that glutamine is an important source of nitrogen for *S. mutans* [77]. Some of the GlnR-modulated genes in response to carolacton treatment included a glutamine transporter gene-cluster (SMU_1519-SMU_1522), the glutamine/glutamate biosynthesis operon (*glnRA-gltAD*), as well as an operon consisting of the putative ammonium transporter coding gene *nrgA* and the nitrogen regulatory protein coding gene *glnB*.

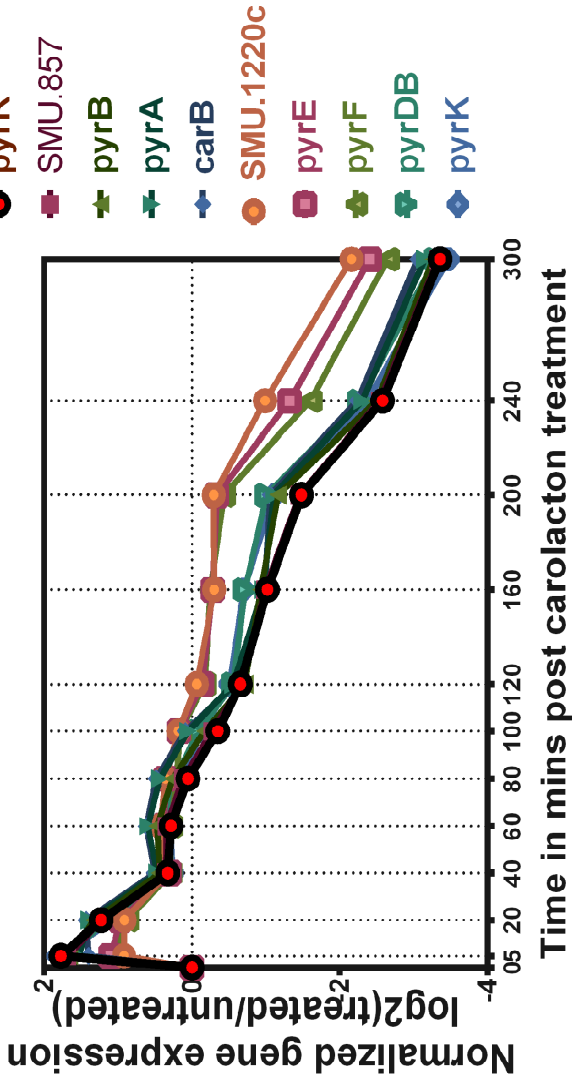


Figure 22. Normalized expression profiles of the genes co-regulated by the pyrimidine biosynthesis regulatory protein in the carolacton treatment context TRRN of *S. mutans*. Two operons containing genes encoding enzymes involved in the biosynthesis of pyrimidine ribonucleotides were upregulated sharply by about 1.8 to 2 log2-fold 5 min post carolacton treatment.

Other glutamate and glutamine metabolism related genes such as *citZH* and *idh* [344], whose modulation was also reported to be mediated by GlnR [76], were downregulated initially until 20 min post carolacton addition. The protein products of the *citZH* and *idh* genes catalyze the steps leading to the formation of alpha-ketoglutarate from oxaloacetate (feeding the de-novo glutamine synthesis pathway) [344]. The expression pattern of the genes predicted to be controlled by GlnR suggests an activational (positive) rather than an inverted (negative) relationship. But GlnR is known to function as a repressor protein. However, this does not seem to be a conflictory observation since GlnR orthologs have been known to have dual activator-repressor functions [345]. Hence, the *S. mutans* GlnR could also be a dual regulator given that several transcriptional regulators with dual functions have been reported earlier in *S. mutans* [97, 346, 378]. This could explain the activational or positive modulation of GlnR-controlled genes. Glutamate is known to play a role in the acid tolerance response (ATR) mechanism of *S. mutans* [347-349] which is crucial for the survival and adaptation of *S. mutans*. Carolacton has the maximum effect in terms of membrane damage and cell death in actively growing biofilm cultures at low pH. But the transcriptional changes (immediate downregulation until 20 min post treatment) of *glnR* and its target genes occurred at a neutral pH (data not shown) thus ruling out ATR as a cause for the modulation of GlnR mediated genes. At later time points which correspond to a low pH, indeed the upregulation of the GlnR-modulated genes occurred although a downregulation should have been the expected trend to enhance the ATR. This suggests that carolacton either abolishes the ATR mediated via glutamate or weakens it due to the disturbed glutamine metabolism. Alternatively, the upregulation of the GlnR sub-network after the initial downregulation could possibly be explained as a means to provide precursors for peptidoglycan biosynthesis. Thus, it is plausible that the pyrimidine metabolism is induced to feed the peptidoglycan biosynthesis during the initial phase (until 20 min after carolacton treatment) followed by the upregulation of the glutamine metabolism related genes during the later phase.

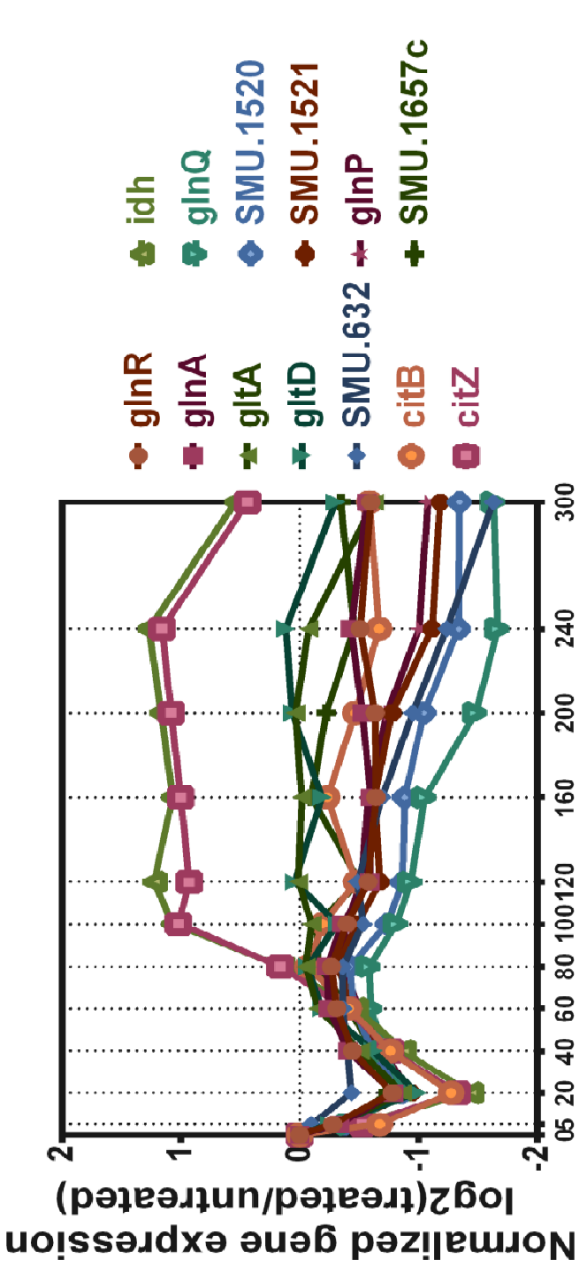


Figure 23. Expression dynamics of the genes co-regulated by the glutamine repressor GlnR in response to carolacton treatment.

4.11 TRRN genes co-regulated by the essential response regulator VicR

The microarray data show that *vicKR*, one of the earliest responding TCSs upon carolacton treatment, was downregulated. The *vicKR* TCS plays a major role in the biofilm formation, competence development, oxidative stress tolerance, acid tolerance, autolysin production, and glucan and fructan metabolism of *S. mutans* [19, 82, 84, 94, 96, 109, 325]. In addition, the response regulator *vicR* has also been identified as being essential for the survival of *S. mutans,* i.e. it is an essential gene which cannot be deleted [81]. The genes co-regulated by VicR (**Figure 24**) in the carolacton treatment context TRRN consisted mostly of downregulated genes including those encoding surface structure and cell wall proteins such as glucosyltransferases (gtf) B and D, glucan binding protein B (gbpB), and the sortase-anchored cell wall protein *wapE* [350] among others.

This is consistent with the previous study [10] where all the genes co-expressed with the downregulated *vicR* had reduced transcript abundances. The evidence put together suggests that carolacton could interfere with the signalling mechanisms related to the activity of this essential response regulator thereby exerting its biofilm inhibitory and membrane damaging effects. Some of the genes such as *gtfB*, *gbpB* and *comC* in the VicR sub-network were already known transcriptional regulatory targets [96, 108, 351] of the VicR protein thus contributing to an independent verification of our target prediction based on microarray data and motif information.

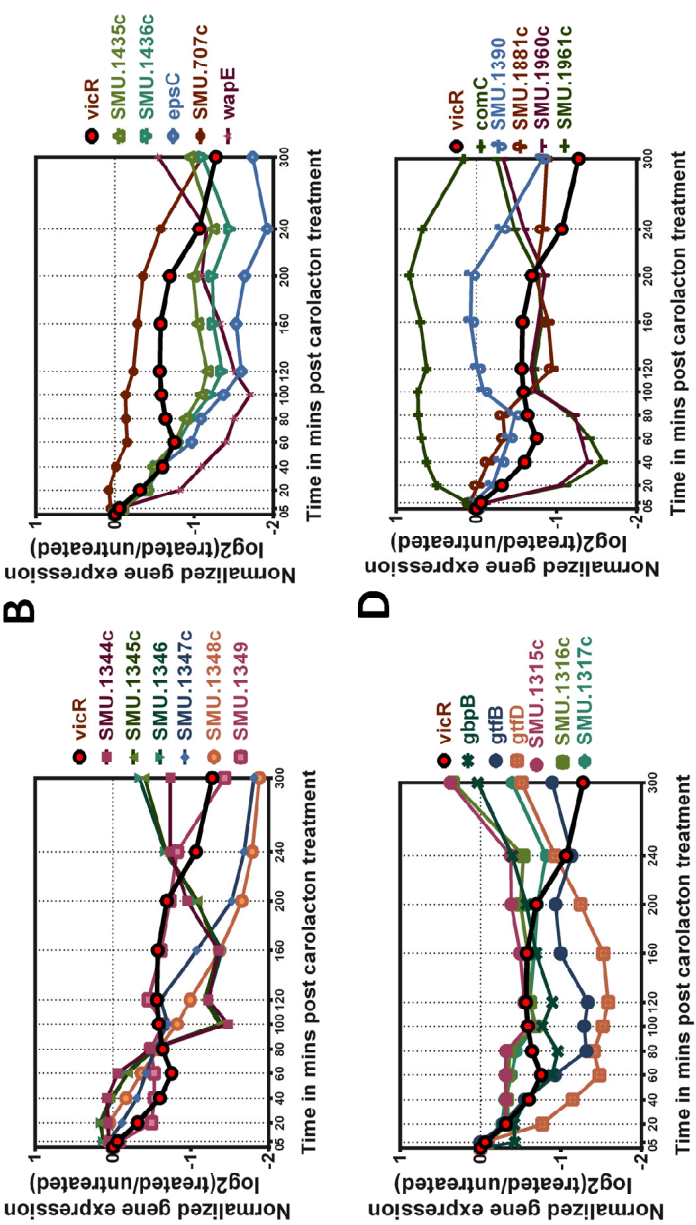


Figure 24 (A-D). Normalized expression profiles of the genes commonly regulated by the downregulated essential TCS response regulator *vicR* in response to carolacton. VicR modulated genes include those encoding virulence attributing products such as glucosyltransferases B and D, cell wall protein WapE among others.

4.12 Modulation of genes co-regulated by the cysteine metabolism controlling transcription factor CysR

Genes connected by edges to *cysR* (encoding a regulatory protein controlling the metabolism of sulfur containing amino acids [352]) in the co-expression network had both inverted and positive activational patterns compared to that of *cysR*. After combining the CysR binding motif information [352] however, directionality was assigned to the edges after which all of the co-regulated genes were having only positive activational expression patterns (**Figure 25**) in relation to *cysR*. CysR was found to modulate 26 TRRN genes (**Figure 26**) making it the transcriptional regulator with the second highest degree (number of modulated genes) in the carolacton context TRRN.

Some of the genes transcriptionally modulated by CysR include SMU_609 encoding a putative-40K cell wall protein precursor, SMU_246 encoding a putative glycosyltransferase-N-acetylglucosaminyltransferase and SMU_984 whose gene product has been predicted to be an uncharacterized autolysin among others. Besides, SMU_1509 whose gene product is a Rgg family [353] regulator protein was also observed to be among the CysR sub-network as were genes encoding putative Nif proteins related to nitrogen metabolism and those such as the SMU_1074-SMU_1077 operon involved in cysteine metabolism. Statistically significant (significance score >=0) enrichment events with respect to functional categories were not observed within the CysR sub-network. Nevertheless, the functional relevance of the individual genes (such as SMU_609 which is one of the strongest modulated genes with > 2 log-fold upregulation in response to carolacton treatment) within the CysR sub-network as well as its connectivity in the TRRN point out the importance of CysR in the response of S. mutans biofilms to carolacton.

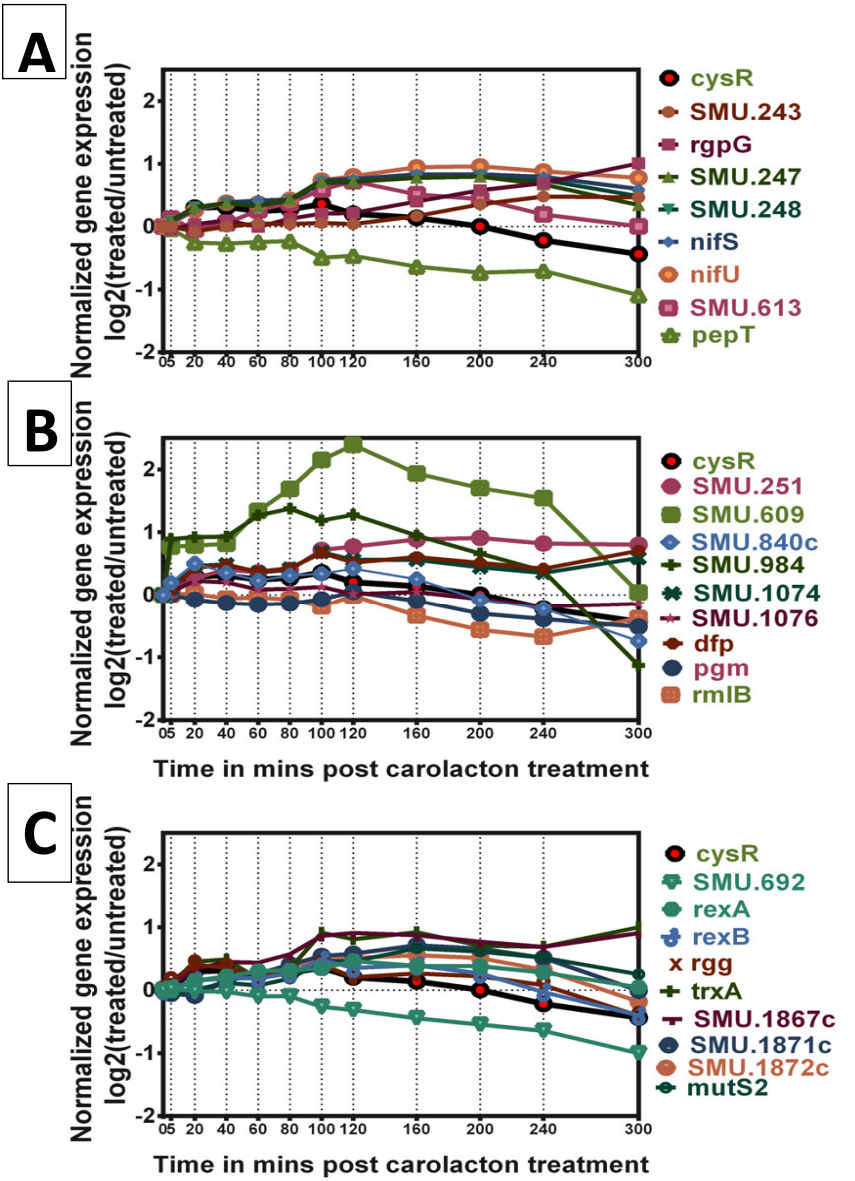


Figure 25 (A-C). Temporal behavior of the 26 genes found within the co-regulated group/subnetwork commonly modulated by SMU.852 encoding the CysR cysteine metabolism regulatory protein.

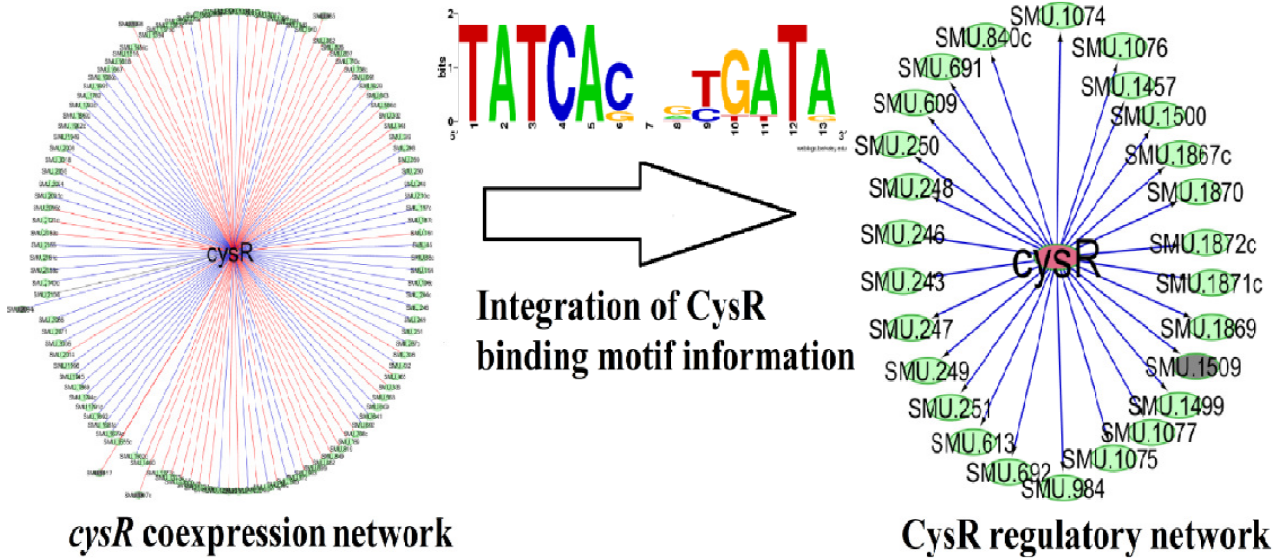


Figure 26. Co-regulated gene groups or sub-networks were constructed by overlaying predicted regulator-binding site maps onto the co-expression network as shown here specifically for the network confined to cysR. The node corresponding to the lone regulator (SMU.1509 encoding a putative Rgg family transcription factor) in the gene group co-regulated by CysR is marked in grey.

Chapter 5. Results: Experimental verification of network predictions

The experiments (EMSA, generation of mutants and their sensitivity testing) were performed by Dr. Reck, Research Group Microbial Communication, Helmholtz-Center for Infection Research, Braunschweig, Germany.

5.1 Regulatory interaction between MbrC (BceR) and the *murMN*-SMU_718 operon

By combining gene expression and motif information, *mbrC* was predicted to regulate four so far unidentified target genes (the SMU_716-SMU_717-SMU_718c operon and SMU_610 encoding the cell surface antigen SpaP) in the carolacton treatment context TRRN. The autoregulatory response regulator MbrC - also known as BceR - is encoded within a four-gene operon (*mbrABCD*/*bceABRS*) and has been shown to regulate cell envelope stress response mechanisms in *S. mutans* [95]. Ouyang et al demonstrated using binding studies that MbrC transcriptionally regulates the genes SMU_302, SMU_862, SMU_1006 and SMU_1856 by binding to their promoter elements via a conserved binding motif. The binding motif consensus consists of a pair inverted repeats separated by 2 nucleotides (TTACAAnnTTGTAA) [95].

This chapter has been adapted from the publication P. Sudhakar, M. Reck, W. Wang, F. He, I. Wagner-Döbler, AP. Zeng. Construction and verification of the transcriptional regulatory network of the human dental pathogen *Streptococcus mutans* upon treatment with the biofilm inhibitor Carolacton. BMC Genomics 2014, 15:362.

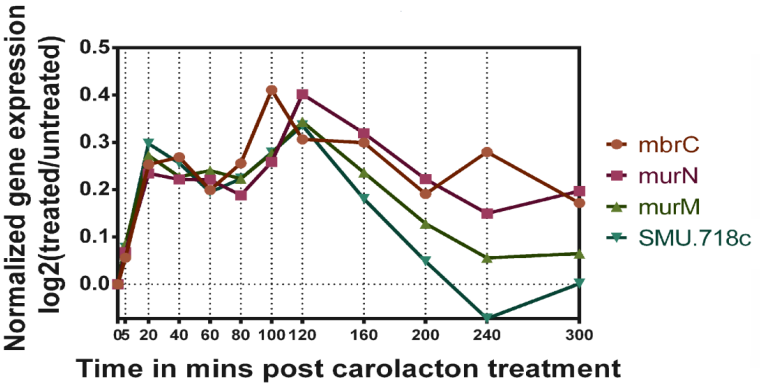


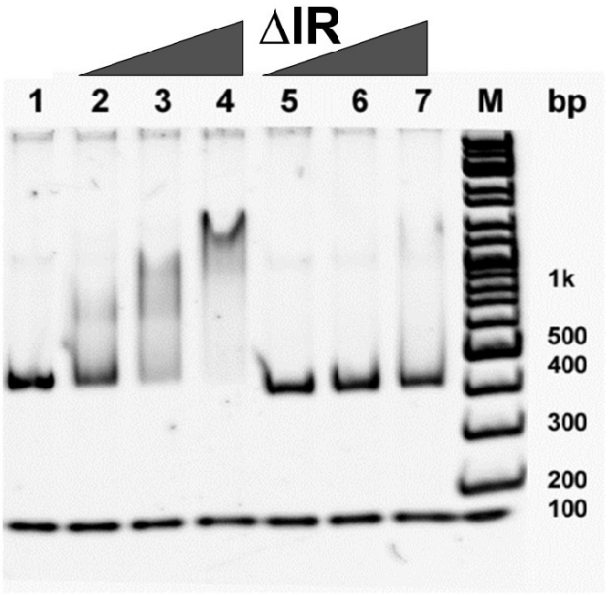
Figure 27. Expression profiles of *mbrC* and the *murMN*-SMU_718 operon genes in response to carolacton treatment.

Among the predicted MbrC target genes within the carolacton context TRRN, SMU_718c codes for a hypothetical protein with a haloacid dehalogenase-like domain, and SMU_716 and SMU_717 encode two different enzymes, MurN and MurM respectively. These enzymes catalyze the last steps of the peptidoglycan biosynthesis pathway and also play a important role in imparting resistance to cell wall-acting antibiotics [354-357]. **Figure 27** illustrates the coexpression of *mbrC* with the genes of the aforementioned operon and **Figure 28** displays the potential MbrC binding site (TTACAA-AT-TTCTAC) upstream of the putative target *murMN*-SMU_718c operon. This potential binding site differs from the motif consensus identified by Ouyang et al. [95] by the presence of two substitutions in the inverted repeat and is located upstream (between -33 and -20) of the transcriptional start site of the *murMN*-SMU_718c operon.

GENE	UPSTREAM REGULATORY SEQUENCE WITH THE SPACED DYAD MbrC BINDING MOTIF	POSITION	STRAND
SMU_1006	TCAATGC*TTACAATTTTGTAA*GCTACGATTCTTTAAGTG	-59	D
SMU_302	TGAGGAC*TTACAAAGGTGTAA*GGTTCTATCTTGATTTTT	-53	D
SMU_1856	GGTTTAT*TTACAAAATTGTAA*GTTATAGTACAAAGACTG	-40	D
SMU_862	TTAAAAC*TTACAATTTTGTAA*GTTTTAACTTTAATAATG	-19	D
SMU_718c	TTTCTAA*TTACAAATTTCTAC*AGACTATCTTACTAAAAA	-33	R

Figure 28. The alignment between the MbrC binding sites in other experimentally verified targets (black) reported by Ouyang et al [95] and the putative site upstream of the predicted target (red) *murMN*-SMU_718 operon is shown. The signature repeats of the MbrC binding motif are italicized, underlined and shown in bold.

To verify this predicted regulatory relationship, Electro Mobility Shift Assays (EMSA) were carried out. The promoter region of SMU_1006 served as a positive control as the binding of MbrC to it was reported previously [95]. An unrelated DNA fragment (around 100-150 bp, lower DNA band in all the lanes) was added in all the samples to prove the specificity of the binding reaction. As shown in **Figure 29,** the results confirmed the binding of MbrC to the SMU_1006 promoter as well as the necessity of the consensus motif for the binding of MbrC to its target. When an unrelated transcriptional regulator VicR was used instead of MbrC in the binding reaction as a control to test the specificity of the assay, no shift of the DNA band was found while with increasing MbrC concentrations, a clear shift was observed (upper band around 350 bp). No significant shift was observed upon deleting the binding motif from the SMU_1006 promoter.



Lane M : Ladder

Lane 1 (control) : 20 pmol VicR protein added to the intact SMU_1006 promoter

Lane 2 : 5 pmol MbrC protein added to intact SMU_1006 promoter

Lane 3 : 10 pmol MbrC protein added to intact SMU_1006 promoter

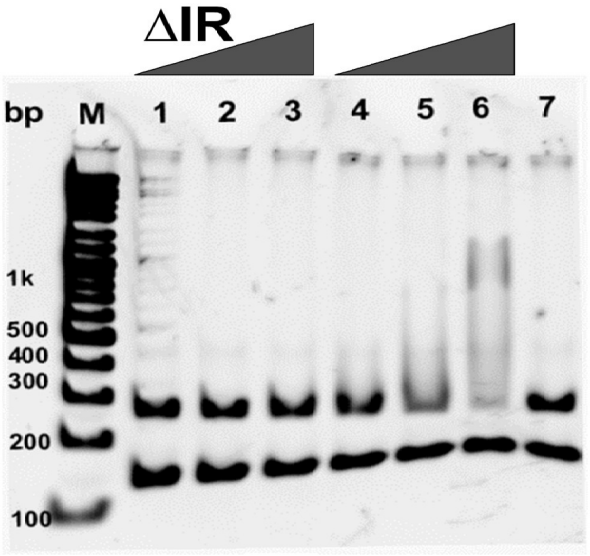
Lane 4 : 20 pmol MbrC protein added to intact SMU_1006 promoter

Lane 5 : 5 pmol MbrC protein added to intact SMU_1006 promoter with deleted MbrC binding motif

Lane 6 : 10 pmol MbrC protein added to intact SMU_1006 promoter with deleted MbrC binding motif

Lane 7 : 20 pmol MbrC protein added to intact SMU_1006 promoter with deleted MbrC binding motif

Figure 29. Binding of MbrC to the promoter region of the gene SMU_1006 (positive control) was verified using Electro Mobility Shift Assays (EMSA), as already reported by Ouyang et al [95].



Lane M : Ladder

Lane 1 : 5 pmol MbrC protein added to the murMN promoter with deleted MbrC binding motif

Lane 2 : 10 pmol MbrC protein added to the murMN promoter with deleted MbrC binding motif

Lane 3 : 20 pmol MbrC protein added to the murMN promoter with deleted MbrC binding motif

Lane 4 : 5 pmol MbrC protein added to the intact murMN promoter

Lane 5 : 10 pmol MbrC protein added to the intact murMN promoter

Lane 6 : 20 pmol MbrC protein added to the intact murMN promoter

Lane 7 (Negative control) : 20 pmol VicR protein added to the intact murMN promoter

Figure 30. EMSA also provided the verification of the in-vitro binding of the MbrC protein to the promoter region of the predicted target murMN-SMU_718c operon via the putative binding site thus confirming that the latter is a transcriptional regulatory target of MbrC. The triangles indicate increasing concentrations of MbrC in the binding reactions. IR indicates target DNA fragments lacking the MbrC binding site.

Similarly, MbrC also binds to the promoter region of SMU_718 as indicated by a gelshift at MbrC concentrations higher than or equal to 5pmol (**Figure 30**). With increasing amounts of MbrC (5-20 pmol) present in the binding reaction mixture, a shift from 250 bp (upper DNA band) to around 1 kbp was observed. No shift was observed in the lane corresponding to the reaction mixture containing an unrelated transcriptional regulator VicR (whose binding site was not present in the *murMN*-SMU_718 promoter) confirming the absence of non-specific binding. Furthermore, the deletion of the MbrC binding site in the promoter region of the *murMN*-SMU_718 operon also abolished the binding of MbrC to its DNA target. Thus the observations from the the EMSA studies demonstrate that the *murMN*-SMU_718 operon is a direct target of the MbrC response regulator as predicted from the regulatory response network. The two substitutions present in the MbrC binding site upstream of the *murMN*-SMU_718c operon could reflect a less stringent binding of MbrC. Accordingly, the shift of the SMU_718c target DNA occurred at a higher MbrC protein concentration (10 pmol) than that (5 pmol) observed for the corresponding SMU_1006 target DNA.

Our finding that MbrC directly controls the expression of two enzymes involved in a central step of cell wall synthesis further highlights the importance of this response regulator for the cell envelope stress response and for maintaining cell wall homeostasis. The reconstitution of the Gram positive bacterial cell wall during growth is intricately linked to the synthesis of peptidoglycan polymer chains [358]. Both *murM and murN* encode alanine adding enzymes involved in the branched peptide peptidoglycan biosynthetic pathway and have also been reported to be primary determinants of the peptidoglycan stem peptide structure [354, 357] which in turn affects various virulence properties such as antibiotic resistance [359-361] and pathogenesis [362]. Thus, although the transcriptional induction of the *murMN*-SMU_718 operon and the *mbrC* gene is relatively weak following carolacton treatment, it might still be significant since only a few copies of the enzymes might potentially be required to catalyze

the peptidoglycan biosynthetic reactions. Post transcriptional regulation of enzymes involved in cell wall synthesis and cell division might also significantly contribute to a quick and efficient adaptation of the cell to maintain cell wall homeostasis. Thus small transcriptional changes might result in strong changes of enzyme activity. Induction of the MbrC regulon by carolacton is reminiscent of the previously observed bacitracin induced cell envelope stress response in *S. mutans* [95] mediated by this regulon. This finding further points out the influence of carolacton on cell wall synthesis as reported previously [10, 379].

5.2 Deletion mutants of "key" transcriptional regulators

Based on the inferred TRRN and co-expression networks, single gene deletion mutants corresponding to five different transcriptional regulators (CysR, Rgg, GlnR, SpxA, and FabT) were constructed. Next to that of CodY, CysR (SMU_852) was identified as the regulator with the highest connectivity in the regulatory response network. *rgg* (SMU_1509) was identified as the only transcriptional regulator encoding gene among the CysR co-regulated genes and hence was chosen as a candidate for deletion. Alterations in the cell wall metabolism could play a critical role in the response of *S. mutans* biofilms to carolacton due to the observed membrane and cell wall damage [6, 10]. GlnR was chosen as a possible knock-out target due to its importance in peptidoglycan biosynthesis and cell wall metabolism via many of the glutamine-metabolism and transport genes which it modulates as observed in the TRRN. *spxA* (SMU_1142) and *fabT* (SMU_1745) were identified as the top two nodes based on connectivities in the co-expression network. All the selected "key" regulators are potentially involved in cell wall metabolism, glutamine metabolism, fatty acid metabolism, or stress tolerance and survival.

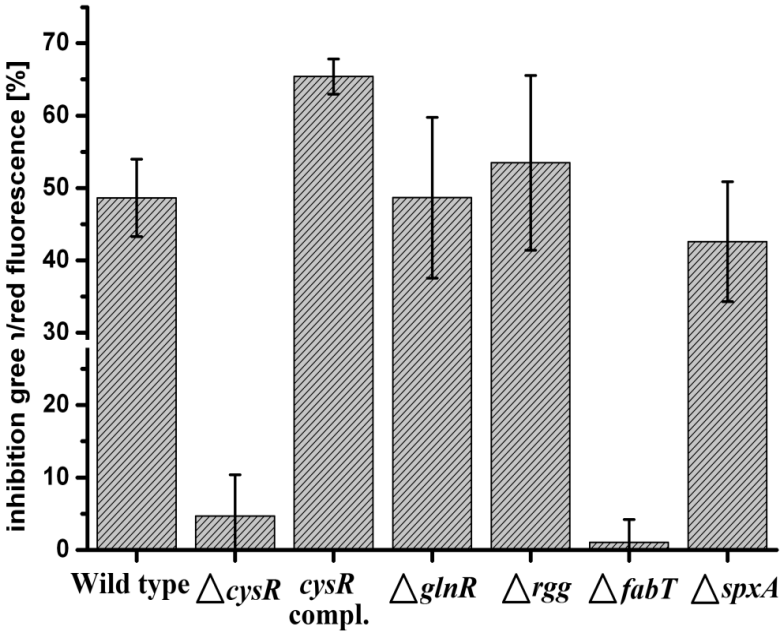


Figure 31. Effect of deleting five "key" transcriptional regulators and sensitivity of the *cysR* deletion mutant to carolacton treatment. Inhibition of viability caused by carolacton treatment was tested for the biofilms of gene deletion mutants of 5 key transcriptional regulators identified from network analysis. The inhibition of viability was determined by live/dead staining of 20h-old static biofilms of mutant, wild type and *cysR* complementation strains under carolacton treatment and is expressed as inhibition of the green/red fluorescence ratio. The bars show the mean of three independent biological replicates.

The susceptibility to carolacton treatment of 20 h old static biofilms of the corresponding key regulator gene deletion mutants was tested using Live/Dead viability staining. As shown in **Figure 31,** carolacton only marginally reduced the viability (~5-10% inhibition) of the biofilms of the *cysR* gene deletion strain, while complementation of *cysR* in trans fully restored the carolacton sensitive phenotype of the wildtype (approximately 45-55% inhibition of viability). It was previously shown that carolacton exclusively damages growing biofilms [10]. The observed strong loss of sensitivity of the *cysR* gene deletion strain to carolacton was not biased due to poor or significantly slower growing mutant biofilms since the final growth

yield and doubling time of the *cysR* gene deletion mutant was only slightly reduced in comparison to the wildtype (data not shown). Strikingly, the *spxA* gene deletion strain displayed impaired ability to grow under acidic conditions [363], a condition known to be essential for the carolacton induced membrane damage in *S. mutans* [10]. However, all the tested strains with the *fabT* or *cysR* deletion showed similar susceptibilities to carolacton relative to the wildtype (**Figure 31**). For the *fabT* deletion mutant, no final conclusion regarding its sensitivity to carolacton treatment could be drawn since the mutant strain grew very poorly under the tested conditions and formed very thin biofilms. Since growth is a prerequisite for the membrane damage caused by carolacton treatment [10], the results of the Live/Dead staining for the *fabT* mutant should be interpreted with caution. However, the membrane integrity of the *fabT* deletion mutant was highly compromised (data not shown), as determined by live/dead staining. The importance of *fabT* for maintaining membrane integrity in *S. mutans* biofilms might be indicative of its involvement in the membrane damage caused by carolacton treatment. Hence from the intial susceptibility studies of mutant biofilms under carolacton treatment, *cysR* could be identified as being essential for the sensitivity of growing *S. mutans* biofilms to carolacton.

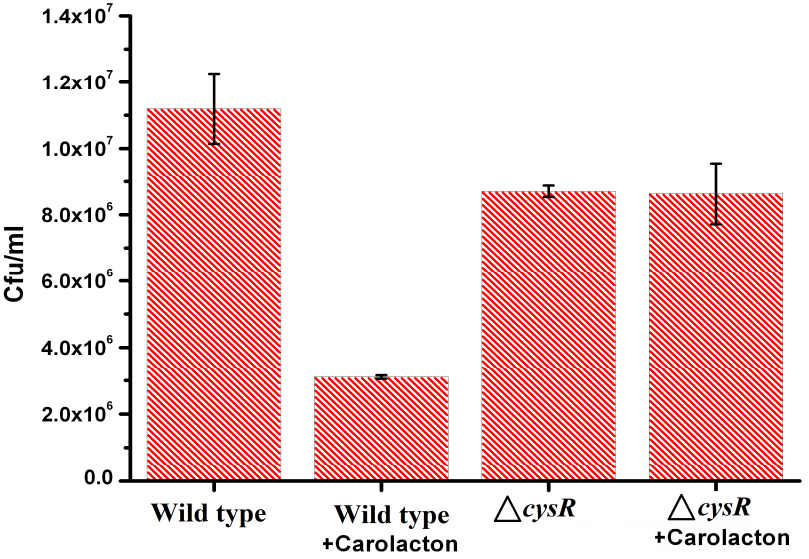


Figure 32. The effect of carolacton treatment on the number of colony forming units (cfu's) of biofilms of the *S. mutans* UA159 wildtype and *cysR* gene deletion mutant was also investigated. The Cfu experiment was repeated in two biological replicates. The sequences of primers used for generating the deletion mutants are given in Table 5.

5.3 The role of CysR in the response of *S. mutans* biofilms to carolacton

The almost complete loss of sensitivity of the *cysR* mutant to carolacton treatment was independently confirmed by the determination of colony forming units of carolacton treated and untreated biofilms of the wild type and *cysR* gene deletion strains. While the wildtype showed a reduction of CFUs of approximately 75%, the carolacton treated and untreated mutant cells showed almost no difference in the amount of colony forming units (**Figure 32**). Thus the experiment clearly confirms the essential role of CysR in the response of *S. mutans* biofilms to carolacton treatment, as implied from the TRRN. The genome of *S. mutans* encodes 4 LysR-type transcriptional regulators, of which 3 (CysR, MetR, HomR) are

phylogenetically linked and control the supply of sulfur amino acids [352, 364]. CysR is known to modulate the expression of genes involved in cysteine metabolism in *S. mutans* [352]. Intial results from the network and experimental analysis show that apart from its role in cysteine metabolism, CysR is essential for the response of *S. mutans* biofilms to the biofilm inhibitor carolacton. This implies a potential role of CysR in cell wall maintenance, as this trait was shown to be impaired in carolacton treated biofilm cells and cell wall changes are most likely responsible for cell death caused under acidic conditions [10].

The inferred transcriptional regulatory targets of CysR, as identified after integration of contextual coexpression correlation and binding motif information, indeed comprise genes involved in maintaining cell wall homeostasis. Genes SMU_984 and SMU_609 encode two autolysins which were not characterised so far. Interestingly, both genes were immediately (5 min after treatment) and constantly upregulated after carolacton treatment and belong to the strongest regulated genes of our time-resolved transcriptome study. SMU_609 encodes a 40K cell wall protein which was shown to exhibit murein hydrolase activity using a peptidoglycan zymogram assay and is likely linked to the cell surface via the sortase system [365, 350]. As cell wall synthesis during growth depends on a balanced interplay between build-up and breakdown of the cell-wall, enhanced autolysin activity might promote cell lysis [366] and thus explain the influence of CysR on cell death. However, deletion of SMU_609 and SMU_984 did not cause a carolacton insensitive phenotype (data not shown). Another CysR regulated gene, *rgpG* (SMU_246), encodes a putative UDP-N-acetylglucosamine undecaprenyl-phosphate GlcNac 1-phosphate transferase. This enzyme catalyses the first membrane localized step of the biosynthesis of various polymers of the bacterial cell wall. Imbalanced expression of a bacteriocin and its corresponding immunity protein could also have a profound impact on cellular viability [367, 368]. However, SMU_613 was found to be only slightly modulated at the transcriptional level. The other potential target genes include

the operon SMU_1074-SMU_1077 which encode putative metabolic enzymes in pathways related to cysteine metabolism and is concurrent with the already known role of *cysR* in cysteine metabolism [352].

It should also be noted that *cysR* and its paralog *cpsY* are located in close vicinity to the genes of the de novo pyrimidine biosynthesis pathway on the genome of *S. mutans*. The pyrimidine biosynthesis genes were shown to be the strongest upregulated genes 5 min after carolacton treatment. The functionality of pyrimidine metabolism, especially of the *pyr* operons upregulated specifically under carolacton treatment, could possibly be linked to the availability and synthesis of cell wall precursors. In prokaryotes, genes encoding products involved in the same or related biological functions are often located in close neighbourhood on the genome [202, 369, 370].

Moreover, for the *S. iniae cysR* ortholog *cpsY*, it was shown that this LysR type transcriptional regulator (LTTR) induced cell wall changes essential for the intracellular survival of this invasive pathogen in neutrophils [371]. Interestingly, in the study of Allen *et al* [371], it was determined that deletion of *cpsY* alters the cell surface charge, muropeptide composition and susceptibility to lysozyme treatment. This is fully in accordance with the current understanding of the mode of action of carolacton [10]. However further experiments are necessary to verify the targets of CysR experimentally (e.g. EMSA experiments with purified CysR protein) and to fully address the role of *cysR* in cell wall metabolism e.g. by knocking out the CysR regulated autolysins. Altogether, our data strongly suggest a so far unidentified role of the CysR transcriptional regulator in cell wall homeostasis.

A highly interesting aspect of the network analysis is the finding that CysR potentially regulates another transcriptional regulator (Rgg). As the carolacton specific regulatory response network contains albeit only two levels of hierarchy, CysR can thus be considered as a potential global regulator. Rgg transcriptional regulators associated with small hydrophobic

peptides (SHP) were identified to represent a part of a novel quorum sensing mechanism in Streptococci [353, 372]. Rgg proteins have been identified and conserved in nearly all streptococci [373]. This mechanism works independently of a TCS, but senses the signalling molecule inside the cell after its internalisation via an oligopeptide permease [353]. But the Rgg knockout strain showed comparable susceptibility to carolacton treatment as the wildtype strain (**Figure 31**) thus excluding the possibility of Rgg being a critical regulator mediating the observed physiological effects of carolacton.

With respect to *cysR* however, its deletion mutant displayed a loss of sensitivity to carolacton as a result of a reduction in the inhibition of viability caused by carolacton. Nevertheless, CysR does not represent the primary target of carolacton as the treated cells of the *cysR* gene deletion mutant still exhibit an inhibition of about 10%. Moreover, phenotypic differences between carolacton treated and untreated *cysR* mutant cells were observed, indicating that carolacton can still bind to its target(s) and induce changes in the cellular morphology. Nevertheless, the lethal effects of carolacton treatment apparently rely to a large extent on the presence of an intact copy of the *cysR* gene. Another interesting question is the link of CysR to the PknB controlled regulon, as *pknB* was also shown to be essential for carolacton mediated cell death [10]. With PknB being a global regulator of cell division [374] and virulence attributes [103], the regulons of PknB and CysR might converge at the level of the controlled autolysins. The CysR coregulated genes SMU_984, and SMU_609 were 2 of the 3 strongest modulated genes in the transcriptome analysis of an exponentially growing *pknB* deletion strain [102]. However, as the deletion mutants of SMU_609 and SMU_984 are still susceptible to carolacton treatment, this potential overlap of the regulatory systems does not provide an explanation why both the *pknB* and *cysR* deletion mutant biofilms exhibited almost no sensitivity to carolacton treatment.

Since LTTRs show a high degree of sequence conservation, control important metabolic pathways of sulphur containing amino acids, and since sulfur is essential for the active sites of many enzymes, the transcriptional regulator CysR might also be an attractive drug target. Coeffector binding is required for its transcriptional activation, and the likely cofactor could be O-Acetylserine (OAS) [352]. Thus, small molecule inhibitors could potentially be designed to compete with the co-effector molecule for the binding site on this LTTR and to finally block its biological function. To conclude, the results from this study points out that the role of CysR may be much more diverse and important than previously expected. Thus CysR might represent an attractive novel drug target in streptococci.

5.4 Linkage between the PknB and CysR regulons

Deletion of *cysR* almost completely prevents cell death in carolacton treated biofilms, a finding previously reported for the gene deletion strain of the Serine Threonine Protein Kinase (STPK) encoding gene *pknB* as well [10]. This instantly raises the question whether the regulons of these two regulator proteins overlap or if PknB is located upstream of CysR in the regulatory cascade and thus controls the latter's activity. Both regulators control the expression of the two carolacton responsive autolysins SMU_984 and SMU_609, which might explain the shared insensitive phenotype to carolacton of their mutant strains.

However, in this study, the focus is only on the transcriptional network analysis, while PknB modulates its target genes via protein phosphorylation at the post-transcriptional level. Thus the question whether and on which level the modulated networks of PknB and CysR converge and if they represent the essential pathway for the carolacton mediated cell death cannot be fully answered from the transcriptional regulatory network analysis. To quantitatively understand the effects of carolacton on cell metabolism and to determine the missing layers in the regulatory cascades of CysR and PknB, it is necessary to consider the regulatory effects of

small non-coding RNAs [375] and other post-transcriptional and post-translational modifications [146, 376] including modulation by other signalling pathways [377]. Transcriptional analysis is generally lacking in information about these important regulatory mechanisms which have been shown to be essential for cellular metabolism and homeostasis [146, 375, 377]. Nevertheless, the high merit of the transcriptional network analysis is demonstrated here to identify a key regulator mediating cell death in *S. mutans* biofilms in response to the biofilm inhibitor carolacton.

Chapter 6. Conclusion

A combination of transcriptional network prediction and experimental verification was used to analyse the response of *S. mutans* biofilms upon treatment with the biofilm inhibitor carolacton. As per current literature, this is the first study inferring a genome-wide transcriptional regulatory response network for *S. mutans* using heterogeneous data including a time-series transcriptomic dataset and transcription factor binding site information. Based on the inferences from the regulatory network, important predicted transcriptional regulatory interactions between the cell envelope stress modulating TCS response regulator MbrC and the operon harbouring the *murMN* genes encoding cell wall metabolism associated enzymes were experimentally verified . CysR, a regulator known to modulate cysteine metabolism, was predicted and experimentally verified to be an essential key regulator for the mode of action of the biofilm inhibitor carolacton. The results [380] from this study strongly suggest a role of *cysR* in cell wall metabolism, cell division and cell surface biogenesis, apart from its already known role in cysteine metabolism and sulfur supply in *S. mutans,* thus highlighting its potential as an attractive novel streptococcal drug target. The high predictive power of the network strategy used was also demonstrated.

This chapter has been adapted from the publication P. Sudhakar, M. Reck, W. Wang, F. He, I. Wagner-Döbler, AP. Zeng. Construction and verification of the transcriptional regulatory network of the human dental pathogen *Streptococcus mutans* upon treatment with the biofilm inhibitor Carolacton. BMC Genomics 2014, 15:362.

References

1. Stoodley P, Sauer K, Davies DG, Costerton JW: **Biofilms as complex differentiated communities**. *Annu. Rev. Microbiol.* 2002, **56**:187-209.
2. Dufour D, Leung V, Levesque CM: **Bacterial biofilm: structure, function, and antimicrobial resistance**. *Endodont. Top.* 2012, **22**: 2-16.
3. Lemos JA, Burne RA: **A model of efficiency: stress tolerance by *Streptococcus mutans***. *Microbiology (Reading, Engl.)* 2008, **154**:3247-3255.
4. Li YH, Tang N, Aspiras MB, Lau PC, Lee JH, Ellen RP, Cvitkovitch DG: **A quorum-sensing signaling system essential for genetic competence in *Streptococcus mutans* is involved in biofilm formation**. *J.Bacteriol.* 2002, **184**:2699-2708.
5. Nomura R, Nakano K, Nemoto H, Fujita K, Inagaki S, Takahashi T, Taniguchi K, Takeda M, Yoshioka H, Amano A, Ooshima T: **Isolation and characterization of *Streptococcus mutans* in heart valve and dental plaque specimens from a patient with infective endocarditis**. *J. Med. Microbiol.* 2006, **55**:1135-1140.
6. Kunze B, Reck M, Dötsch A, Lemme A, Schummer D, Irschik H, Steinmetz H, Wagner-Döbler I: **Damage of *Streptococcus mutans* biofilms by carolacton, a secondary metabolite from the myxobacterium *Sorangium cellulosum***. *BMC Microbiol.* 2010, **10**:199.
7. Jansen R, Irschik H, Huch V, Schummer D, Steinmetz H, Bock M, Schmidt T, Kirschning A, Müller R: **Carolacton – A Macrolide Ketocarbonic Acid that Reduces Biofilm Formation by the Caries- and Endocarditis-Associated Bacterium *Streptococcus mutans***. *Eur. J. Org. Chem.* 2010, **7**: 1284-1289.
8. Schmidt T, Kirschning A: **Total synthesis of carolacton, a highly potent biofilm inhibitor**. *Angew. Chem. Int. Ed. Engl.* 2012, **51**:1063-1066.
9. Roychoudhury S, Zielinski NA, Ninfa AJ, Allen NE, Jungheim LN, Nicas TI, Chakrabarty AM: **Inhibitors of two-component signal transduction systems: inhibition of alginate gene activation in *Pseudomonas aeruginosa***. *Proc. Natl. Acad. Sci. U.S.A.* 1993, **90**:965-969.
10. Reck M, Rutz K, Kunze B, Tomasch J, Surapaneni SK, Schulz S, Wagner-Döbler I: **The biofilm inhibitor carolacton disturbs membrane integrity and cell division of *Streptococcus mutans* through the serine/threonine protein kinase PknB**. *J. Bacteriol.* 2011,**193**: 5692-5706.
11. He F, Balling R, Zeng A: **Reverse engineering and verification of gene networks: principles, assumptions, and limitations of present methods and future perspectives**. *J. Biotechnol.* 2009, **144**: 190-203.
12. Bansal M, Belcastro V, Ambesi-Impiombato A, Di Bernardo D: **How to infer gene networks from expression profiles**. *Mol. Syst. Biol.* 2007, **3**:78.
13. Cabusora L, Sutton E, Fulmer A, Forst CV: **Differential network expression during drug and stress response**. *Bioinformatics* 2005, **21**:2898-2905.
14. Bansal M, Della Gatta G, Di Bernardo D: **Inference of gene regulatory networks and compound mode of action from time course gene expression profiles**. *Bioinformatics* 2006, **22**:815-822.
15. Aluru M, Zola J, Nettleton D, Aluru S: **Reverse engineering and analysis of large genome-scale gene networks**. *Nucleic Acids Res.* 2013, **41**:e24.
16. Faith JJ, Hayete B, Thaden JT, Mogno I, Wierzbowski J, Cottarel G, Kasif S, Collins JJ, Gardner TS: **Large-scale mapping and validation of *Escherichia coli* transcriptional regulation from a compendium of expression profiles**. *PLoS Biol.* 2007, **5**:e8.
17. He F, Chen H, Probst-Kepper M, Geffers R, Eifes S, Del Sol A, Schughart K, Zeng A, Balling R: **PLAU inferred from a correlation network is critical for suppressor function of regulatory T cells**. *Mol. Syst. Biol.* 2012, **8**:624.
18. O'Rourke KP, Shaw JD, Pesesky MW, Cook BT, Roberts SM, Bond JP, Spatafora GA: **Genome-wide characterization of the SloR metalloregulome in *Streptococcus mutans***. *J. Bacteriol.* 2010, **192**:1433-1443.
19. Ahn S, Wen ZT, Burne RA: **Effects of oxygen on virulence traits of *Streptococcus mutans***. *J. Bacteriol.* 2007,**189**: 8519-8527.

20. Ajdić D, Pham VTT: **Global transcriptional analysis of *Streptococcus mutans* sugar transporters using microarrays**. *J. Bacteriol.* 2007,**189**: 5049-5059.
21. Liu J, Wu C, Huang I, Merritt J, Qi F: **Differential response of *Streptococcus mutans* towards friend and foe in mixed-species cultures**. *Microbiology (Reading, Engl.)* 2011, **157**:2433-2444.
22. He F, Zeng A: **In search of functional association from time-series microarray data based on the change trend and level of gene expression**. *BMC Bioinformatics* 2006, **7**:69.
23. Ajdić D, McShan WM, McLaughlin RE, Savić G, Chang J, Carson MB, Primeaux C, Tian R, Kenton S, Jia H, Lin S, Qian Y, Li S, Zhu H, Najar F, Lai H, White J, Roe BA, Ferretti JJ: **Genome sequence of *Streptococcus mutans* UA159, a cariogenic dental pathogen**. *Proc. Natl. Acad. Sci. U.S.A.* 2002, **99**:14434-14439.
24. Song L, Sudhakar P, Wang W, Conrads G, Brock A, Sun J, Wagner-Döbler I, Zeng A: **A genome-wide study of two-component signal transduction systems in eight newly sequenced mutans streptococci strains**. *BMC Genomics* 2012, **13**:128.
25. Song L, Wang W, Conrads G, Rheinberg A, Sztajer H, Reck M, Wagner-Döbler I, Zeng A: **Genetic variability of mutans streptococci revealed by wide whole-genome sequencing**. *BMC Genomics* 2013, **14**:430.
26. Aikawa C, Furukawa N, Watanabe T, Minegishi K, Furukawa A, Eishi Y, Oshima K, Kurokawa K, Hattori M, Nakano K, Maruyama F, Nakagawa I, Ooshima T: **Complete genome sequence of the serotype k *Streptococcus mutans* strain LJ23**. *J. Bacteriol.* 2012,**194**: 2754-2755.
27. Maruyama F, Kobata M, Kurokawa K, Nishida K, Sakurai A, Nakano K, Nomura R, Kawabata S, Ooshima T, Nakai K, Hattori M, Hamada S, Nakagawa I: **Comparative genomic analyses of *Streptococcus mutans* provide insights into chromosomal shuffling and species-specific content**. *BMC Genomics* 2009, **10**:358.
28. Loesche WJ: **Role of *Streptococcus mutans* in human dental decay**. *Microbiol. Rev.* 1986, **50**:353-380.
29. Lemos JAC, Abranches J, Burne RA: **Responses of cariogenic streptococci to environmental stresses**. *Curr. Issues. Mol. Biol.* 2005, **7**:95-107.
30. Hale JDF, Ting Y, Jack RW, Tagg JR, Heng NCK: **Bacteriocin (mutacin) production by *Streptococcus mutans* genome sequence reference strain UA159: elucidation of the antimicrobial repertoire by genetic dissection**. *Appl. Environ. Microbiol.* 2005, **71**:7613-7617.
31. Balakrishnan M, Simmonds RS, Kilian M, Tagg JR: **Different bacteriocin activities of *Streptococcus mutans* reflect distinct phylogenetic lineages**. *J. Med. Microbiol.* 2002, **51**:941-948.
32. Xie Z, Okinaga T, Niu G, Qi F, Merritt J: **Identification of a novel bacteriocin regulatory system in *Streptococcus mutans***. *Mol. Microbiol.* 2010, **78**:1431-1447.
33. Kolenbrander PE: **Oral microbial communities: biofilms, interactions, and genetic systems**. *Annu. Rev. Microbiol.* 2000, **54**:413-437.
34. Kolenbrander PE, Palmer RJ, Rickard AH, Jakubovics NS, Chalmers NI, Diaz PI: **Bacterial interactions and successions during plaque development**. *Periodontol. 2000* 2006, **42**:47-79.
35. Costerton JW, Cheng KJ, Geesey GG, Ladd TI, Nickel JC, Dasgupta M, Marrie TJ: **Bacterial biofilms in nature and disease**. *Annu. Rev. Microbiol.* 1987,**41**: 435-464.
36. Costerton JW, Stewart PS, Greenberg EP: **Bacterial biofilms: a common cause of persistent infections**. *Science* 1999, **284**:1318-1322.
37. Jefferson KK: **What drives bacteria to produce a biofilm?** *FEMS Microbiol. Lett.* 2004, **236**:163-173.
38. Spoering AL, Gilmore MS: **Quorum sensing and DNA release in bacterial biofilms**. *Curr. Opin. Microbiol.* 2006, **9**:133-137.
39. Li Y, Tian X, Layton G, Norgaard C, Sisson G: **Additive attenuation of virulence and cariogenic potential of *Streptococcus mutans* by simultaneous inactivation of the ComCDE quorum-sensing system and HK/RR11 two-component regulatory system**. *Microbiology (Reading, Engl.)* 2008, **154**:3256-3265.
40. Senadheera D, Cvitkovitch DG: **Quorum sensing and biofilm formation by *Streptococcus mutans***. *Adv. Exp. Med. Biol.* 2008, **631**:178-188.
41. Costerton JW: **Introduction to biofilm**. *Int. J. Antimicrob. Agents* 1999, **11**:217-21; discussion 237-9.

42. Donlan RM, Costerton JW: **Biofilms: survival mechanisms of clinically relevant microorganisms**. *Clin. Microbiol. Rev.* 2002, **15**:167-193.
43. Keren I, Kaldalu N, Spoering A, Wang Y, Lewis K: **Persister cells and tolerance to antimicrobials**. *FEMS Microbiol. Lett.* 2004, **230**:13-18.
44. Lewis K: **Riddle of biofilm resistance**. *Antimicrob. Agents Chemother.* 2001, **45**:999-1007.
45. Høiby N, Bjarnsholt T, Givskov M, Molin S, Ciofu O: **Antibiotic resistance of bacterial biofilms**. *Int. J. Antimicrob. Agents* 2010, **35**:322-332.
46. Yoshida A, Kuramitsu HK: **Multiple *Streptococcus mutans* Genes Are Involved in Biofilm Formation**. *Appl. Environ. Microbiol.* 2002, **68**:6283-6291.
47. Chia JS, Yeh CY, Chen JY: **Identification of a fibronectin binding protein from *Streptococcus mutans***. *Infect. Immun.* 2000, **68**:1864-1870.
48. Ray CA, Gfell LE, Buller TL, Gregory RL: **Interactions of *Streptococcus mutans* fimbria-associated surface proteins with salivary components**. *Clin. Diagn. Lab. Immunol.* 1999, **6**:400-404.
49. Mitchell TJ: **The pathogenesis of streptococcal infections: from tooth decay to meningitis**. *Nat. Rev. Microbiol.* 2003, **1**:219-230.
50. Gibbons RJ: **Role of adhesion in microbial colonization of host tissues: a contribution of oral microbiology**. *J. Dent. Res.* 1996, **75**:866-870.
51. Lee SF, Progulske-Fox A, Erdos GW, Piacentini DA, Ayakawa GY, Crowley PJ, Bleiweis AS: **Construction and characterization of isogenic mutants of *Streptococcus mutans* deficient in major surface protein antigen P1 (I/II)**. *Infect. Immun.* 1989, **57**:3306-3313.
52. Bowen WH, Schilling K, Giertsen E, Pearson S, Lee SF, Bleiweis A, Beeman D: **Role of a cell surface-associated protein in adherence and dental caries**. *Infect. Immun.* 1991, **59**:4606-4609.
53. Zhu L, Kreth J, Cross SE, Gimzewski JK, Shi W, Qi F: **Functional characterization of cell-wall-associated protein WapA in *Streptococcus mutans***. *Microbiology (Reading, Engl.)* 2006, **152**:2395-2404.
54. Hirasawa M, Kiyono H, Shiota T, Hull RA, Curtiss R, Michalek SM, McGhee JR: **Virulence of *Streptococcus mutans*: restoration of pathogenesis of a glucosyltransferase-defective mutant (C4)**. *Infect. Immun.* 1980, **27**:915-921.
55. Hamada S, Slade HD: **Biology, immunology, and cariogenicity of *Streptococcus mutans***. *Microbiol. Rev.* 1980, **44**:331-384.
56. Tsumori H, Kuramitsu H: **The role of the *Streptococcus mutans* glucosyltransferases in the sucrose-dependent attachment to smooth surfaces: essential role of the GtfC enzyme**. *Oral Microbiol. Immunol.* 1997, **12**:274-280.
57. Ooshima T, Matsumura M, Hoshino T, Kawabata S, Sobue S, Fujiwara T: **Contributions of three glycosyltransferases to sucrose-dependent adherence of *Streptococcus mutans***. *J. Dent. Res.* 2001, **80**:1672-1677.
58. Banas JA, Vickerman MM: **Glucan-binding proteins of the oral streptococci**. *Crit. Rev. Oral Biol. Med.* 2003, **14**:89-99.
59. Russell RR: **Glucan-binding proteins of *Streptococcus mutans* serotype c**. *J. Gen. Microbiol.* 1979, **112**:197-201.
60. Banas JA, Russell RR, Ferretti JJ: **Sequence analysis of the gene for the glucan-binding protein of *Streptococcus mutans* Ingbritt**. *Infect. Immun.* 1990, **58**:667-673.
61. Mattos-Graner RO, Jin S, King WF, Chen T, Smith DJ, Duncan MJ: **Cloning of the *Streptococcus mutans* gene encoding glucan binding protein B and analysis of genetic diversity and protein production in clinical isolates**. *Infect. Immun.* 2001, **69**:6931-6941.
62. Mattos-Graner RO, Porter KA, Smith DJ, Hosogi Y, Duncan MJ: **Functional analysis of glucan binding protein B from *Streptococcus mutans***. *J. Bacteriol.* 2006, **188**:3813-3825.
63. Ma Y, Lassiter MO, Banas JA, Galperín MY, Taylor KG, Doyle RJ: **Multiple glucan-binding proteins of *Streptococcus sobrinus***. *J. Bacteriol.* 1996, **178**:1572-1577.
64. Shah DSH, Russell RRB: **A novel glucan-binding protein with lipase activity from the oral pathogen *Streptococcus mutans***. *Microbiology (Reading, Engl.)* 2004, **150**:1947-1956.
65. Russell RR, Donald AC, Douglas CW: **Fructosyltransferase activity of a glucan-binding protein from *Streptococcus mutans***. *J. Gen. Microbiol.* 1983, **129**:3243-3250.

66. Webb AJ, Homer KA, Hosie AHF: **A phosphoenolpyruvate-dependent phosphotransferase system is the principal maltose transporter in *Streptococcus mutans***. *J. Bacteriol.* 2007, **189**:3322-3327.
67. Poy F, Jacobson GR: **Evidence that a low-affinity sucrose phosphotransferase activity in *Streptococcus mutans* GS-5 is a high-affinity trehalose uptake system**. *Infect. Immun.* 1990, **58**:1479-1480.
68. Postma PW, Lengeler JW, Jacobson GR: **Phosphoenolpyruvate:carbohydrate phosphotransferase systems of bacteria**. *Microbiol. Rev.* 1993, **57**:543-594.
69. Slee AM, Tanzer JM: **Effect of growth conditions on sucrose phosphotransferase activity of *Streptococcus mutans***. *Infect. Immun.* 1980, **27**:922-927.
70. Vadeboncoeur C, Thibault L, Neron S, Halvorson H, Hamilton IR: **Effect of growth conditions on levels of components of the phosphoenolpyruvate:sugar phosphotransferase system in *Streptococcus mutans* and *Streptococcus sobrinus* grown in continuous culture**. *J. Bacteriol.* 1987, **169**:5686-5691.
71. Abranches J, Nascimento MM, Zeng L, Browngardt CM, Wen ZT, Rivera MF, Burne RA: **CcpA regulates central metabolism and virulence gene expression in *Streptococcus mutans***. *J. Bacteriol.* 2008, **190**:2340-2349.
72. Abranches J, Chen YM, Burne RA: **Characterization of *Streptococcus mutans* strains deficient in EIIAB Man of the sugar phosphotransferase system**. *Appl. Environ. Microbiol.* 2003, **69**:4760-4769.
73. Cvitkovitch DG, Boyd DA, Thevenot T, Hamilton IR: **Glucose transport by a mutant of *Streptococcus mutans* unable to accumulate sugars via the phosphoenolpyruvate phosphotransferase system**. *J. Bacteriol.* 1995, **177**:2251-2258.
74. Sheng J, Marquis RE: **Enhanced acid resistance of oral streptococci at lethal pH values associated with acid-tolerant catabolism and with ATP synthase activity**. *FEMS Microbiol. Lett.* 2006, **262**:93-98.
75. Griswold AR, Jameson-Lee M, Burne RA: **Regulation and physiologic significance of the agmatine deiminase system of *Streptococcus mutans* UA159**. *J. Bacteriol.* 2006, **188**:834-841.
76. Chen P, Chen YM, Yu S, Sher S, Lai C, Chia J: **Role of GlnR in acid-mediated repression of genes encoding proteins involved in glutamine and glutamate metabolism in *Streptococcus mutans***. *Appl. Environ. Microbiol.* 2010, **76**:2478-2486.
77. Dashper SG, Riley PF, Reynolds EC: **Characterization of glutamine transport in *Streptococcus mutans***. *Oral Microbiol. Immunol.* 1995, **10**:183-187.
78. Korithoski B, Krastel K, Cvitkovitch DG: **Transport and metabolism of citrate by *Streptococcus mutans***. *J. Bacteriol.* 2005, **187**:4451-4456.
79. Sheng J, Marquis RE: **Malolactic fermentation by *Streptococcus mutans***. *FEMS Microbiol. Lett.* 2007, **272**:196-201.
80. Sheng J, Baldeck JD, Nguyen PTM, Quivey RG, Marquis RE: **Alkali production associated with malolactic fermentation by oral streptococci and protection against acid, oxidative, or starvation damage**. *Can. J. Microbiol.* 2010, **56**:539-547.
81. Bhagwat SP, Nary J, Burne RA: **Effects of mutating putative two-component systems on biofilm formation by *Streptococcus mutans* UA159**. *FEMS Microbiol. Lett.* 2001, **205**:225-230.
82. Liu Y, Burne RA: **Multiple two-component systems of *Streptococcus mutans* regulate agmatine deiminase gene expression and stress tolerance**. *J. Bacteriol.* 2009, **191**:7363-7366.
83. Kawada-Matsuo M, Shibata Y, Yamashita Y: **Role of two component signaling response regulators in acid tolerance of *Streptococcus mutans***. *Oral Microbiol. Immunol.* 2009, **24**:173-176.
84. Lee SF, Delaney GD, Elkhateeb M: **A two-component covRS regulatory system regulates expression of fructosyltransferase and a novel extracellular carbohydrate in *Streptococcus mutans***. *Infect. Immun.* 2004, **72**:3968-3973.
85. Lévesque CM, Mair RW, Perry JA, Lau PCY, Li Y, Cvitkovitch DG: **Systemic inactivation and phenotypic characterization of two-component systems in expression of *Streptococcus mutans* virulence properties**. *Lett. Appl. Microbiol.* 2007, **45**:398-404.
86. Li Y, Lau PCY, Tang N, Svensäter G, Ellen RP, Cvitkovitch DG: **Novel two-component regulatory system involved in biofilm formation and acid resistance in *Streptococcus mutans***. *J. Bacteriol.* 2002, **184**:6333-6342.

87. Casino P, Rubio V, Marina A: **The mechanism of signal transduction by two-component systems**. *Curr. Opin. Struct. Biol.* 2010, **20**:763-771.
88. Capra EJ, Laub MT: **Evolution of Two-Component Signal Transduction Systems**. *Annual review of microbiology* 2012, **66**:325-347.
89. Mitrophanov AY, Groisman EA: **Signal integration in bacterial two-component regulatory systems**. *Genes Dev.* 2008, **22**:2601-2611.
90. Idone V, Brendtro S, Gillespie R, Kocaj S, Peterson E, Rendi M, Warren W, Michalek S, Krastel K, Cvitkovitch D, Spatafora G: **Effect of an orphan response regulator on *Streptococcus mutans* sucrose-dependent adherence and cariogenesis**. *Infect. Immun.* 2003, **71**:4351-4360.
91. Biswas S, Biswas I: **Regulation of the glucosyltransferase (gtfBC) operon by CovR in *Streptococcus mutans***. *J. Bacteriol.* 2006, **188**:988-998.
92. Biswas I, Drake L, Biswas S: **Regulation of gbpC expression in *Streptococcus mutans***. *J. Bacteriol.* 2007, **189**:6521-6531.
93. Dmitriev A, Mohapatra SS, Chong P, Neely M, Biswas S, Biswas I: **CovR-controlled global regulation of gene expression in *Streptococcus mutans***. *PLoS ONE* 2011, **6**:e20127.
94. Deng DM, Liu MJ, Cate JM ten, Crielaard W: **The VicRK system of *Streptococcus mutans* responds to oxidative stress**. *J. Dent. Res.* 2007, **86**:606-610.
95. Ouyang J, Tian X, Versey J, Wishart A, Li Y: **The BceABRS four-component system regulates the bacitracin-induced cell envelope stress response in *Streptococcus mutans***. *Antimicrob. Agents Chemother.* 2010, **54**:3895-3906.
96. Senadheera MD, Guggenheim B, Spatafora GA, Huang YC, Choi J, Hung DCI, Treglown JS, Goodman SD, Ellen RP, Cvitkovitch DG: **A VicRK signal transduction system in *Streptococcus mutans* affects gtfBCD, gbpB, and ftf expression, biofilm formation, and genetic competence development**. *J. Bacteriol.* 2005, **187**:4064-4076.
97. Kreth J, Hung DCI, Merritt J, Perry J, Zhu L, Goodman SD, Cvitkovitch DG, Shi W, Qi F: **The response regulator ComE in *Streptococcus mutans* functions both as a transcription activator of mutacin production and repressor of CSP biosynthesis**. *Microbiology (Reading, Engl.)* 2007, **153**:1799-1807.
98. Ahn S, Burne RA: **Effects of oxygen on biofilm formation and the AtlA autolysin of *Streptococcus mutans***. *J. Bacteriol.* 2007, **189**:6293-6302.
99. van der Ploeg JR: **Regulation of bacteriocin production in *Streptococcus mutans* by the quorum-sensing system required for development of genetic competence**. *J. Bacteriol.* 2005, **187**:3980-3989.
100. Lemos JA, Lin VK, Nascimento MM, Abranches J, Burne RA: **Three gene products govern (p)ppGpp production by *Streptococcus mutans***. *Mol. Microbiol.* 2007, **65**:1568-1581.
101. Li Y, Tang N, Aspiras MB, Lau PCY, Lee JH, Ellen RP, Cvitkovitch DG: **A quorum-sensing signaling system essential for genetic competence in *Streptococcus mutans* is involved in biofilm formation**. *J. Bacteriol.* 2002, **184**:2699-2708.
102. Banu LD, Conrads G, Rehrauer H, Hussain H, Allan E, van der Ploeg JR: **The *Streptococcus mutans* serine/threonine kinase, PknB, regulates competence development, bacteriocin production, and cell wall metabolism**. *Infect. Immun.* 2010, **78**:2209-2220.
103. Hussain H, Branny P, Allan E: **A eukaryotic-type serine/threonine protein kinase is required for biofilm formation, genetic competence, and acid resistance in *Streptococcus mutans***. *J. Bacteriol.* 2006, **188**:1628-1632.
104. Abranches J, Miller JH, Martinez AR, Simpson-Haidaris PJ, Burne RA, Lemos JA: **The collagen-binding protein Cnm is required for *Streptococcus mutans* adherence to and intracellular invasion of human coronary artery endothelial cells**. *Infect. Immun.* 2011, **79**:2277-2284.
105. Knight JA: **Review: Free radicals, antioxidants, and the immune system**. *Ann. Clin. Lab. Sci.* 2000, **30**:145-158.
106. Fridovich I: **Superoxide radical: an endogenous toxicant**. *Annu. Rev. Pharmacol. Toxicol.* 1983, **23**:239-257.
107. Taylor BL, Zhulin IB: **PAS domains: internal sensors of oxygen, redox potential, and light**. *Microbiol. Mol. Biol. Rev.* 1999, **63**:479-506.
108. Duque C, Stipp RN, Wang B, Smith DJ, Hofling JF, Kuramitsu HK, Duncan MJ, Mattos-Graner RO: **Downregulation of GbpB, a component of the VicRK regulon, affects biofilm formation**

and cell surface characteristics of *Streptococcus mutans*. *Infection and immunity* 2011, **79**:786-796.
109. Tremblay YDN, Lo H, Li Y, Halperin SA, Lee SF: **Expression of the *Streptococcus mutans* essential two-component regulatory system VicRK is pH and growth-phase dependent and controlled by the LiaFSR three-component regulatory system**. *Microbiology (Reading, Engl.)* 2009, **155**:2856-2865.
110. Loo CY, Mitrakul K, Jaafar S, Gyurko C, Hughes CV, Ganeshkumar N: **Role of a nosX homolog in *Streptococcus gordonii* in aerobic growth and biofilm formation**. *J. Bacteriol.* 2004, **186**:8193-8206.
111. Resch A, Rosenstein R, Nerz C, Götz F: **Differential gene expression profiling of Staphylococcus aureus cultivated under biofilm and planktonic conditions**. *Appl. Environ. Microbiol.* 2005, **71**:2663-2676.
112. Qi F, Merritt J, Lux R, Shi W: **Inactivation of the ciaH Gene in *Streptococcus mutans* diminishes mutacin production and competence development, alters sucrose-dependent biofilm formation, and reduces stress tolerance**. *Infect. Immun.* 2004, **72**:4895-4899.
113. Ahn S, Wen ZT, Burne RA: **Multilevel control of competence development and stress tolerance in *Streptococcus mutans* UA159**. *Infect. Immun.* 2006, **74**:1631-1642.
114. Chong P, Drake L, Biswas I: **LiaS regulates virulence factor expression in *Streptococcus mutans***. *Infect. Immun.* 2008, **76**:3093-3099.
115. Suntharalingam P, Senadheera MD, Mair RW, Lévesque CM, Cvitkovitch DG: **The LiaFSR system regulates the cell envelope stress response in *Streptococcus mutans***. *J. Bacteriol.* 2009, **191**:2973-2984.
116. Takahashi N, Nyvad B: **The role of bacteria in the caries process: ecological perspectives**. *J. Dent. Res.* 2011, **90**:294-303.
117. Kreth J, Merritt J, Shi W, Qi F: **Competition and coexistence between *Streptococcus mutans* and Streptococcus sanguinis in the dental biofilm**. *J. Bacteriol.* 2005, **187**:7193-7203.
118. Ma H, Zeng A: **Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms**. *Bioinformatics* 2003, **19**:270-277.
119. Zeng L, Wen ZT, Burne RA: **A novel signal transduction system and feedback loop regulate fructan hydrolase gene expression in *Streptococcus mutans***. *Mol. Microbiol.* 2006, **62**:187-200.
120. Burne RA, Schilling K, Bowen WH, Yasbin RE: **Expression, purification, and characterization of an exo-beta-D-fructosidase of *Streptococcus mutans***. *J. Bacteriol.* 1987, **169**:4507-4517.
121. Burne RA, Chen YY, Wexler DL, Kuramitsu H, Bowen WH: **Cariogenicity of *Streptococcus mutans* strains with defects in fructan metabolism assessed in a program-fed specific-pathogen-free rat model**. *J. Dent. Res.* 1996, **75**:1572-1577.
122. Lemos JA, Nascimento MM, Lin VK, Abranches J, Burne RA: **Global regulation by (p)ppGpp and CodY in *Streptococcus mutans***. *J. Bacteriol.* 2008, **190**:5291-5299.
123. Brückner R, Titgemeyer F: **Carbon catabolite repression in bacteria: choice of the carbon source and autoregulatory limitation of sugar utilization**. *FEMS Microbiol. Lett.* 2002, **209**:141-148.
124. Jankovic I, Brückner R: **Carbon catabolite repression of sucrose utilization in Staphylococcus xylosus: catabolite control protein CcpA ensures glucose preference and autoregulatory limitation of sucrose utilization**. *J. Mol. Microbiol. Biotechnol.* 2007, **12**:114-120.
125. Tobisch S, Zühlke D, Bernhardt J, Stülke J, Hecker M: **Role of CcpA in regulation of the central pathways of carbon catabolism in *Bacillus subtilis***. *J. Bacteriol.* 1999, **181**:6996-7004.
126. Henkin TM: **The role of CcpA transcriptional regulator in carbon metabolism in Bacillus subtilis**. *FEMS Microbiol. Lett.* 1996, **135**:9-15.
127. Titgemeyer F, Hillen W: **Global control of sugar metabolism: a gram-positive solution**. *Antonie Van Leeuwenhoek* 2002, **82**:59-71.
128. Lemos JA, Luzardo Y, Burne RA: **Physiologic effects of forced down-regulation of dnaK and groEL expression in *Streptococcus mutans***. *J. Bacteriol.* 2007, **189**:1582-1588.
129. Wen ZT, Suntharaligham P, Cvitkovitch DG, Burne RA: **Trigger factor in *Streptococcus mutans* is involved in stress tolerance, competence development, and biofilm formation**. *Infect. Immun.* 2005, **73**:219-225.

130. Ahn S, Lemos JAC, Burne RA: **Role of HtrA in growth and competence of *Streptococcus mutans* UA159**. *J. Bacteriol.* 2005, **187**:3028-3038.
131. Quivey RG, Faustoferri RC, Clancy KA, Marquis RE: **Acid adaptation in *Streptococcus mutans* UA159 alleviates sensitization to environmental stress due to RecA deficiency**. *FEMS Microbiol. Lett.* 1995, **126**:257-261.
132. Li Y, Tian X: **Quorum sensing and bacterial social interactions in biofilms**. *Sensors (Basel)* 2012, **12**:2519-2538.
133. Irie Y, Parsek MR: **Quorum sensing and microbial biofilms**. *Curr. Top. Microbiol. Immunol.* 2008, **322**:67-84.
134. Li YH, Lau PC, Lee JH, Ellen RP, Cvitkovitch DG: **Natural genetic transformation of *Streptococcus mutans* growing in biofilms**. *J. Bacteriol.* 2001, **183**:897-908.
135. Li YH, Hanna MN, Svensäter G, Ellen RP, Cvitkovitch DG: **Cell density modulates acid adaptation in *Streptococcus mutans*: implications for survival in biofilms**. *J. Bacteriol.* 2001, **183**:6875-6884.
136. Kreth J, Merritt J, Zhu L, Shi W, Qi F: **Cell density- and ComE-dependent expression of a group of mutacin and mutacin-like genes in *Streptococcus mutans***. *FEMS Microbiol. Lett.* 2006, **265**:11-17.
137. Petersen FC, Pecharki D, Scheie AA: **Biofilm mode of growth of *Streptococcus intermedius* favored by a competence-stimulating signaling peptide**. *J. Bacteriol.* 2004, **186**:6327-6331.
138. Seshasayee ASN, Bertone P, Fraser GM, Luscombe NM: **Transcriptional regulatory networks in bacteria: from input signals to output responses**. *Curr. Opin. Microbiol.* 2006, **9**:511-519.
139. Hoebeke M, Chiapello H, Noirot P, Bessières P: **SPiD: a subtilis protein interaction database**. *Bioinformatics* 2001, **17**:1209-1212.
140. Noirot P, Noirot-Gros M: **Protein interaction networks in bacteria**. *Curr. Opin. Microbiol.* 2004, **7**:505-512.
141. Marchadier E, Carballido-López R, Brinster S, Fabret C, Mervelet P, Bessières P, Noirot-Gros M, Fromion V, Noirot P: **An expanded protein-protein interaction network in *Bacillus subtilis* reveals a group of hubs: Exploration by an integrative approach**. *Proteomics* 2011, **11**:2981-2991.
142. Francke C, Siezen RJ, Teusink B: **Reconstructing the metabolic network of a bacterium from its genome**. *Trends Microbiol.* 2005, **13**:550-558.
143. Sun MGF, Kim PM: **Evolution of biological interaction networks: from models to real data**. *Genome Biol.* 2011, **12**:235.
144. Winter G, Krömer JO: **Fluxomics - connecting 'omics analysis and phenotypes**. *Environ. Microbiol.* 2013, **15**:1901-1916.
145. Hathout Y: **Approaches to the study of the cell secretome**. *Expert Rev. Proteomics* 2007, **4**:239-248.
146. Bhatt S, Romeo T, Kalman D: **Honing the message: post-transcriptional and post-translational control in attaching and effacing pathogens**. *Trends Microbiol.* 2011, **19**:217-224.
147. Djordjevic M: **Efficient transcription initiation in bacteria: an interplay of protein-DNA interaction parameters**. *Integr. Biol. (Camb).* 2013, **5**:796-806.
148. Adamson DN, Lim HN: **Rapid and robust signaling in the CsrA cascade via RNA-protein interactions and feedback regulation**. *Proc. Natl. Acad. Sci. U.S.A.* 2013, **110**:13120-13125.
149. Wade JT, Struhl K, Busby SJW, Grainger DC: **Genomic analysis of protein-DNA interactions in bacteria: insights into transcription and chromosome organization**. *Mol. Microbiol.* 2007, **65**:21-26.
150. Sahr T, Buchrieser C: **Co-immunoprecipitation: protein-RNA and protein-DNA interaction**. *Methods Mol. Biol.* 2013, **954**:583-593.
151. Caldelari I, Chao Y, Romby P, Vogel J: **RNA-mediated regulation in pathogenic bacteria**. *Cold Spring Harb. Perspect. Med.* 2013, **3**:a010298.
152. Control of Gene expression: *http://www.biologyreference.com/Ce-Co/Control-of-Gene-Expression.html#b*.
153. Browning DF, Busby SJ: **The regulation of bacterial transcription initiation**. *Nat. Rev. Microbiol.* 2004, **2**:57-65.

154. Hellman LM, Fried MG: **Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions**. *Nat. Protoc.* 2007, **2**:1849-1861.
155. Gaudreault M, Gingras M, Lessard M, Leclerc S, Guérin SL: **Electrophoretic mobility shift assays for the analysis of DNA-protein interactions**. *Methods Mol. Biol.* 2009, **543**:15-35.
156. Molloy PL: **Electrophoretic mobility shift assays**. *Methods Mol. Biol.* 2000, **130**:235-246.
157. Galas DJ, Schmitz A: **DNAse footprinting: a simple method for the detection of protein-DNA binding specificity**. *Nucleic Acids Res.* 1978, **5**:3157-3170.
158. Massie CE, Mills IG: **Mapping protein-DNA interactions using ChIP-sequencing**. *Methods Mol. Biol.* 2012, **809**:157-173.
159. Ho JWK, Bishop E, Karchenko PV, Nègre N, White KP, Park PJ: **ChIP-chip versus ChIP-seq: lessons for experimental design and data analysis**. *BMC Genomics* 2011, **12**:134.
160. Winkler WC, Breaker RR: **Genetic control by metabolite-binding riboswitches**. *Chembiochem* 2003, **4**:1024-1032.
161. Winkler WC: **Riboswitches and the role of noncoding RNAs in bacterial metabolic control**. *Curr. Opin. Chem. Biol.* 2005, **9**:594-602.
162. Winkler WC, Breaker RR: **Regulation of bacterial gene expression by riboswitches**. *Annu. Rev. Microbiol.* 2005, **59**:487-517.
163. Dambach MD, Winkler WC: **Expanding roles for metabolite-sensing regulatory RNAs**. *Curr. Opin. Microbiol.* 2009, **12**:161-169.
164. Yakhnin H, Zhang H, Yakhnin AV, Babitzke P: **The trp RNA-binding attenuation protein of *Bacillus subtilis* regulates translation of the tryptophan transport gene trpP (yhaG) by blocking ribosome binding**. *J. Bacteriol.* 2004, **186**:278-286.
165. Winkler WC, Cohen-Chalamish S, Breaker RR: **An mRNA structure that controls gene expression by binding FMN**. *Proc. Natl. Acad. Sci. U.S.A.* 2002, **99**:15908-15913.
166. Kim JN, Breaker RR: **Purine sensing by riboswitches**. *Biol. Cell* 2008, **100**:1-11.
167. Kubodera T, Watanabe M, Yoshiuchi K, Yamashita N, Nishimura A, Nakai S, Gomi K, Hanamoto H: **Thiamine-regulated gene expression of *Aspergillus oryzae* thiA requires splicing of the intron containing a riboswitch-like domain in the 5'-UTR**. *FEBS Lett.* 2003, **555**:516-520.
168. Lu Y, Switzer RL: **Evidence that the *Bacillus subtilis* pyrimidine regulatory protein PyrR acts by binding to pyr mRNA at three sites in vivo**. *J. Bacteriol.* 1996, **178**:5806-5809.
169. Remy I, Michnick SW: **Mapping biochemical networks with protein-fragment complementation assays**. *Methods Mol. Biol.* 2004, **261**:411-426.
170. Michnick SW: **Protein fragment complementation strategies for biochemical network mapping**. *Curr. Opin. Biotechnol.* 2003, **14**:610-617.
171. Kim ED, Sabharwal A, Vetta AR, Blanchette M: **Predicting direct protein interactions from affinity purification mass spectrometry data**. *Algorithms Mol. Biol.* 2010, **5**:34.
172. Rohila JS, Chen M, Chen S, Chen J, Cerny R, Dardick C, Canlas P, Xu X, Gribskov M, Kanrar S, Zhu J, Ronald P, Fromm ME: **Protein-protein interactions of tandem affinity purification-tagged protein kinases in rice**. *Plant J.* 2006, **46**:1-13.
173. Rohila JS, Chen M, Chen S, Chen J, Cerny RL, Dardick C, Canlas P, Fujii H, Gribskov M, Kanrar S, Knoflicek L, Stevenson B, Xie M, Xu X, Zheng X, Zhu J, Ronald P, Fromm ME: **Protein-protein interactions of tandem affinity purified protein kinases from rice**. *PLoS ONE* 2009, **4**:e6685.
174. Kenworthy AK: **Imaging protein-protein interactions using fluorescence resonance energy transfer microscopy**. *Methods* 2001, **24**:289-296.
175. Rüdiger AH, Rüdiger M, Carl UD, Chakraborty T, Roepstorff P, Wehland J: **Affinity mass spectrometry-based approaches for the analysis of protein-protein interaction and complex mixtures of peptide-ligands**. *Anal. Biochem.* 1999, **275**:162-170.
176. Zhao Y, Muir TW, Kent SB, Tischer E, Scardina JM, Chait BT: **Mapping protein-protein interactions by affinity-directed mass spectrometry**. *Proc. Natl. Acad. Sci. U.S.A.* 1996, **93**:4020-4024.
177. Griffiths AE, Wang W, Hagen FK, Fay PJ: **Use of affinity-directed liquid chromatography-mass spectrometry to map the epitopes of a factor VIII inhibitor antibody fraction**. *J. Thromb. Haemost.* 2011, **9**:1534-1540.

178. Hall DA, Ptacek J, Snyder M: **Protein microarray technology**. *Mech. Ageing Dev.* 2007, **128**:161-167.
179. Wienken CJ, Baaske P, Rothbauer U, Braun D, Duhr S: **Protein-binding assays in biological liquids using microscale thermophoresis**. *Nat Commun* 2010, **1**:100.
180. Bonifacino JS, Dell'Angelica EC, Springer TA: **Immunoprecipitation**. *Curr. Protoc. Neurosci.* 2006, **46**:5.28.1-5.28.14.
181. Guo D, Rajamäki M, Valkonen J: **Protein-protein interactions: the yeast two-hybrid system**. *Methods Mol. Biol.* 2008, **451**:421-439.
182. Legrain P, Wojcik J, Gauthier JM: **Protein--protein interaction maps: a lead towards cellular functions**. *Trends Genet.* 2001, **17**:346-352.
183. Procaccini A, Lunt B, Szurmant H, Hwa T, Weigt M: **Dissecting the specificity of protein-protein interaction in bacterial two-component signaling: orphans and crosstalks**. *PLoS ONE* 2011, **6**:e19729.
184. Weigt M, White RA, Szurmant H, Hoch JA, Hwa T: **Identification of direct residue contacts in protein-protein interaction by message passing**. *Proc. Natl. Acad. Sci. U.S.A.* 2009, **106**:67-72.
185. http://noxclass.bioinf.mpi-inf.mpg.de/help.php.
186. Whitworth DE, Millard A, Hodgson DA, Hawkins PF: **Protein-protein interactions between two-component system transmitter and receiver domains of *Myxococcus xanthus***. *Proteomics* 2008, **8**:1839-1842.
187. Frieden C: **Protein-protein interaction and enzymatic activity**. *Annu. Rev. Biochem.* 1971, **40**:653-696.
188. Bachhawat P, Swapna GVT, Montelione GT, Stock AM: **Mechanism of activation for transcription factor PhoB suggested by different modes of dimerization in the inactive and active states**. *Structure* 2005, **13**:1353-1363.
189. Li X, Gianoulis TA, Yip KY, Gerstein M, Snyder M: **Extensive in vivo metabolite-protein interactions revealed by large-scale systematic analyses**. *Cell* 2010, **143**:639-650.
190. Li X, Snyder M: **Metabolites as global regulators: a new view of protein regulation: systematic investigation of metabolite-protein interactions may help bridge the gap between genome-wide association studies and small molecule screening studies**. *Bioessays* 2011, **33**:485-489.
191. GERHART JC, PARDEE AB: **The enzymology of control by feedback inhibition**. *J. Biol. Chem.* 1962, **237**:891-896.
192. Stahel PF, Smith WR, Bruchis J, Rabb CH: **Peroxisome proliferator-activated receptors: "key" regulators of neuroinflammation after traumatic brain injury**. *PPAR Res.* 2008, **2008**:538141.
193. Gaballa A, MacLellan S, Helmann JD: **Transcription activation by the siderophore sensor Btr is mediated by ligand-dependent stimulation of promoter clearance**. *Nucleic Acids Res.* 2012, **40**:3585-3595.
194. Walsh CT, Garneau-Tsodikova S, Gatto GJ: **Protein posttranslational modifications: the chemistry of proteome diversifications**. *Angew. Chem. Int. Ed. Engl.* 2005, **44**:7342-7372.
195. Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, Doerks T, Stark M, Muller J, Bork P, Jensen LJ, Mering C von: **The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored**. *Nucleic Acids Res.* 2011, **39**:D561-8.
196. Snel B, Bork P, Huynen MA: **The identification of functional modules from the genomic association of genes**. *Proc. Natl. Acad. Sci. U.S.A.* 2002, **99**:5890-5895.
197. Korbel JO, Jensen LJ, Mering C von, Bork P: **Analysis of genomic context: prediction of functional associations from conserved bidirectionally transcribed gene pairs**. *Nat. Biotechnol.* 2004, **22**:911-917.
198. Pellegrini M, Thompson M, Fierro J, Bowers P: **Computational method to assign microbial genes to pathways**. *J. Cell. Biochem. Suppl.* 2001, **37**:106-109.
199. Marcotte EM: **Computational genetics: finding protein function by nonhomology methods**. *Curr. Opin. Struct. Biol.* 2000, **10**:359-365.
200. Dandekar T, Snel B, Huynen M, Bork P: **Conservation of gene order: a fingerprint of proteins that physically interact**. *Trends Biochem. Sci.* 1998, **23**:324-328.

201. Tamames J: **Evolution of gene order conservation in prokaryotes**. *Genome Biol.* 2001, **2**: research0020-research0020.11.
202. Huynen M, Snel B, Lathe W, Bork P: **Predicting protein function by genomic context: quantitative evaluation and qualitative inferences**. *Genome Res.* 2000, **10**:1204-1210.
203. Strong M, Mallick P, Pellegrini M, Thompson MJ, Eisenberg D: **Inference of protein function and protein linkages in *Mycobacterium tuberculosis* based on prokaryotic genome organization: a combined computational approach**. *Genome Biol.* 2003, **4**:R59.
204. Bowers PM, Pellegrini M, Thompson MJ, Fierro J, Yeates TO, Eisenberg D: **Prolinks: a database of protein functional linkages derived from coevolution**. *Genome Biol.* 2004, **5**:R35.
205. Wu J, Kasif S, DeLisi C: **Identification of functional links between genes using phylogenetic profiles**. *Bioinformatics* 2003, **19**:1524-1530.
206. Pellegrini M, Marcotte EM, Thompson MJ, Eisenberg D, Yeates TO: **Assigning protein functions by comparative genome analysis: protein phylogenetic profiles**. *Proc. Natl. Acad. Sci. U.S.A.* 1999, **96**:4285-4288.
207. Huynen MA, Bork P: **Measuring genome evolution**. *Proc. Natl. Acad. Sci. U.S.A.* 1998, **95**:5849-5856.
208. Marcotte CJV, Marcotte EM: **Predicting functional linkages from gene fusions with confidence**. *Appl. Bioinformatics* 2002, **1**:93-100.
209. Enright AJ, Iliopoulos I, Kyrpides NC, Ouzounis CA: **Protein interaction maps for complete genomes based on gene fusion events**. *Nature* 1999, **402**:86-90.
210. Overbeek R, Fonstein M, D'Souza M, Pusch GD, Maltsev N: **The use of gene clusters to infer functional coupling**. *Proc. Natl. Acad. Sci. U.S.A.* 1999, **96**:2896-2901.
211. Overbeek R, Fonstein M, D'Souza M, Pusch GD, Maltsev N: **Use of contiguity on the chromosome to predict functional coupling**. *In Silico Biol. (Gedrukt)* 1999, **1**:93-108.
212. Jensen LJ, Lagarde J, Mering C von, Bork P: **ArrayProspector: a web resource of functional associations inferred from microarray expression data**. *Nucleic Acids Res.* 2004, **32**:W445-8.
213. Gama-Castro S, Salgado H, Peralta-Gil M, Santos-Zavaleta A, Muñiz-Rascado L, Solano-Lira H, Jimenez-Jacinto V, Weiss V, García-Sotelo JS, López-Fuentes A, Porrón-Sotelo L, Alquicira-Hernández S, Medina-Rivera A, Martínez-Flores I, Alquicira-Hernández K, Martínez-Adame R, Bonavides-Martínez C, Miranda-Ríos J, Huerta AM, Mendoza-Vargas A, Collado-Torres L, Taboada B, Vega-Alvarado L, Olvera M, Olvera L, Grande R, Morett E, Collado-Vides J: **RegulonDB version 7.0: transcriptional regulation of *Escherichia coli* K-12 integrated within genetic sensory response units (Gensor Units)**. *Nucleic Acids Res.* 2011, **39**:D98-105.
214. Sierro N, Makita Y, Hoon M de, Nakai K: **DBTBS: a database of transcriptional regulation in *Bacillus subtilis* containing upstream intergenic conservation information**. *Nucleic Acids Res.* 2008, **36**:D93-6.
215. Münch R, Hiller K, Barg H, Heldt D, Linz S, Wingender E, Jahn D: **PRODORIC: prokaryotic database of gene regulation**. *Nucleic Acids Res.* 2003, **31**:266-269.
216. Rodionov DA: **Comparative genomic reconstruction of transcriptional regulatory networks in bacteria**. *Chem. Rev.* 2007, **107**:3467-3497.
217. Faria JP, Overbeek R, Xia F, Rocha M, Rocha I, Henry CS: **Genome-scale bacterial transcriptional regulatory networks: reconstruction and integrated analysis with metabolic models**. *Brief. Bioinformatics* 2013.
218. Fadda A, Fierro AC, Lemmens K, Monsieurs P, Engelen K, Marchal K: **Inferring the transcriptional network of *Bacillus subtilis***. *Mol. Biosyst.* 2009, **5**:1840-1852.
219. Castro-Melchor M, Charaniya S, Karypis G, Takano E, Hu W: **Genome-wide inference of regulatory networks in *Streptomyces coelicolor***. *BMC Genomics* 2010, **11**:578.
220. Bonneau R, Reiss DJ, Shannon P, Facciotti M, Hood L, Baliga NS, Thorsson V: **The Inferelator: an algorithm for learning parsimonious regulatory networks from systems-biology data sets de novo**. *Genome Biol.* 2006, **7**:R36.
221. Fu Y, Jarboe LR, Dickerson JA: **Reconstructing genome-wide regulatory network of E. coli using transcriptome data and predicted transcription factor activities**. *BMC Bioinformatics* 2011, **12**:233.
222. Guelzim N, Bottani S, Bourgine P, Képès F: **Topological and causal structure of the yeast transcriptional regulatory network**. *Nat. Genet.* 2002, **31**:60-63.

223. Altwasser R, Linde J, Buyko E, Hahn U, Guthke R: **Genome-Wide Scale-Free Network Inference for Candida albicans**. *Front. Microbiol.* 2012, **3**:51.
224. Bailly-Bechet M, Braunstein A, Pagnani A, Weigt M, Zecchina R: **Inference of sparse combinatorial-control networks from gene-expression data: a message passing approach**. *BMC Bioinformatics* 2010, **11**:355.
225. Pollack JR: **DNA microarray technology. Introduction**. *Methods Mol. Biol.* 2009, **556**:1-6.
226. Schena M, Shalon D, Davis RW, Brown PO: **Quantitative monitoring of gene expression patterns with a complementary DNA microarray**. *Science* 1995, **270**:467-470.
227. Kogenaru S, Qing Y, Guo Y, Wang N: **RNA-seq and microarray complement each other in transcriptome profiling**. *BMC Genomics* 2012, **13**:629.
228. Wang Z, Gerstein M, Snyder M: **RNA-Seq: a revolutionary tool for transcriptomics**. *Nat. Rev. Genet.* 2009, **10**:57-63.
229. Kazakov AE, Cipriano MJ, Novichkov PS, Minovitsky S, Vinogradov DV, Arkin A, Mironov AA, Gelfand MS, Dubchak I: **RegTransBase--a database of regulatory sequences and interactions in a wide range of prokaryotic genomes**. *Nucleic Acids Res.* 2007, **35**:D407-12.
230. Park PJ: **ChIP-seq: advantages and challenges of a maturing technology**. *Nat. Rev. Genet.* 2009, **10**:669-680.
231. Novichkov PS, Laikova ON, Novichkova ES, Gelfand MS, Arkin AP, Dubchak I, Rodionov DA: **RegPrecise: a database of curated genomic inferences of transcriptional regulatory interactions in prokaryotes**. *Nucleic Acids Res.* 2010, **38**:D111-8.
232. Rodionov DA, Novichkov PS, Stavrovskaya ED, Rodionova IA, Li X, Kazanov MD, Ravcheev DA, Gerasimova AV, Kazakov AE, Kovaleva GY, Permina EA, Laikova ON, Overbeek R, Romine MF, Fredrickson JK, Arkin AP, Dubchak I, Osterman AL, Gelfand MS: **Comparative genomic reconstruction of transcriptional networks controlling central metabolism in the Shewanella genus**. *BMC Genomics* 2011, **12 Suppl 1**:S3.
233. Bustin SA, Benes V, Nolan T, Pfaffl MW: **Quantitative real-time RT-PCR--a perspective**. *J. Mol. Endocrinol.* 2005, **34**:597-601.
234. Liang P, PARDEE AB: **Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction**. *Science* 1992, **257**:967-971.
235. Levsky JM, Singer RH: **Fluorescence in situ hybridization: past, present and future**. *J. Cell. Sci.* 2003, **116**:2833-2838.
236. Josefsen K, Nielsen H: **Northern blotting analysis**. *Methods Mol. Biol.* 2011, **703**:87-105.
237. Fu Y, Xiao W: **Study of transcriptional regulation using a reporter gene assay**. *Methods Mol. Biol.* 2006, **313**:257-264.
238. Iida K, Nishimura I: **Gene expression profiling by DNA microarray technology**. *Crit. Rev. Oral. Biol. Med.* 2002, **13**:35-50.
239. Yamamoto M, Wakatsuki T, Hada A, Ryo A: **Use of serial analysis of gene expression (SAGE) technology**. *J. Immunol. Methods* 2001, **250**:45-66.
240. Ramsköld D, Kavak E, Sandberg R: **How to analyze gene expression using RNA-sequencing data**. *Methods Mol. Biol.* 2012, **802**:259-274.
241. Ng ST, Sanusi Jangi M, Shirley MW, Tomley FM, Wan KL: **Comparative EST analyses provide insights into gene expression in two asexual developmental stages of Eimeria tenella**. *Exp. Parasitol.* 2002, **101**:168-173.
242. Quijada L, Soto M, Requena JM: **Genomic DNA macroarrays as a tool for analysis of gene expression in Leishmania**. *Exp. Parasitol.* 2005, **111**:64-70.
243. Brenner S, Johnson M, Bridgham J, Golda G, Lloyd DH, Johnson D, Luo S, McCurdy S, Foy M, Ewan M, Roth R, George D, Eletr S, Albrecht G, Vermaas E, Williams SR, Moon K, Burcham T, Pallas M, DuBridge RB, Kirchner J, Fearon K, Mao J, Corcoran K: **Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays**. *Nat. Biotechnol.* 2000, **18**:630-634.
244. Obayashi T, Kinoshita K: **Rank of correlation coefficient as a comparable measure for biological significance of gene coexpression**. *DNA Res.* 2009, **16**:249-260.
245. Belcastro V, Di Bernardo D: **Reverse engineering transcriptional gene networks**. *Methods Mol. Biol.* 2014, **1101**:179-196.
246. Hache H, Lehrach H, Herwig R: **Reverse engineering of gene regulatory networks: a comparative study**. *EURASIP J Bioinform. Syst. Biol* 2009:617281.

247. Kim D, Wang X, Yang C, Gao J: **Learning biological network using mutual information and conditional independence**. *BMC Bioinformatics* 2010, **11 Suppl 3**:S9.
248. Magwene PM, Kim J: **Estimating genomic coexpression networks using first-order conditional independence**. *Genome Biol.* 2004, **5**:R100.
249. Hache H, Wierling C, Lehrach H, Herwig R: **GeNGe: systematic generation of gene regulatory networks**. *Bioinformatics* 2009, **25**:1205-1207.
250. Friedman N, Linial M, Nachman I, Pe'er D: **Using Bayesian networks to analyze expression data**. *J. Comput. Biol.* 2000, **7**:601-620.
251. Husmeier D, Werhli AV: **Bayesian integration of biological prior knowledge into the reconstruction of gene regulatory networks with Bayesian networks**. *Comput Syst Bioinformatics Conf.* 2007, **6**:85-95.
252. Werhli AV, Grzegorczyk M, Husmeier D: **Comparative evaluation of reverse engineering gene regulatory networks with relevance networks, graphical gaussian models and bayesian networks**. *Bioinformatics* 2006, **22**:2523-2531.
253. Werhli AV, Husmeier D: **Reconstructing gene regulatory networks with bayesian networks by combining expression data with multiple sources of prior knowledge**. *Stat. Appl. Genet. Mol. Biol.* 2007, **6**:Article15.
254. La Fuente A de, Bing N, Hoeschele I, Mendes P: **Discovery of meaningful associations in genomic data using partial correlation coefficients**. *Bioinformatics* 2004, **20**:3565-3574.
255. Margolin AA, Nemenman I, Basso K, Wiggins C, Stolovitzky G, Dalla Favera R, Califano A: **ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context**. *BMC Bioinformatics* 2006, **7 Suppl 1**:S7.
256. Qian J, Dolled-Filhart M, Lin J, Yu H, Gerstein M: **Beyond synexpression relationships: local clustering of time-shifted and inverted gene expression profiles identifies new, biologically relevant interactions**. *J. Mol. Biol.* 2001, **314**:1053-1066.
257. Kabir M, Noman N, Iba H: **Reverse engineering gene regulatory network from microarray data using linear time-variant model**. *BMC Bioinformatics* 2010, **11 Suppl 1**:S56.
258. Noman N, Iba H: **Reverse engineering genetic networks using evolutionary computation**. *Genome Inform.* 2005, **16**:205-214.
259. Repsilber D, Liljenström H, Andersson SGE: **Reverse engineering of regulatory networks: simulation studies on a genetic algorithm approach for ranking hypotheses**. *BioSystems* 2002, **66**:31-41.
260. Xiong H, Choe Y: **Structural systems identification of genetic regulatory networks**. *Bioinformatics* 2008, **24**:553-560.
261. Tejera E, Bernardes J, Rebelo I: **Co-expression network analysis and genetic algorithms for gene prioritization in preeclampsia**. *BMC Med Genomics* 2013, **6**:51.
262. Villa-Vialaneix N, Liaubet L, Laurent T, Cherel P, Gamot A, SanCristobal M: **The structure of a gene co-expression network reveals biological functions underlying eQTLs**. *PLoS ONE* 2013, **8**:e60045.
263. Zhang J, Lu K, Xiang Y, Islam M, Kotian S, Kais Z, Lee C, Arora M, Liu H, Parvin JD, Huang K: **Weighted frequent gene co-expression network mining to identify genes involved in genome stability**. *PLoS Comput. Biol.* 2012, **8**:e1002656.
264. Willsey AJ, Sanders SJ, Li M, Dong S, Tebbenkamp AT, Muhle RA, Reilly SK, Lin L, Fertuzinhos S, Miller JA, Murtha MT, Bichsel C, Niu W, Cotney J, Ercan-Sencicek AG, Gockley J, Gupta AR, Han W, He X, Hoffman EJ, Klei L, Lei J, Liu W, Liu L, Lu C, Xu X, Zhu Y, Mane SM, Lein ES, Wei L *et al.*: **Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism**. *Cell* 2013, **155**:997-1007.
265. Puniya BL, Kulshreshtha D, Verma SP, Kumar S, Ramachandran S: **Integrated gene co-expression network analysis in the growth phase of *Mycobacterium tuberculosis* reveals new potential drug targets**. *Mol. Biosyst.* 2013, **9**:2798-2815.
266. Filteau M, Pavey SA, St-Cyr J, Bernatchez L: **Gene coexpression networks reveal key drivers of phenotypic divergence in lake whitefish**. *Mol. Biol. Evol.* 2013, **30**:1384-1396.
267. Basso K, Margolin AA, Stolovitzky G, Klein U, Dalla-Favera R, Califano A: **Reverse engineering of regulatory networks in human B cells**. *Nat. Genet.* 2005, **37**:382-390.

268. Licata L, Briganti L, Peluso D, Perfetto L, Iannuccelli M, Galeota E, Sacco F, Palma A, Nardozza AP, Santonico E, Castagnoli L, Cesareni G: **MINT, the molecular interaction database: 2012 update**. *Nucleic Acids Res.* 2012, **40**:D857-61.
269. Kerrien S, Aranda B, Breuza L, Bridge A, Broackes-Carter F, Chen C, Duesbury M, Dumousseau M, Feuermann M, Hinz U, Jandrasits C, Jimenez RC, Khadake J, Mahadevan U, Masson P, Pedruzzi I, Pfeiffenberger E, Porras P, Raghunath A, Roechert B, Orchard S, Hermjakob H: **The IntAct molecular interaction database in 2012**. *Nucleic Acids Res.* 2012, **40**:D841-6.
270. Xenarios I, Rice DW, Salwinski L, Baron MK, Marcotte EM, Eisenberg D: **DIP: the database of interacting proteins**. *Nucleic Acids Res.* 2000, **28**:289-291.
271. Stark C, Breitkreutz B, Reguly T, Boucher L, Breitkreutz A, Tyers M: **BioGRID: a general repository for interaction datasets**. *Nucleic Acids Res.* 2006, **34**:D535-9.
272. Bader GD, Betel D, Hogue CWV: **BIND: the Biomolecular Interaction Network Database**. *Nucleic Acids Res.* 2003, **31**:248-250.
273. Pagel P, Kovac S, Oesterheld M, Brauner B, Dunger-Kaltenbach I, Frishman G, Montrone C, Mark P, Stümpflen V, Mewes H, Ruepp A, Frishman D: **The MIPS mammalian protein-protein interaction database**. *Bioinformatics* 2005, **21**:832-834.
274. Mewes HW, Dietmann S, Frishman D, Gregory R, Mannhaupt G, Mayer KFX, Münsterkötter M, Ruepp A, Spannagl M, Stümpflen V, Rattei T: **MIPS: analysis and annotation of genome information in 2007**. *Nucleic Acids Res.* 2008, **36**:D196-201.
275. Tetko IV, Brauner B, Dunger-Kaltenbach I, Frishman G, Montrone C, Fobo G, Ruepp A, Antonov AV, Surmeli D, Mewes H: **MIPS bacterial genomes functional annotation benchmark dataset**. *Bioinformatics* 2005, **21**:2520-2521.
276. Frolkis A, Knox C, Lim E, Jewison T, Law V, Hau DD, Liu P, Gautam B, Ly S, Guo C an, Xia J, Liang Y, Shrivastava S, Wishart DS: **SMPDB: The Small Molecule Pathway Database**. *Nucleic Acids Res.* 2010, **38**:D480-7.
277. Joshi-Tope G, Gillespie M, Vastrik I, D'Eustachio P, Schmidt E, Bono B de, Jassal B, Gopinath GR, Wu GR, Matthews L, Lewis S, Birney E, Stein L: **Reactome: a knowledgebase of biological pathways**. *Nucleic Acids Res.* 2005, **33**:D428-32.
278. Salek RM, Haug K, Conesa P, Hastings J, Williams M, Mahendraker T, Maguire E, González-Beltrán AN, Rocca-Serra P, Sansone S, Steinbeck C: **The MetaboLights repository: curation challenges in metabolomics**. *Database (Oxford)* 2013, **2013**:bat029.
279. Hanzlik RP, Koen YM, Theertham B, Dong Y, Fang J: **The reactive metabolite target protein database (TPDB)--a web-accessible resource**. *BMC Bioinformatics* 2007, **8**:95.
280. Bauer RA, Günther S, Jansen D, Heeger C, Thaben PF, Preissner R: **SuperSite: dictionary of metabolite and drug binding sites in proteins**. *Nucleic Acids Res.* 2009, **37**:D195-200.
281. Winnenburg R, Baldwin TK, Urban M, Rawlings C, Köhler J, Hammond-Kosack KE: **PHI-base: a new database for pathogen host interactions**. *Nucleic Acids Res.* 2006, **34**:D459-64.
282. Kamburov A, Pentchev K, Galicka H, Wierling C, Lehrach H, Herwig R: **ConsensusPathDB: toward a more complete picture of cell biology**. *Nucleic Acids Res.* 2011, **39**:D712-7.
283. Ma S, Shah S, Bohnert HJ, Snyder M, Dinesh-Kumar SP: **Incorporating motif analysis into gene co-expression networks reveals novel modular expression pattern and new signaling pathways**. *PLoS Genet.* 2013, **9**:e1003840.
284. Clements M, van Someren EP, Knijnenburg TA, Reinders MJT: **Integration of known transcription factor binding site information and gene expression data to advance from co-expression to co-regulation**. *Genomics Proteomics Bioinformatics* 2007, **5**:86-101.
285. Kharchenko P, Chen L, Freund Y, Vitkup D, Church GM: **Identifying metabolic enzymes with multiple types of association evidence**. *BMC Bioinformatics* 2006, **7**:177.
286. Leyn SA, Kazanov MD, Sernova NV, Ermakova EO, Novichkov PS, Rodionov DA: **Genomic Reconstruction of Transcriptional Regulatory Network in *Bacillus subtilis***. *J. Bacteriol.* 2013.
287. Cornish JP, Matthews F, Thomas JR, Erill I: **Inference of self-regulated transcriptional networks by comparative genomics**. *Evol. Bioinform. Online* 2012, **8**:449-461.
288. Thomas-Chollier M, Defrance M, Medina-Rivera A, Sand O, Herrmann C, Thieffry D, van Helden J: **RSAT 2011. regulatory sequence analysis tools**. *Nucleic acids research* 2011, **39**:W86-91.

289. Janky R, van Helden J, Babu MM: **Investigating transcriptional regulation: from analysis of complex networks to discovery of cis-regulatory elements**. *Methods* 2009, **48**:277-286.
290. Defrance M, Janky R, Sand O, van Helden J: **Using RSAT oligo-analysis and dyad-analysis tools to discover regulatory signals in nucleic sequences**. *Nat. Protoc.* 2008, **3**:1589-1603.
291. Janky R, van Helden J: **Evaluation of phylogenetic footprint discovery for predicting bacterial cis-regulatory elements and revealing their evolution**. *BMC Bioinformatics* 2008, **9**:37.
292. Sand O, van Helden J: **Discovery of motifs in promoters of coregulated genes**. *Methods Mol. Biol.* 2007, **395**:329-348.
293. Janky R, van Helden J: **Discovery of conserved motifs in promoters of orthologous genes in prokaryotes**. *Methods Mol. Biol.* 2007, **395**:293-308.
294. Wasserman WW, Sandelin A: **Applied bioinformatics for the identification of regulatory elements**. *Nat. Rev. Genet.* 2004, **5**:276-287.
295. van Helden J: **Regulatory sequence analysis tools**. *Nucleic Acids Res.* 2003, **31**:3593-3596.
296. Turatsinze J, Thomas-Chollier M, Defrance M, van Helden J: **Using RSAT to scan genome sequences for transcription factor binding sites and cis-regulatory modules**. *Nat. Protoc.* 2008, **3**:1578-1588.
297. Sadler JR, Waterman MS, Smith TF: **Regulatory pattern identification in nucleic acid sequences**. *Nucleic Acids Res.* 1983, **11**:2221-2231.
298. Pevzner PA, Borodovsky MYu, Mironov AA: **Linguistics of nucleotide sequences. I: The significance of deviations from mean statistical characteristics and prediction of the frequencies of occurrence of words**. *J. Biomol. Struct. Dyn.* 1989, **6**:1013-1026.
299. Pevzner PA, Borodovsky MYu, Mironov AA: **Linguistics of nucleotide sequences. II: Stationary words in genetic texts and the zonal structure of DNA**. *J. Biomol. Struct. Dyn.* 1989, **6**:1027-1038.
300. Hertz GZ, Stormo GD: **Identifying DNA and protein patterns with statistically significant alignments of multiple sequences**. *Bioinformatics* 1999, **15**:563-577.
301. Schneider TD: **Reading of DNA sequence logos: prediction of major groove binding by information theory**. *Meth. Enzymol.* 1996, **274**:445-455.
302. Schneider TD, Stephens RM: **Sequence logos: a new way to display consensus sequences**. *Nucleic Acids Res.* 1990, **18**:6097-6100.
303. Shultzaberger RK, Schneider TD: **Using sequence logos and information analysis of Lrp DNA binding sites to investigate discrepancies between natural selection and SELEX**. *Nucleic Acids Res.* 1999, **27**:882-887.
304. Schneider TD, Stormo GD, Gold L, Ehrenfeucht A: **Information content of binding sites on nucleotide sequences**. *J. Mol. Biol.* 1986, **188**:415-431.
305. Schneider TD, Stormo GD, Yarus MA, Gold L: **Delila system tools**. *Nucleic Acids Res.* 1984, **12**:129-140.
306. Lawrence CE, Altschul SF, Boguski MS, Liu JS, Neuwald AF, Wootton JC: **Detecting subtle sequence signals: a Gibbs sampling strategy for multiple alignment**. *Science* 1993, **262**:208-214.
307. Stormo GD, Hartzell GW: **Identifying protein-binding sites from unaligned DNA fragments**. *Proc. Natl. Acad. Sci. U.S.A.* 1989, **86**:1183-1187.
308. Thijs G, Lescot M, Marchal K, Rombauts S, Moor B de, Rouzé P, Moreau Y: **A higher-order background model improves the detection of promoter regulatory elements by Gibbs sampling**. *Bioinformatics* 2001, **17**:1113-1122.
309. van Helden J, André B, Collado-Vides J: **Extracting regulatory sites from the upstream region of yeast genes by computational analysis of oligonucleotide frequencies**. *J. Mol. Biol.* 1998, **281**:827-842.
310. Hermsen R, Tans S, Wolde PR ten: **Transcriptional regulation by competing transcription factor modules**. *PLoS Comput. Biol.* 2006, **2**:e164.
311. Buchler NE, Gerland U, Hwa T: **On schemes of combinatorial transcription logic**. *Proc. Natl. Acad. Sci. U.S.A.* 2003, **100**:5136-5141.
312. Karp PD, Paley SM, Krummenacker M, Latendresse M, Dale JM, Lee TJ, Kaipa P, Gilham F, Spaulding A, Popescu L, Altman T, Paulsen I, Keseler IM, Caspi R: **Pathway Tools version**

13.0. integrated software for pathway/genome informatics and systems biology. *Briefings in bioinformatics* 2010, **11**:40-79.
313. Ideker T, Krogan NJ: **Differential network biology**. *Mol. Syst. Biol.* 2012, **8**:565.
314. Valcárcel B, Würtz P, Seich al Basatena N, Tukiainen T, Kangas AJ, Soininen P, Järvelin M, Ala-Korpela M, Ebbels TM, Iorio M de: **A differential network approach to exploring differences between biological states: an application to prediabetes**. *PLoS ONE* 2011, **6**:e24702.
315. Li L, Stoeckert CJ, Roos DS: **OrthoMCL: identification of ortholog groups for eukaryotic genomes**. *Genome Res.* 2003, **13**:2178-2189.
316. Medina-Rivera A, Abreu-Goodger C, Thomas-Chollier M, Salgado H, Collado-Vides J, van Helden J: **Theoretical and empirical quality assessment of transcription factor-binding motifs**. *Nucleic Acids Res.* 2011, **39**:808-824.
317. Thomas-Chollier M, Sand O, Turatsinze J, Janky R, Defrance M, Vervisch E, Brohée S, van Helden J: **RSAT: regulatory sequence analysis tools**. *Nucleic Acids Res.* 2008, **36**:W119-27.
318. Gillespie JJ, Wattam AR, Cammer SA, Gabbard JL, Shukla MP, Dalay O, Driscoll T, Hix D, Mane SP, Mao C, Nordberg EK, Scott M, Schulman JR, Snyder EE, Sullivan DE, Wang C, Warren A, Williams KP, Xue T, Yoo HS, Zhang C, Zhang Y, Will R, Kenyon RW, Sobral BW: **PATRIC: the comprehensive bacterial bioinformatics resource with a focus on human pathogenic species**. *Infect. Immun.* 2011, **79**:4286-4298.
319. Killcoyne S, Carter GW, Smith J, Boyle J: **Cytoscape: a community-based framework for network modeling**. *Methods Mol. Biol.* 2009, **563**:219-239.
320. Kashtan N, Itzkovitz S, Milo R, Alon U: **Efficient sampling algorithm for estimating subgraph concentrations and detecting network motifs**. *Bioinformatics* 2004, **20**:1746-1758.
321. Battke F, Symons S, Nieselt K: **Mayday--integrative analytics for expression data**. *BMC Bioinformatics* 2010, **11**:121.
322. Dunny GM, Lee LN, LeBlanc DJ: **Improved electroporation and cloning vector system for gram-positive bacteria**. *Appl. Environ. Microbiol.* 1991, **57**:1194-1201.
323. Lemme A, Gröbe L, Reck M, Tomasch J, Wagner-Döbler I: **Subpopulation-specific transcriptome analysis of competence-stimulating-peptide-induced *Streptococcus mutans***. *J. Bacteriol.* 2011, **193**:1863-1877.
324. Perry JA, Lévesque CM, Suntharaligam P, Mair RW, Bu M, Cline RT, Peterson SN, Cvitkovitch DG: **Involvement of *Streptococcus mutans* regulator RR11 in oxidative stress response during biofilm growth and in the development of genetic competence**. *Lett. Appl. Microbiol.* 2008, **47**:439-444.
325. Senadheera D, Krastel K, Mair R, Persadmehr A, Abranches J, Burne RA, Cvitkovitch DG: **Inactivation of VicK affects acid production and acid survival of *Streptococcus mutans***. *J. Bacteriol.* 2009, **191**:6415-6424.
326. Isalan M, Lemerle C, Michalodimitrakis K, Horn C, Beltrao P, Raineri E, Garriga-Canut M, Serrano L: **Evolvability and hierarchy in rewired bacterial gene networks**. *Nature* 2008, **452**:840-845.
327. Ma H, Buer J, Zeng A: **Hierarchical structure and modules in the *Escherichia coli* transcriptional regulatory network revealed by a new top-down approach**. *BMC Bioinformatics* 2004, **5**:199.
328. Galán-Vásquez E, Luna B, Martínez-Antonio A: **The Regulatory Network of *Pseudomonas aeruginosa***. *Microb. Inform. Exp.* 2011, **1**:3.
329. Dufour YS, Kiley PJ, Donohue TJ: **Reconstruction of the core and extended regulons of global transcription factors**. *PLoS Genet.* 2010, **6**:e1001027.
330. Chaturongakul S, Raengpradub S, Palmer ME, Bergholz TM, Orsi RH, Hu Y, Ollinger J, Wiedmann M, Boor KJ: **Transcriptomic and phenotypic analyses identify coregulated, overlapping regulons among PrfA, CtsR, HrcA, and the alternative sigma factors sigmaB, sigmaC, sigmaH, and sigmaL in *Listeria monocytogenes***. *Appl. Environ. Microbiol.* 2011, **77**:187-200.
331. Konagurthu AS, Lesk AM: **Single and multiple input modules in regulatory networks**. *Proteins* 2008, **73**:320-324.
332. Mangan S, Alon U: **Structure and function of the feed-forward loop network motif**. *Proc. Natl. Acad. Sci. U.S.A.* 2003, **100**:11980-11985.

333. Kalir S, McClure J, Pabbaraju K, Southward C, Ronen M, Leibler S, Surette MG, Alon U: **Ordering genes in a flagella pathway by analysis of expression kinetics from living bacteria**. *Science* 2001, **292**:2080-2083.
334. Kashtan N, Itzkovitz S, Milo R, Alon U: **Topological generalizations of network motifs**. *Phys. Rev. E. Stat. Nonlin. Soft. Matter. Phys.* 2004, **70**:31909.
335. Ma H, Kumar B, Ditges U, Gunzer F, Buer J, Zeng A: **An extended transcriptional regulatory network of *Escherichia coli* and analysis of its hierarchical structure and network motifs**. *Nucleic Acids Res.* 2004, **32**:6643-6649.
336. Zeng L, Das S, Burne RA: **Genetic analysis of the functions and interactions of components of the LevQRST signal transduction complex of *Streptococcus mutans***. *PLoS ONE* 2011, **6**:e17335.
337. Webb AJ, Homer KA, Hosie AHF: **Two closely related ABC transporters in *Streptococcus mutans* are involved in disaccharide and/or oligosaccharide uptake**. *J. Bacteriol.* 2008, **190**:168-178.
338. Touger-Decker R, van Loveren C: **Sugars and dental caries**. *Am. J. Clin. Nutr.* 2003, **78**:881S-892S.
339. van Houte J, Upeslacis VN, Jordan HV, Skobe Z, Green DB: **Role of sucrose in colonization of *Streptococcus mutans* in conventional Sprague-Dawley rats**. *J. Dent. Res.* 1976, **55**:202-215.
340. Cury JA, Francisco SB, Del Bel Cury AA, Tabchoury CP: **In situ study of sucrose exposure, mutans streptococci in dental plaque and dental caries**. *Braz. Dent. J* 2001, **12**:101-104.
341. Nobre dos Santos M, Melo dos Santos L, Francisco SB, Cury JA: **Relationship among dental plaque composition, daily sugar exposure and caries in the primary dentition**. *Caries Res.* 2002, **36**:347-352.
342. Sadykov MR, Mattes TA, Luong TT, Zhu Y, Day SR, Sifri CD, Lee CY, Somerville GA: **Tricarboxylic acid cycle-dependent synthesis of *Staphylococcus aureus* Type 5 and 8 capsular polysaccharides**. *J. Bacteriol.* 2010, **192**:1459-1462.
343. Howden BP, McEvoy CRE, Allen DL, Chua K, Gao W, Harrison PF, Bell J, Coombs G, Bennett-Wood V, Porter JL, Robins-Browne R, Davies JK, Seemann T, Stinear TP: **Evolution of multidrug resistance during *Staphylococcus aureus* infection involves mutation of the essential two component regulator WalKR**. *PLoS Pathog.* 2011, **7**:e1002359.
344. Cvitkovitch DG, Gutierrez JA, Bleiweis AS: **Role of the citrate pathway in glutamate biosynthesis by *Streptococcus mutans***. *J. Bacteriol.* 1997, **179**:650-655.
345. Yu H, Yao Y, Liu Y, Jiao R, Jiang W, Zhao G: **A complex role of *Amycolatopsis mediterranei* GlnR in nitrogen metabolism and related antibiotics production**. *Arch. Microbiol.* 2007, **188**:89-96.
346. Chattoraj P, Mohapatra SS, Rao JLUM, Biswas I: **Regulation of transcription by SMU.1349, a TetR family regulator, in *Streptococcus mutans***. *J. Bacteriol.* 2011, **193**:6605-6613.
347. Krastel K, Senadheera DB, Mair R, Downey JS, Goodman SD, Cvitkovitch DG: **Characterization of a glutamate transporter operon, glnQHMP, in *Streptococcus mutans* and its role in acid tolerance**. *J. Bacteriol.* 2010, **192**:984-993.
348. Matsui R, Cvitkovitch D: **Acid tolerance mechanisms utilized by *Streptococcus mutans***. *Future Microbiol.* 2010, **5**:403-417.
349. Svensäter G, Larsson UB, Greif EC, Cvitkovitch DG, Hamilton IR: **Acid tolerance response and survival by oral bacteria**. *Oral Microbiol. Immunol.* 1997, **12**:266-273.
350. Lévesque CM, Voronejskaia E, Huang YC, Mair RW, Ellen RP, Cvitkovitch DG: **Involvement of sortase anchoring of cell wall proteins in biofilm formation by *Streptococcus mutans***. *Infect. Immun.* 2005, **73**:3773-3777.
351. Senadheera DB, Cordova M, Ayala EA, Chávez Paz LE de, Singh K, Downey JS, Svensäter G, Goodman SD, Cvitkovitch DG: **Regulation of bacteriocin production and cell death by the VicRK signaling system in *Streptococcus mutans***. *J. Bacteriol.* 2012, **194**:1307-1316.
352. Sperandio B, Gautier C, Pons N, Ehrlich DS, Renault P, Guedon E: **Three paralogous LysR-type transcriptional regulators control sulfur amino acid supply in *Streptococcus mutans***. *J. Bacteriol.* 2010, **192**:3464-3473.
353. Fleuchot B, Gitton C, Guillot A, Vidic J, Nicolas P, Besset C, Fontaine L, Hols P, Leblond-Bourget N, Monnet V, Gardan R: **Rgg proteins associated with internalized small**

hydrophobic peptides: a new quorum-sensing mechanism in streptococci. *Mol. Microbiol.* 2011, **80**:1102-1119.
354. Filipe SR, Severina E, Tomasz A: **Functional analysis of *Streptococcus pneumoniae* MurM reveals the region responsible for its specificity in the synthesis of branched cell wall peptides**. *J. Biol. Chem.* 2001, **276**:39618-39628.
355. Filipe SR, Severina E, Tomasz A: **The role of murMN operon in penicillin resistance and antibiotic tolerance of *Streptococcus pneumoniae***. *Microb. Drug Resist.* 2001, **7**:303-316.
356. Filipe SR, Severina E, Tomasz A: **The murMN operon: a functional link between antibiotic resistance and antibiotic tolerance in *Streptococcus pneumoniae***. *Proc. Natl. Acad. Sci. U.S.A.* 2002, **99**:1550-1555.
357. Filipe SR, Pinho MG, Tomasz A: **Characterization of the murMN operon involved in the synthesis of branched peptidoglycan peptides in *Streptococcus pneumoniae***. *J. Biol. Chem.* 2000, **275**:27768-27774.
358. Typas A, Banzhaf M, Gross CA, Vollmer W: **From the regulation of peptidoglycan synthesis to bacterial growth and morphology**. *Nat. Rev. Microbiol.* 2012, **10**:123-136.
359. Jonge BLM de, Gage D, Xu N: **The carboxyl terminus of peptidoglycan stem peptides is a determinant for methicillin resistance in *Staphylococcus aureus***. *Antimicrob. Agents Chemother.* 2002, **46**:3151-3155.
360. Garcia-Bustos JF, Chait BT, Tomasz A: **Altered peptidoglycan structure in a pneumococcal transformant resistant to penicillin**. *J. Bacteriol.* 1988, **170**:2143-2147.
361. Garcia-Bustos J, Tomasz A: **A biological price of antibiotic resistance: major changes in the peptidoglycan structure of penicillin-resistant pneumococci**. *Proc. Natl. Acad. Sci. U.S.A.* 1990, **87**:5415-5419.
362. Simelyte E, Rimpiläinen M, Zhang X, Toivanen P: **Role of peptidoglycan subtypes in the pathogenesis of bacterial cell wall arthritis**. *Ann. Rheum. Dis.* 2003, **62**:976-982.
363. Kajfasz JK, Rivera-Ramos I, Abranches J, Martinez AR, Rosalen PL, Derr AM, Quivey RG, Lemos JA: **Two Spx proteins modulate stress tolerance, survival, and virulence in *Streptococcus mutans***. *J. Bacteriol.* 2010, **192**:2546-2556.
364. Sperandio B, Gautier C, McGovern S, Ehrlich DS, Renault P, Martin-Verstraete I, Guédon E: **Control of methionine synthesis and uptake by MetR and homocysteine in *Streptococcus mutans***. *J. Bacteriol.* 2007, **189**:7032-7044.
365. Catt DM, Gregory RL: ***Streptococcus mutans* murein hydrolase**. *J. Bacteriol.* 2005, **187**:7863-7865.
366. Dufour D, Lévesque CM: **Cell death of *Streptococcus mutans* induced by a quorum-sensing peptide occurs via a conserved streptococcal autolysin**. *J. Bacteriol.* 2013, **195**:105-114.
367. Dufour D, Cordova M, Cvitkovitch DG, Lévesque CM: **Regulation of the competence pathway as a novel role associated with a streptococcal bacteriocin**. *J. Bacteriol.* 2011, **193**:6552-6559.
368. Perry JA, Jones MB, Peterson SN, Cvitkovitch DG, Lévesque CM: **Peptide alarmone signalling triggers an auto-active bacteriocin necessary for genetic competence**. *Mol. Microbiol.* 2009, **72**:905-917.
369. Wu H, Su Z, Mao F, Olman V, Xu Y: **Prediction of functional modules based on comparative genome analysis and Gene Ontology application**. *Nucleic Acids Res.* 2005, **33**:2822-2837.
370. Mavromatis K, Chu K, Ivanova N, Hooper SD, Markowitz VM, Kyrpides NC: **Gene context analysis in the Integrated Microbial Genomes (IMG) data management system**. *PLoS ONE* 2009, **4**:e7979.
371. Allen JP, Neely MN: **CpsY influences *Streptococcus iniae* cell wall adaptations important for neutrophil intracellular survival**. *Infect. Immun.* 2012, **80**:1707-1715.
372. Federle MJ, Morrison DA: **One if by land, two if by sea: signalling to the ranks with CSP and XIP**. *Mol. Microbiol.* 2012, **86**:241-245.
373. Mashburn-Warren L, Morrison DA, Federle MJ: **A novel double-tryptophan peptide pheromone controls competence in Streptococcus spp. via an Rgg regulator**. *Mol. Microbiol.* 2010, **78**:589-606.
374. Beilharz K, Nováková L, Fadda D, Branny P, Massidda O, Veening J: **Control of cell division in Streptococcus pneumoniae by the conserved Ser/Thr protein kinase StkP**. *Proc. Natl. Acad. Sci. U.S.A.* 2012, **109**:E905-13.

375. Güell M, Yus E, Lluch-Senar M, Serrano L: **Bacterial transcriptomics: what is beyond the RNA horiz-ome?** *Nat. Rev. Microbiol.* 2011, **9**:658-669.
376. Koyanagi S, Lévesque CM: **Characterization of a *Streptococcus mutans* intergenic region containing a small toxic peptide and its cis-encoded antisense small RNA antitoxin**. *PLoS ONE* 2013, **8**:e54291.
377. Hasona A, Crowley PJ, Levesque CM, Mair RW, Cvitkovitch DG, Bleiweis AS, Brady LJ: **Streptococcal viability and diminished stress tolerance in mutants lacking the signal recognition particle pathway or YidC2**. *Proc. Natl. Acad. Sci. U.S.A.* 2005, **102**:17466-17471.
378. Rolerson E, Swick A, Newlon L, Palmer C, Pan Y, Keeshan B, Spatafora G: **The SloR/Dlg metalloregulator modulates *Streptococcus mutans* virulence gene expression**. *J. Bacteriol.* 2006,**188**:5033-5044.
379. Li J, Wang W, Wang Y, Zeng AP: **Two-dimensional gel-based proteomic of the caries causing bacterium *Streptococcus mutans* UA159 and insight into the inhibitory effect of carolacton**. *Proteomics* 2013,**13**: 3470-3477.
380. Sudhakar P, Reck M, Wang W, He FQ, Wagner-Döbler I, Zeng AP. **Construction and verification of the transcriptional regulatory network of the human dental pathogen *Streptococcus mutans* upon treatment with the biofilm inhibitor Carolacton.** *BMC Genomics* 2014,**15**:362.