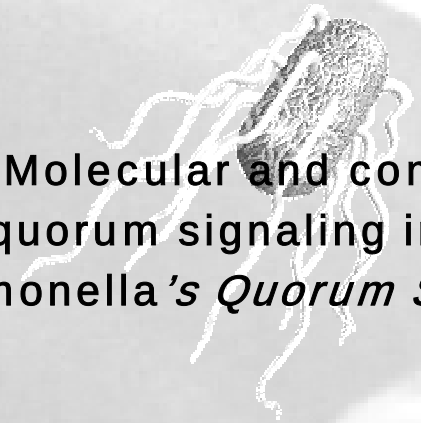
Period 01.10.02-28.02.07
Final report

GBOU-project 20160 SQUAD



Molecular and computational biology of quorum signaling in bacterial pathogens: Salmonella*'s Quorum Sensing to Attack Disease*

Prof. J. Vanderleyden (CMPG) –
Prof. B. De Moor (ESAT-SISTA) – Prof. J. Van Impe (BioTeC)

Contents

1. SCIENTIFIC AND TECHNICAL REPORT 4

1.1. Introduction 4

1.2. Overview of the work packages 5
Work Package 1: Development of molecular biological techniques for studying quorum sensing in *Salmonella* 5
Task 1.1 Construction of *S.* Typhimurium microarray 5
Task 1.2 Optimizing high-throughput 2D Difference in gel electrophoresis of *Salmonella* 6
Task 1.3 Optimizing phosphoproteomics for *Salmonella* 7
Task 1.4 Optimizing low-throughput reporter techniques for quorum sensing in *Salmonella* 7
Task 1.5 Development of chromatin immunoprecipitation (ChIP) DNA microarray (ChIP-chip) technology 8
Work Package 2: Development of Bioinformatics tools for *Salmonella* 9
Task 2.1 Development of a method for genetic network inference in *Salmonella* 9
Task 2.2 Development of a knowledge system of *Salmonella* 12
Task 2.3. Development of a method for prokaryotic phylogenetic footprinting 12
Task 2.4a Development of a preprocessing methodology for proteomics data 14
Task 2.4b Development of a preprocessing methodology for microarray data 14
Task 2.4c Development of a preprocessing methodology for ChIP-Chip data (additional task) 14
Application of tools on bacterial model systems 14
Work Package 3: Design and quantification of computer-controlled bioreactor experiments 15
Task 3.1 Macroscopic mass balance type modeling 15
Task 3.2 Microscopic individual based modeling (IBM) 19
Task 3.3 Optimal experiment design (OED) 19
Work package 4: Studying the influence of AI-2 mediated quorum sensing in *Salmonella* 22
Task 4.1 Optimizing the growth conditions for maximal expression of AI-2 synthesis and AI-2 dependent pathways in relevant environment 23
Task 4.2 Profiling of the dynamic behavior of *Salmonella* by high-throughput measurements in conditions eliciting AI-2 dependent quorum sensing 25
Task 4.3. Identification of AI-2 synthesis and AI-2 dependent pathways 31
Task 4.4. Additional experiments to validate the network structure 32
Task 4.5. Building a single cell model of the AI-2 dependent cell behavior 35
Task 4.6. Identification of additional targets of AI-2-signaling pathway 35
Work package 5: Identification of *Salmonella* specific drug targets by *in silico* genomics 35
Task 5.1. Motif finding and genome wide scanning 35
Task 5.2. Phylogenetic footprinting of the identified motifs 36
Task 5.3. Studying the evolutionary aspect of the spread of quorum sensing 37

1.3. Overall conclusions 39

1.4. List of deliverables 41

1.5. References 43

2. INTERACTION WITH USER COMMITTEE 47

2.1. Representative of IWT 47

2.2. Members of user committee 47

2.3. Subcontractors 47

2.4. Consortium members attending the user committee meetings 47

2.5. Overview of user committee meetings organized 47

2.6. General overview of user committee meetings 48

2.7. Perspectives for further utilization of results 48

3. VALORIZATION REPORT 50

3.1. Dissemination of results 50
International peer-reviewed papers 50
Conference proceedings 52
Abstracts and presentations 54
Dissertations (finished) 57
MSc Theses 57
Workshop organization 58

3.2. Contact with industries 59

3.3. SQUAD as inspiration for product development 59

3.4. SQUAD as seed for new research initiatives 59

1. Scientific and technical report

1.1. Introduction

The acquirement of resistance of pathogenic bacteria to common antibiotics and the development of multidrug resistant strains is raising alarms in healthcare and fueling demand for new antibiotics. Despite the apparent need for new and effective antibiotics, few novel drug targets have been identified and no new classes of antibiotics have been introduced in the last 20 years. Moreover, antibiotics have a broad range effect, killing also the beneficial intestinal microflora. Alternative more sustainable approaches to cope with these infectious bacteria are needed. Efforts are already made by several research groups to come up with alternative ways to combat bacterial infections. The prophylactic and therapeutic use of probiotics is in line with this context. A boom in functional food R&D activity is a result of this. In this same sector of functional food and nutraceuticals, a new industry is arising, based on the breakthrough discovery that a wide variety of pathogenic bacteria communicate via a common chemical "language" required to coordinate and regulate the onset of pathogenesis. This bacterial communication is termed quorum sensing. Most of the bacteria thus far identified that utilize quorum sensing are associated in some way with a host. Unraveling the basic mechanisms of quorum sensing is a major challenge in microbiology.

A major goal of this project is to gain insight into the mechanism of quorum sensing since it indefinitely is the clue to a plethora of distinct intriguing phenomena such as bacterial pathogenesis. This opens perspectives for the development of an entirely new class of broad-spectrum anti-infective drugs that disable disease-causing mechanisms in pathogenic bacteria. This goal will be obtained by dissecting a test case in which a global study of the influence of quorum sensing on *Salmonella* Typhimurium pathogenicity will be performed. In humans *S.* Typhimurium causes gastroenteritis. *Salmonella* applies a multistage pathogenesis. The expression of its virulence genes is controlled by complex regulatory networks in which quorum sensing is expected to play a crucial role. A broad variety of quorum-sensing signal molecules has been identified in the past 20 years which can roughly be divided in (i) acyl-homoserinelactones in Gram-negatives Whitehead *et al.*, 2001; (ii) processed oligopeptides in Gram-positives Sturme *et al.*, 2002; and (iii) AI-2 in both Gram-negatives and Gram-positives Xavier and Bassler, 2003. In contrast to the first two types of signal molecules which are usually involved in intraspecies communication, the AI-2 molecule has been proposed to serve as a ’universal’ signal for interspecies communication Surette *et al.*, 1999;McNab *et al.*, 2003. The synthase for AI-2 production, LuxS, is widely conserved among Gram-negative and Gram-positive bacteria Surette *et al.*, 1999. Many genera of bacteria, including *S.* Typhimurium, produce AI-2 Surette and Bassler, 1998;Ahmer, 2004.

Global expression profiling in combination with novel developments in bioinformatics affords an unprecedented opportunity for the genome-scale study of bacterial expression and will yield a portrait of interacting regulatory pathways. In this project we will study, by using such global approach, quorum sensing in *S.* Typhimurium. To achieve such goal the combination of the latest knowledge gained in research domains of molecular biology, bioinformatics, bioreactor technology and cell modeling is a prerequisite.

1.2. Overview of the work packages

Work Package 1: Development of molecular biological techniques for studying quorum sensing in *Salmonella*

In this work package the essential molecular biological tools to study AI-2 mediated quorum sensing in *Salmonella* at a global level were developed, i.e. *Salmonella* microarrays, 2D-DIGE, phosphoproteomics and low-throughput reporter techniques.

Task 1.1 Construction of* S. *Typhimurium microarray

Salmonella microarrays are not commercially available. Via our subcontractant Dr. P. Van Hummelen (MAF, VIB), we purchased a *Salmonella* oligo-set (70-mers representing the complete genome of *Salmonella,* i.e. ± 5500 genes) (Operon, Qiagen) to be spotted on CodeLink slides (Amersham). The hybridization protocol was optimized.

Different protocols for RNA isolation have been compared and adapted to our needs. To stabilize the RNA prior to isolation, the laboratory-made RNA protect reagens (5% acidic phenol in EtOH, v/v) was used. This resulted in pure, high quality RNA.

For downstream validation experiments, different protocols were optimized (gel shift analysis, Northern blotting etc.). In order to study gene expression in biofilms, a protocol for RNA isolation in these conditions was optimized. Moreover, since we extended our research to small, non-coding regulatory RNA molecules (sRNA, see WP4), we tested other RNA isolation procedures to be able to isolate sRNAs as well. Moreover, we developed a procedure to detect bacterial sRNAs by an adapted Northern analysis.

In addition to expression data, the complete genome sequence of *S.* Typhimurium (McClelland *et al.*, 2001) and related organisms is available. These sequence data reflect, at least in theory, the direct transcriptional interactions between regulatory proteins and their respective target genes. Indeed, regulatory motifs (short DNA sequence present in the promoter region of a gene acting as the recognition site of a specific regulatory protein) determine how basic networks are active in a cell. As such, the motif data are an important source of information for the inference of transcriptional networks. Motif detection exclusively based on sequence data has the advantage of being independent of the experimental condition. Indeed, the identified motifs mirror all possible interactions between regulatory proteins and their dependent genes. Concomitantly, the sensitivity of such an approach is high compared to a condition-dependent method (such as the identification of motifs based on expression data). However, without additional experimental validation, the risk of false positive motif predictions (low selectivity) exists. Until recently, it was only possible to prove the *in vivo* function of a given motif upstream of an open reading frame by means of labor intensive, low-throughput approaches (i.e. gene fusion experiments, site-specific mutagenesis, knock-outs). A recently implemented genome-wide location analysis, the chromatin immunoprecipitation (ChIP) DNA microarray (ChIP-chip) technology (see also Task 1.5), allows directly mapping of *in vivo* physical interactions between transcriptional regulators and their binding sites at a high-throughput level (Laub *et al.*, 2000; Lee *et al.*, 2002; Pollack *et al.*, 1999; Laub *et al.*, 2002) (Fig. 1). Briefly, a cell (bacteria, yeast, etc.) expressing a specific tagged transcriptional regulator is constructed. After growth under experimental conditions, DNA fragments bound to the tagged transcriptional regulator are recovered by a ChIP assay and hybridized to DNA microarrays containing the complete set of the cell's intergenic regions. Strong hybridization in a region proximal to a gene would indicate transcription factor binding to that gene's promoter site (Zhang and Gerstein, 2003). Unlike the genetic approach, which uses mutants, ChIP has the advantage of analysing transcriptional regulator occupancy under natural conditions. Consequently, integration of ChIP-chip location data with motif information and

expression data is a very powerful method to reconstruct regulons and basic network components in a cell (see also WP 2.1) (Li and Wang, 2003; Banerjee and Zhang, 2002; Zhang and Gerstein, 2003). Part of the ChIP-chip experiments has been performed using cDNA microarrays of our collaborator Prof. M. McClelland (SKCC, San Diego) (see also Task 1.5). However, since the motifs to which the regulatory protein of interest binds are situated in the upstream region of the coding sequence, arrays that represent the complete genome of *Salmonella*, i.e. tiling arrays, are more suited for this ChIP-chip technology. Therefore, *Salmonella* whole genome tiling arrays (NimbleGen Systems, Inc.) were implemented.

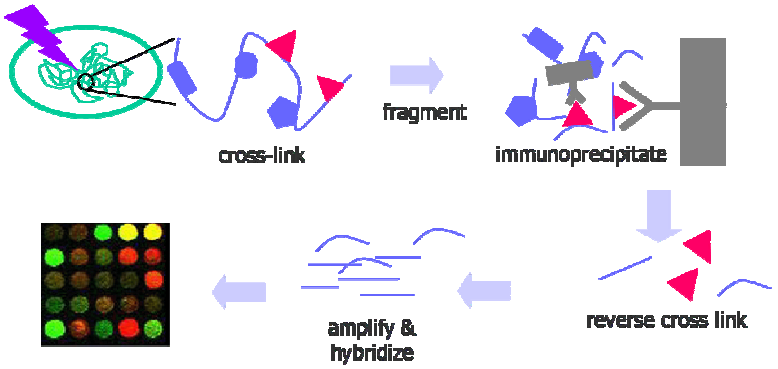


Fig. 1. ChIP-chip technology (for details, see text).

Task 1.2 Optimizing high-throughput 2D Difference in gel electrophoresis of Salmonella

At the start of the SQUAD project, the first technological objective concerning high-throuhgput proteomics research on *Salmonella* Typhimurium was to implement the 2D-DIGE technique at the CMPG. To this end a company screening for equipment was performed and a full proteomics platform was purchased, including an IPGphor, EttanDaltSix, Typhoon laser scanner and DeCyder Software (Fig. 2). During the timecourse of the project, upgrades for the software were implemented and the equipment was extended with an EttanDaltTwelve tank.



Fig. 2. Proteomics equipment implemented for SQUAD.

Optimization of the protocols needed for 2D-DIGE was performed. Different protocols for the different phases in the process were evaluated such as protocols for protein preparation (different cell fractions, different protein sources e.g. biofilms or mixed cultures), protocols for iso-electric focussing (different pI ranges, different protein application methods, different running protocols), protocols for protein labeling (different amounts of the cy-dyes), protocols for the second dimension (different gel types), and protocols for visualization (different staining methods). During this process of optimization we gained insight in the strong and weak points of the 2D-DIGE technology and were able to select a set of general standard operations protocols suited for the proteomic analysis of *Salmonella* Typhimurium (Fig. 3). Moreover, with the experience and technical knowledge obtained, the technology and protocols can now be easily adapted and applied in research of other micro-organisms.



Fig. 3. 2D-gel of *Salmonella* total proteins.

In order to validate the biological relevance of the 2D-DIGE technique for *Salmonella* Typhimurium an experiment was set up that compared protein profiles of a wild type strain grown under standard laboratory conditions and under *in vivo*-mimicking conditions. The latter represent a micro-aerobe, high osmolarity environment and genes that are influenced by these factors are well known and described in literature. From this experiment we could conclude that the proteins we picked up as being differentially expressed represented very well the applied culture conditions. Genes of which in literature transcription was shown to be influenced by either oxygen status or osmotic pressure reflected this effect qualitatively on the protein level (Sonck *et al.*, submitted). As such, proof was obtained that 2D-DIGE can be used as a solid technique to detect biological relevant changes in protein expression.

Task 1.3 Optimizing phosphoproteomics for Salmonella

Investigating the phosphorylation state of proteins at different growth phases is a very logic step as many proteins can be phosphorylated during their lifetime. Methods to determine the degree of phosphorylation were optimized (1-D and 2-D separation of protein samples, followed by Western blotting with antiphosphotyrosine, antiphosphoserine, antiphosphothreonine antibodies; Pro-Q Diamond stain (Molecular Probes)). In parallel, an in silico analysis of the published *Salmonella* genome was carried out in order to reveal motifs for phosphorylation of *Salmonella* proteins (Master thesis of S. Yellajosyula, K.U. Leuven, 2003).

As elaborated in WP 4, we found indications that LuxS occurs as a phosphorylated protein in the cell. To identify the phosphorylation site(s) of the LuxS protein, we started several collaborations with research groups active in this domain (Prof. E. Waelkens and Dr. R. Derua (Biochemistry, Fac. Medicine, K.U. Leuven; Prof. P. Proost (Moleculaire Immunologie, REGA, KUL; Prof. M. Bollen and Prof. M. Beullens (Dep. Molecular cell biology – Biochemistry, Fac. Medicine, KUL); Prof. Jef Rozenski (Laboratory for Medicinal Chemistry, KUL).

Task 1.4 Optimizing low-throughput reporter techniques for quorum sensing in Salmonella

Reporter fusions (with GFP, LacZ, Lux) were constructed to monitor expression level of *Salmonella luxS*, *pfs, lsrR* and *lsrA* genes at a low-throughput level. All methods required to monitor gene expression based on *gfp* or *lacZ* fusions have been optimized using other test cases, ie. the HilA (*lacZ*) and PmrA (*gfp*) regulon in *Salmonella*.
Besides the fusion reporter based measurements, expression of all these key genes can be determined by using RT-PCR.
During the project, different *Salmonella* mutants of genes involved in the AI-2 pathway have been constructed using the linear transformation methods of Datsenko and Wanner (Datsenko and Wanner, 2000). For all mutants, complementation constructs have been made.
In collaboration with Prof. D. De Vos (C.O.K., K.U. Leuven), we succeeded in synthesizing AI-2 molecules that can also be used for complementation (De Keersmaecker *et al.*, 2005; see also WP 4).

To monitor AI-2 levels, the standard *Vibrio harveyi* bioassay (Bassler *et al.*, 1997) has been adapted. Attempts have been made to use *Salmonella* as a bioreporter (using the *lsrA* reporter fusion), however, this assay turned out to be less sensitive than the *Vibrio* one. Moreover, the bioassay only allows a qualitative assessment of the AI-2 levels present in a cell-free culture supernatant. The lack of an effective, highly quantitative AI-2 assay resulted in an additional challenge in the modelling work (see WP 3).
To monitor gene expression at the protein level, LuxS antibodies (His_6-tagged LuxS protein) have been raised. Methods for Western blotting have been optimized.

Task 1.5 Development of chromatin immunoprecipitation (ChIP) DNA microarray (ChIP-chip) technology

This task was not in the original project proposal. However, given the power of this technique and the impact on data analysis, especially for the integration of heterogeneous data (see also WP 2.1), we decided to develop the ChIP-chip technology for *S.* Typhimurium.

The ChIP-chip procedure was implemented in *S.* Typhimurium. As a test system, the *Salmonella* HilA regulator was selected. A construct was developed to tag the regulator with a Myc epitope. This Myc epitope can be recognized by an anti-Myc antibody. This procedure is preferred over the use of a regulator-specific antibody to perform the immunoprecipitation (IP), since it is more universal. The correct tagging of the regulator was verified by sequencing and Western blotting. The functionality of the tagged regulator was validated by using reporter gene fusions.
Protocols for the IP were developed and optimized. PCR was used to verify IP of known target genes.

By combining ChIP-chip and transcriptomics assays, comparing gene expression in a wildtype and a *hilA* mutant strain, a number of new target genes of the HilA regulator could be identified, thereby extending our knowledge of the regulatory network controlling the invasion of epithelial cells (Thijs *et al.*, accepted for publication in J. Bacteriol.).

For LsrR, the same set-up has been applied, with the improvement that *Salmonella* whole genome tiling arrays (NimbleGen Systems, Inc.) are used for ChIP-chip; the application of the full ORF microarrays (Prof. M. McClelland, SKCC, San Diego), that proved to be useful for HilA, was ineffective for LsrR. Succesful immunoprecipitation of the known LsrR target gene *lsrA*, however, could be confirmed by PCR. Figure 4 shows the presence of the *lsrA* promoter fragment in the genomic input DNA and in the immunoprecipitated (IP) DNA of the ChIP samples (three replicates, Fig. 4 B), but not in the control samples (three replicates, Fig. 4 A), thus confirming the specificity of our ChIP assay. We are currently waiting for the results of the ChIP samples that have been shipped to NimbleGen Systems, Inc. for hybridization onto whole genome tiling arrays.

The ChIP-chip technology is also being applied for the YijC regulator (see also WP4).

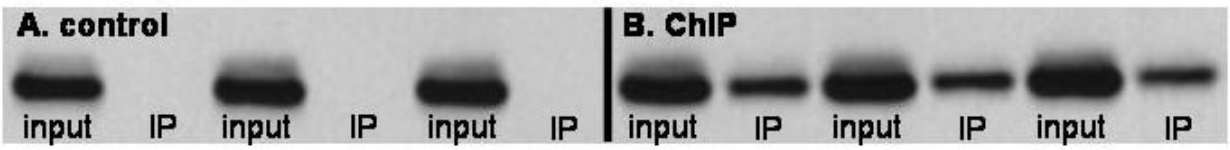


Fig. 4: Presence of the promoter region of *lsrA* in each three replicates of the control (A.) and ChIP samples (B.), as assessed by PCR. The *lsrA* promoter region is detected in the genomic input DNA (A. and B. input) and in the immunoprecipitated DNA of the ChIP samples (B. IP), but not in that of the control samples (A. IP).

Work Package 2: Development of Bioinformatics tools for *Salmonella*

In this work package the essential bioinformatics tools to study AI-2 mediated quorum sensing in *Salmonella* at a global level, were developed, i.e. a method for genetic network inference, knowledge system for *Salmonella*, a method for prokaryotic phylogenetic footprinting and a method for preprocessing data. The different progresses on bioinformatics development have been summarized in a schematic representation (Fig. 5).

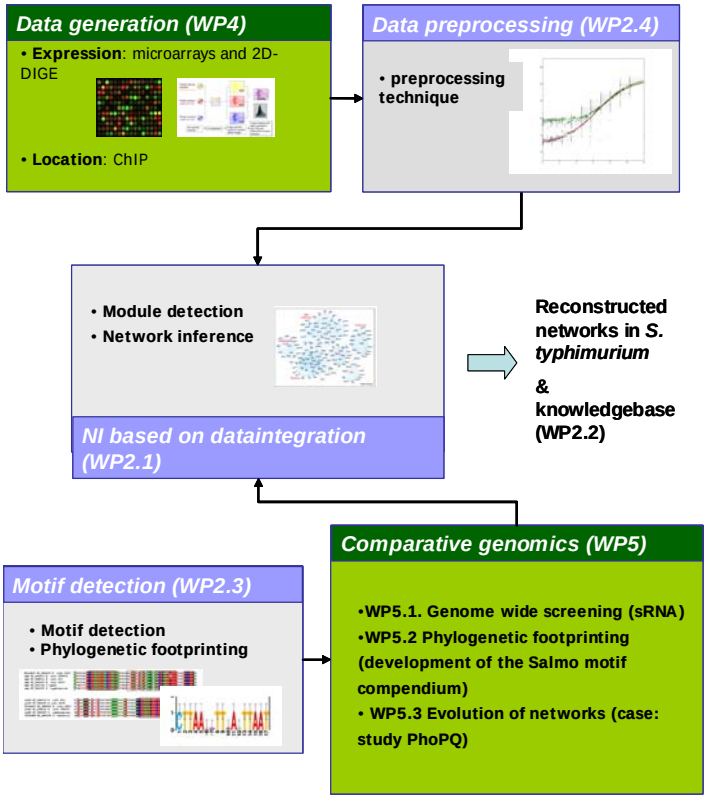


Fig. 5 Overview of the SQUAD-bioinformatics and situation within the remainder of the project

Task 2.1 Development of a method for genetic network inference in Salmonella

1) As network inference is a very competitive domain, we performed at regular time intervals during our algorithmic work, **exhaustive literature overviews** in order to situate and tune our research according to recent developments in the field. These reviews have been published:

T. Van den Bulcke, K. Lemmens, Y. Van de Peer, K. Marchal. Inferring transcriptional networks by mining 'omics' data. Current Bioinformatics, 2006, 1: 301-313.

S. C. De Keersmaecker, I. M. Thijs, J. Vanderleyden, and K. Marchal. Integration of omics data: how well does it work for bacteria? Molecular Microbiology, 2006, 62:1239-50.

The global idea of our work flow is depicted in the figure 6.

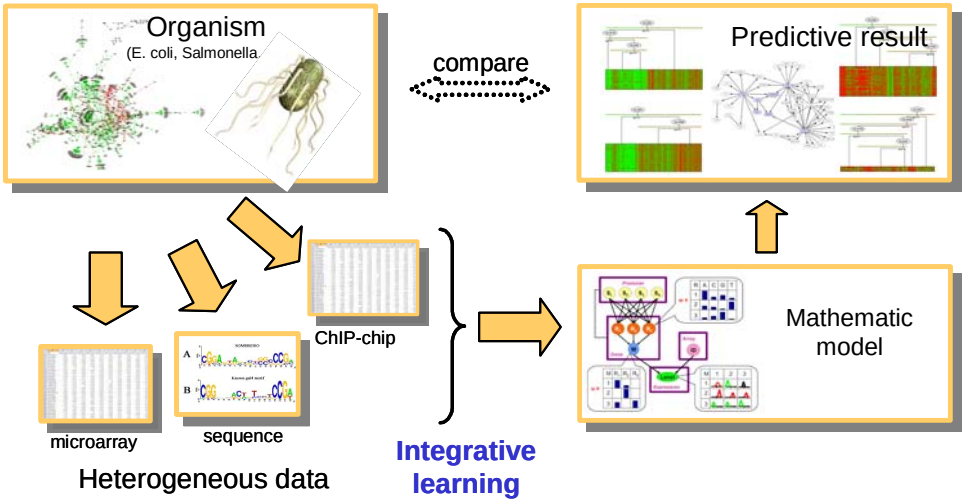


Fig 6. Data on a specific model organism will be collected and properly stored and preprocessed. They will subsequently be integrated by means of an appropriate model (Own model or model previously decribed in literature). A prediction will be generated and the reliability (specificity, sensitivity of the results need to be assessed). As biological systems are still sparsely characterised we will mainly make use of a biologically relavnt generator to test and benchmark our algorithms.

2) We constructed ***a simulator*** to generate biologically relevant artificial data to test the performance of our network inference tools to be developed in the near future. Using artificial data is essential to exhaustively test the sensitivity, specificity of an algorithm under different experimental settings.

T. G. Van den Bulcke, K. M.F. Van Leemput, B. Naudts, P. van Remortel, H. Ma, A. Verschoren, B. De Moor, and K. Marchal. SynTReN: A generator of synthetic gene expression data for design and analysis of structure learning algorithms. BMC Bioinformatics, 2006, 26:7-43

In the mean time, this work has been presented on several conferences.

T. Van den Bulcke, K. Van Leemput, K. Marchal, P. van Remortel, A. Verschoren, B. Naudts, and B. De Moor A synthetic gene expression data generator for validation of reverse engineering methods. ECCB, 2005. 28 september-1 October.

T. Van den Bulcke, K. Van Leemput, P. van Remortel, B. De Moor, K. Marchal. SynTreN: a synthetic gene expression data generator for validation of reverse engineering methods. 29-31 May, 2006, Amsterdam, International Symposium on networks in bioinformatics. Oral presentation.

T. Van den Bulcke, K. Van Leemput, P. van Remortel, B. Naudts, B. De Moor, and K. Marchal. Benchmarking gene network inference algorithms using synthetic gene expression data. Benelearn-Workshop on bioinformatics & nature inspired computing, May 10 - Ghent – Belgium. Oral presentation.

T. Van den Bulcke, K. Van Leemput, T. Dhollander, B. De Moor, P. van Remortel, K. Marchal. Using synthetic gene expression data to assess characteristics of gene network inference algorithms. ISMB 2006, Aug 6-10, Fortaleza, Brasil.

T. Van den Bulcke, K. Van Leemput, T. Dhollander, B. De Moor, P. van Remortel and K. Marchal. Assessing characteristics of gene network inference algorithms using synthetic gene expression data. Workshop: Probabilistic Modeling and Machine Learning in Structural and Systems Biology June 17-18. Finland.

3) Based on a literature search, we could pinpoint a few interesting tools for network inference. To have a better grip on their performance and potential drawbacks we analysed them using SynTReN. These comparative papers have been published recently:

K. Van Leemput, T. Van den Bulcke, T. Dhollander, B. De Moor, K. Marchal and P. van Remortel. Exploring the operational characteristics of inference algorithms for transcriptional networks by means of synthetic data. In press, Artificial Life.

4) We implemented a first ***tool for module inference*** based on the integration of heterogeneous omics datasources (ReMoDiscovery) (Fig. 7).

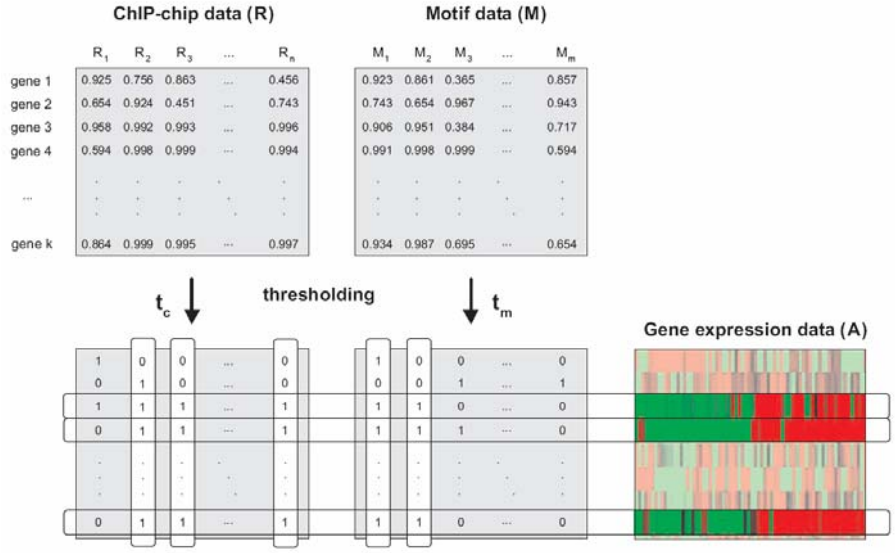


Fig. 7. RemoDiscovery is based on the a priori algorithm for combining item sets. It searches for the maximal gene set for which the expression profile is suffiently correlated, the genes share a similar set of motifs and are bound by the same regulators according to the ChIP-Chip data.

The software of this tool is availble and has been described in the following publications

T. Debie, P. Monsieurs, K. Engelen, B. De Moor, N. Christianini, and K. Marchal. Discovering regulatory modules from heterogenous information sources. Proceedings in Pacific Symposium on Biocomputing, Hawai 4-8 jan, 2005, http://helix-web.stanford.edu/psb05/..

K. Lemmens, T. Dhollander, T. De Bie, P. Monsieurs, K. Engelen, B. Smets, J. Winderickx, B. De Moor, and K. Marchal. Inferring transcriptional module networks from ChIP-chip-, motif- and microarray data. Genome Biology, 2006, 7:R37.

In the mean time, this work has been presented on several conferences:

K. Lemmens, T. De Bie, P. Monsieurs, K. Engelen, B. De Moor, N. Cristianini , and K. Marchal. Discovering regulatory modules from heterogeneous information sources. ECCB, 2005. 28 september-1 October. Oral presentation student symposium.

K. Lemmens, T. Dhollander, T. De Bie, P. Monsieurs, K. Engelen, B. Smets, J. Winderickx, B. De Moor, and K. Marchal. Inferring transcriptional module networks from ChIP-chip-, motif- and microarray data, 29-31 May, 2006, Amsterdam, International Symposium on networks in bioinformatics. Oral presentation.

K. Lemmens, T. Dhollander, T. De Bie, P. Monsieurs, K. Engelen, B. De Moor, K. Marchal. Inferring transcriptional modules from ChIP-chip-, motif- and microarray data. ISMB 2006, Aug 6-10, Fortaleza, Brasil.

Work in progress on RemoDiscovery

1) Further extension of ReMoDiscovery:
    a. for the combined detection of regulatory modules and motifs using suffix trees (still under development)
    b. to take into account condition dependency i.e., searching for modules that are only active in a subset of conditions (implementation available, in testing phase)

4) RemoDiscovery is a deterministic framework. This offers the advantage of being highly specific but at the expense of having a very low sensitivity. Therefore we are currently working on the extension of the deterministic framework ***into a probabilistic one***.

In the framework of module detection, we have decided to use the probabilistic relational framework originally developed by Koller and Pfeffer (1998) and to extend it for the integration of heterogeneous data such as ChIP-chip data, protein interaction data etc.

- So far we have set up an interuniversity collaboration to tackle this challenging problem. We started with the implementation of the basic framework as described by Segal *et al.* (2003) in a concerted effort. This resulted in a first implemementation which has been tested using SynTReN. This resulted in the following publication:

  T. Michoel, S. Maere, E. Bonnet, A. Joshi, T. Van den Bulcke, K. Van Leemput, P. van Remortel, M. Kuiper, K. Marchal and Y. Van de Peer. Validating module network learning algorithms using simulated data. BMC Bioinformatics, accepted.

Work in progress on PRM based models
A basic implementation of the probabilistic relational model is ready by now (within the timeframe of this GBOU project). However, further optimalization and extensions will be the work of follow up projects (SBO BioFrame approved).

- We will identify the most appropriate way to integrate heterogeneous information (constructing the right ontology for the PRM model).
- We will tackle the problem of experiment design.

***Task 2.2 Development of a knowledge system of* Salmonella**

To start with a knowledge base, we first constructed ***a motif database***. This database should store in a comprehensive way all information related to transcriptional motifs in *Salmonella*. We use the *Bacillus* database as a model (http://dbtbs.hgc.jp/). Some initial steps towards the construction of a *Salmonella* motif database were undertaken.

- A database scheme was constructed
- A first implementation is ready for β testing

This work is still in progress. Currently, *Salmonella* data is being loaded into the database (local server).

We also started with the construction of a ***microarray database***. This database contains all publicly available data on prokaryotic model organisms, together with relevant information. It is tuned to facilitate the interface with our preprocessing and dataintegration scripts. The database is in mySQL. In collaboration with the computer science department, we are setting up an automatic update system for our database.

These databases contain the basic information but should be compatible with the ontologies underlying the dataintegration models. Further extension of these databases is planned within the SBO project BioFrame.

Task 2.3. Development of a method for prokaryotic phylogenetic footprinting

1) In the framework of this project, we have constructed an automatic analysis flow for phylogenetic footprinting (ortholog detection by Blast, selection of intergenics, interation with Motif (Block Sampler), downstream analysis of results). Application of this toolbox to *S.* Typhimurium has allowed us to study the PmrA, PhoP and HilA regulons (Marchal *et al.*, 2004; Monsieurs *et al.*, 2005; Thijs et al. (accepted)) (see also WP 5).

2) We developed improved motif detection tools for phylogenetic footprinting: BlockSampler, is an extension of MotifSampler (Thijs *et al.*, 2002) providing a more advanced implementation of a higher-order background model (Thijs *et al.*, 2001; Marchal *et al.*, 2003). The tool BlockSampler has been extensively described in Van Hellemont *et al.* (2005), 2) BlockAligner, is meant to align motif matrices using dynamic programming and an adapted scoring scheme based on the information content. The tool BlockSampler has been extensively described in Monsieurs *et al.*

(2005). Both tools are made available as a stand alone software tools and are described in the following publications:

P. Monsieurs, G. Thijs, A. A. Fadda, S. C.J. De Keersmaecker, J. Vanderleyden, B. De Moor and K. Marchal. More robust detection of motifs in coexpressed genes by using phylogenetic information. BMC Bioinformatics, 2006, 7:160.

R. Van Hellemont, P. Monsieurs, G. Thijs, B. De Moor, Y. Van de Peer, K. Marchal. A novel approach to identify regulatory motifs in distantly related genomes. Genome Biology, 2005, 6:R113.1-R113.17, Epub.

This work has been presented on the following conferences:

P. Monsieurs, G. Thijs, A. A Fadda, S. C. J. De Keersmaecker, J. Vanderleyden. B. De Moor, K. Marchal. More robust detection of motifs in coexpressed genes by using phylogenetic information. ASM 2006, Sept 9-13, Victoria, Canada.

A. Fadda, P. Monsieurs, K. Marchal. A compendium of predicted motifs for Bacillus subtilis. ESF 2006, San Feliu, Spain.

3) Our latest research is focusing on the genome wide scale application of the methodology to construct a compendium of potential *S.* Typhimurium and *E. coli* motifs (See Task 5.2). The analysis flow for this application makes use of both BlockSampler and BlockAligner. To this end a fully automated analysis flow was made which is represented in Fig 8. This work is still in progress. Retrieved motifs are currently being benchmarked with motif databases (RegulonDB). The final result will serve as input, together with the *Salmonella* microarray and ChIP-chip data for our data integration tools.

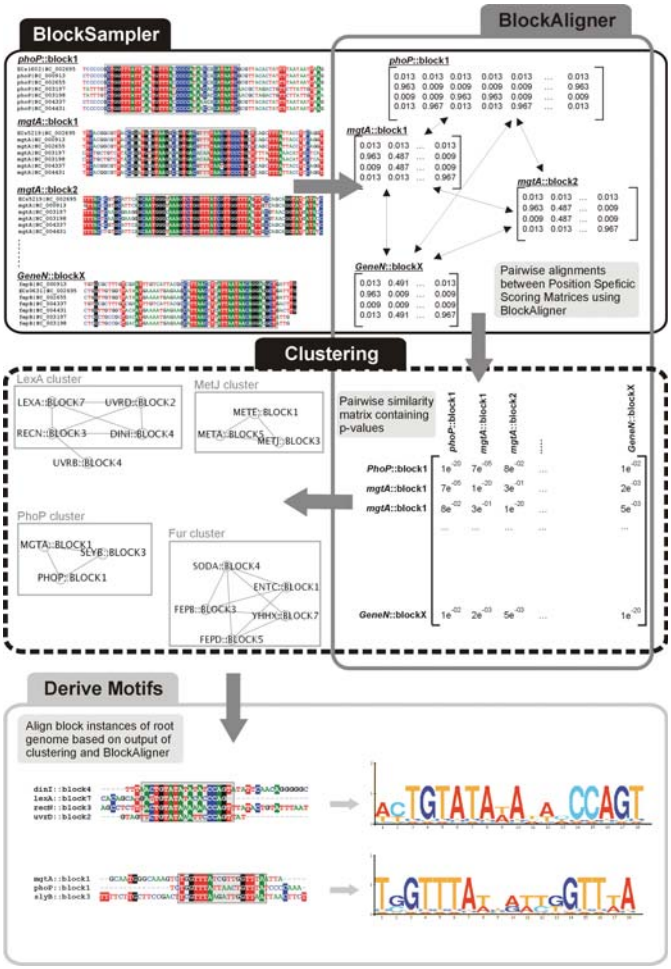


Fig 8. In a first step, conserved regions in orthologous promoter regions (i.e. blocks) are identified. In a second step, those conserved promoter regions, retrieved from different sets of orthologs and stored as position specific scoring matrices, are aligned to each other using a newly developed tool BlockAligner. In a third step, a clustering algorithm tries to group those blocks that share a common regulatory motif and derives a motif model out of these clusters. The last step consists of a screening of all input promoter regions with the detected motif models to retrieve those instances that could not be detected because of the absence of orthologs in closely related species.

Task 2.4a Development of a preprocessing methodology for proteomics data

The data of the proteomics experiments have been analysed using the DeCyder software (Amersham). This software allows high quality data analysis. Additional preprocessing methods are not necessarily required at this stage.

However, for microarray data, preprocessing is crucial for high-quality data analysis. Therefore, an additional task has been performed.
The problem with the 2D DIGE gels, however, is the mapping of spots between experiments. This step is essential in order to generate a protein profile, comparable to an mRNA expression profile, i.e, the changes of a particular protein over time or over conditions. This task will be tackled within the new SBO proposal.

Task 2.4b Development of a preprocessing methodology for microarray data

Normalization of microarray measurements is the first step in a microarray analysis flow. It aims at removing consistent sources of variations to make measurements mutually comparable. Reliable normalization is essential since the results of all subsequent analyses (such as e.g. clustering) might largely be influenced by the normalization procedure. In this proposal we developed a new way of normalizing cDNA microarray data, especially appropriate for prokaryotes for which large changes in gene expression can be expected. Details of our methodology can be found in

K. Engelen, B. Naudts, B. De Moor, K. Marchal. A calibration method for estimating absolute expression levels from microarray data. Bioinformatics, 2006, 22:1251-8.

The implementation has been made available as an R Bioconductor open source package "CALIB". The implementation has been accepted for publication in an international peer-reviewed journal.

H. Zhao, K. Engelen, B. De Moor, K. Marchal. CALIB: a Bioconductor package for estimating absolute expression levels from two-color microarray data. Availability: http://bioconductor.org/packages/2.0/bioc Open Source

H. Zhao, K. Engelen, B. De Moor, K. Marchal. CALIB: a Bioconductor package for estimating absolute expression levels from two-color microarray data. Bioinformatics, accepted for publication

The work has been presented on the following conferences:

K. Engelen, B. Naudts, K. Van Leemput, B. De Moor and K. Marchal. Normalization of cDNA microarrays using external control spikes. ECCB, 2005. Madrid, 28 september-1 October.

K. Engelen, B. De Moor, K. Marchal. A calibration method for estimating abolute expression levels from microarray data. 29-31 May, 2006, Amsterdam, International Symposium on networks in bioinformatics. Oral presentation.

K. Engelen, B. De Moor, K. Marchal. A calibration method for estimating absolute expression levels from microarray data. ISMB 2006, Aug 6-10, Fortaleza, Brasil.

Task 2.4c Development of a preprocessing methodology for ChIP-Chip data (additional task)

Although ChIP-Chip data are related to microarray data, they require a different design and adapted normalization procedures. These were developed in the framework of this project. Details on the preprocessing flow of the ChIP-chip data can be found in Thijs et al. (accepted) (see also Task 1.5).

Application of tools on bacterial model systems

So far we have been field testing our tools on yeast. However, yeast, being a eukaryotic organism might complicate the finetuning of algorithms towards prokaryotes, for instance typical operon

structure is not taken into account so far. As our tools should be applicable to *Salmonella*, we decided to work on designing a benchmark set for a prokaryotic model organism *E. coli*. To this end we collaborate with Prof. Collado-Vides.
(http://www.cifn.unam.mx/Computational_Genomics/computational_genomics.html).

So far we have downloaded, parsed and preprocessed more than 800 *E. coli* microarrays from different databases (Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/), Stanford microarray database (http://genome-www5.stanford.edu/) and ArrayExpress (http://www.ebi.ac.uk/arrayexpress/)) and different platforms (Affymetrix and cDNA microarrays). These data have been stored in local databases (Task 2.2) and have been properly preprocessed (Task 2.4).
An internal compendium of 153 regulatory motifs was built up by motif detection in *E. coli*. Another 95 motifs were downloaded from the RegulonDB database (http://regulondb.ccg.unam.mx/) and added to this collection of motifs (Task 2.2). The *E. coli* genome was screened with the compendium of regulatory motifs to find more and possibly new hits for these motifs (Task 2.3).
This collection of microarray data and motif data will be used to finetune the developed algorithms towards prokaryotes. Currently we are applying the extended version of RemoDiscovery on these data (Task 2.1).

Remark also that in the other work packages (WP 1, WP 5), we have been focusing on the generation of additional data sources (besides the microarray data) e.g. the motif compendium, the ChIP-chip data, proteomics data. These data can be used in our data integration tools to infer modules in *Salmonella.* The goal is to collect on the long run more data on *Salmonella*, as such that we can apply al our tools on this system as well. To collect a compendium of microarray data we plan a visit to the lab of Prof. McClelland (Sidney Kimmel Center San Diego). He has the primary database with over 300 microarrays in *Salmonella*. Motif compendia are constructed as outlined in WP 2.3.

Work Package 3: Design and quantification of computer-controlled bioreactor experiments

Task 3.1 Macroscopic mass balance type modeling

Bioreactor experiments were delayed due to the move of BioTeC, and effectively started in November 2003. At first, time was invested in the definition of appropriate experimental conditions, especially, the selection of carbon substrate and the dissolved oxygen (DO) level. Some changes were made with respect to the original proposal. (*i*) In a first stage, glucose was used as a carbon source but, as shown within this project, glucose hampers the *Vibrio* AI-2 bioassay for measurement of AI-2 (De Keersmaecker and Vanderleyden, 2003). Therefore, it was decided to omit glucose from the LB medium and to add no alternative carbon source. However, due the large heterogeneity in the growth dynamics between repeated experiments of *S.* Typhimurium in LB and the difficulties in identifying the growth-determining substrate(s), it was decided since the start of the second year of the research project to add galactose as a carbon source to LB medium. (ii) An additional change to the experimental conditions is made at the level of the set point of the dissolved oxygen. The dissolved oxygen (DO) concentration is augmented to 30% (more details are given in Task 4.1). A dissolved oxygen concentration of 30 % does not reflect the actual oxygen concentration in the gut. However, we realize that the bioreactor conditions will not be sufficient to study AI-2 mediated quorum sensing in *Salmonella*. As elaborated in WP 4, we will have to more closely mimic the natural niche of *Salmonella*, such as biofilms, if we want to unravel the *in vivo* function of AI-2 in *Salmonella*. For this reason, we opted to design the bioreactor experiments in such a way that they are highly reproducible and can be used to disentangle the AI-2 pathway in free-living *Salmonella* (which are the conditions

used in most of the reports in literature) under very carefully controlled environmental conditions (pH, oxygen etc.). This has not been done yet for *Salmonella*. At 30 % DO, deviations in the dissolved oxygen concentration do not cross the boundary between (micro-)aerobic and anaerobic conditions initiating switches between metabolic pathways. Furthermore, the set point of 30 % DO can be realized more accurately within the bioreactors.

In a first modeling stage, growth kinetics were supposed to be driven by the galactose concentration (as a growth-limiting substrate) according to a Monod kinetics. Optimized fed-batch experiments (designed under Task 3.3) and additional batch (with different galactose levels) clearly pointed out that this model structure cannot track the biomass data properly. In a second modeling, additional informative experiments were designed and performed to elucidate the model structure. Additional experiments included continuous reactor experiments during which the dilution rate was repeatedly changed (more details given in WP 4).

Based on all available *macroscopic* bioreactor data, a macroscopic mass balance type model has been developed for growth and AI-2 profiles of *S.* Typhimurium growing in LB supplemented with galactose, at 37°C, pH 6 and 30% dissolved oxygen. Mechanistic knowledge and observations from experiments have been embedded in a semi-structured (*semi-mechanistic*) macroscopic mass balance type model. The model structure contains five mass balance equations for the following state variables: biomass, galactose, storage carbohydrate (possibly pyruvate, not measured), acetic acid and AI-2. Briefly, the model structure reflects the following dynamics: (*i*) galactose is metabolized, as a *preferential* substrate (and not as a growth-limiting substrate), into storage carbohydrate, (*ii*) the storage carbohydrate is partly used for biomass growth (entering the TCA cycle) and partly secreted as acetic acid (i.e., oxidative overflow), (*iii*) acetic acid is first produced and, after complete depletion of the internal pool of pyruvate, internalized for biomass formation, (*iv*) AI-2 production and intake are linked to the preferential substrate metabolism and the biomass concentration and AI-2 concentration, respectively. Figure 9 depicts a schematic overview of the model for growth and overflow. The model equations are presented in Table1.

The model equations in Table 1 can be explained as follows. (*i*) Growth is modeled in two phases. Substrate (C_S), i.e., galactose, is internalized and used for both growth (first growth phase) and acetate formation (overflow metabolism). The first growth phase is modeled by a logistic equation. Once all the internalized substrate or storage carbohydrate (C_I) is consumed, the bacterial population starts growing on acetate (C_A) (second growth phase). The switch from one substrate (i.e., galactose) to another (i.e., acetate) is implemented in the model by Monod-type equations with a very low Monod constant (ε). (*ii*) The measured AI-2 concentration in the medium is modeled as the overall result of production and uptake. Production of AI-2 is, in accordance with the experimental data, assumed to be correlated with preferential substrate metabolism. The Monod-term with C_{AI2} in the differential equation for AI-2 reflects the fact that AI-2 is needed to suspend the repression of LsrR on the AI-2 transporter and by doing so, to regulate its own uptake.
Simulations were compared with experimental data. Results are displayed in Figures 10, 11 and 12. For the equations describing growth and overflow metabolism (see Table 1), the same parameter values are employed for all simulations. For the description of AI-2, parameters had to be adapted for each data set. The need for different parameters is a result of the non-quantitative nature of the AI-2 measurement (Turovskiy and Chikindas, 2006).
Sensitivity analysis as described by Smets *et al.* (2002) was carried out for the model describing growth and AI-2 production and uptake. All 10 model parameters were found to be essential to describe the states (i.e., cell density, galactose, acetic acid and AI-2). Hence, no model reduction is needed. This result is not surprising since we focused on keeping the model as simple as possible during the model building process.

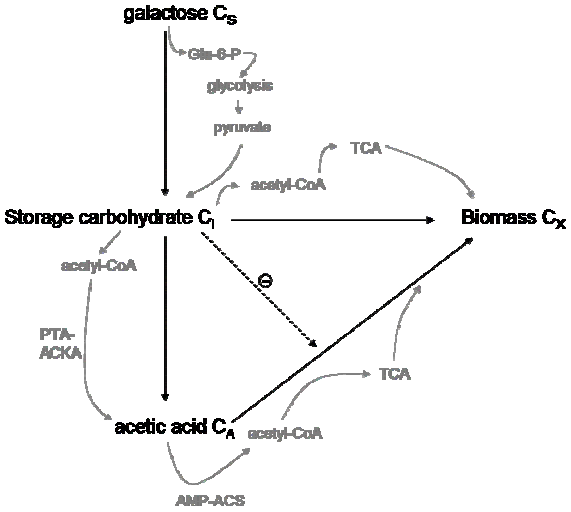


Fig. 9 Schematic overview of the model for growth and overflow metabolism. The states and reaction described by the macroscopic model are depicted in black. Reaction in gray represent underlying biochemical pathways.

Differential equations:

$$\frac{dC_X}{dt} = \left(\mu_{C_I} \cdot (1 - \frac{C_X}{C_{X,max}}) + \mu_{C_A}\right) \cdot C_X - D \cdot C_X$$

$$\frac{dC_S}{dt} = -\sigma_{C_S} \cdot C_X + D \cdot (C_{S,in} - C_S)$$

$$\frac{dC_I}{dt} = \sigma_{C_S} \cdot C_X - \sigma_{C_I} \cdot C_X - D \cdot C_I$$

$$\frac{dC_A}{dt} = Y_{A,X} \cdot \mu_{C_I} \cdot C_X - \sigma_{C_A} \cdot C_X - D \cdot C_A$$

$$\frac{dC_{AI2}}{dt} = \frac{\sigma_{C_S}}{Y_{AI2}} \cdot C_X - \sigma_{m,AI2} \cdot \frac{C_{AI2}}{C_{AI2} + \epsilon} \cdot C_X - D \cdot C_{AI2}$$

$(D = \frac{U}{V} = \text{dilution rate})$

Specific rates:

$$\mu_{C_I} = \mu_{m,I} \cdot \frac{C_I}{C_I + \epsilon}$$

$$\mu_{C_A} = \mu_{m,A} \cdot \frac{C_A}{C_A + \epsilon} \cdot (1 - \frac{C_I}{C_I + \epsilon})$$

$$\sigma_{C_S} = (\sigma_{m,S} \cdot (1 - \frac{C_X}{C_{X,max}}) + m) \cdot \frac{C_S}{C_S + \epsilon}$$

$$\sigma_{C_A} = \frac{\mu_{C_A}}{Y_{XA}}$$

$$\sigma_{C_I} = \frac{\mu_{C_I}}{Y_{XI}}$$

$(\epsilon = 10^{-3})$

List of symbols

C_X	[OD]	cell density	σ_{C_S}	[g/(OD·L)·h^{-1}]	specific consumption rate for C_S
C_S	[g/L]	galactose concentration	$\sigma_{m,S}$	[g/(OD·L)·h^{-1}]	coefficient consumption C_S
C_I	[g/L]	storage carbohydrate	m	[g/(OD·L)·h^{-1}]	maintenance factor
C_A	[g/L]	acetate concentration	σ_{C_I}	[g/(OD·L)·h^{-1}]	specific consumption rate for C_I
C_{AI2}	[-]	AI-2 concentration	$Y_{X,I}$	[(OD ·L)/g]	yield coefficient of C_X on C_I
$C_{X,max}$	[OD]	maximum cell density	$Y_{A,X}$	[-]	yield coefficient for acetate production
μ_{C_I}	[h^{-1}]	specific growth rate on C_I	σ_{C_A}	[g/(OD·L)·h^{-1}]	specific consumption rate for C_A
$\mu_{m,I}$	[h^{-1}]	coefficient growth on C_I	$Y_{X,A}$	[(OD ·L)/g]	yield coefficient of C_X on C_A
μ_{C_A}	[h^{-1}]	specific growth rate on C_A	Y_{AI2}	[(g/L)$^{-1}$]	yield coefficient for AI-2 production
$\mu_{m,A}$	[h^{-1}]	coefficient growth on C_A	$\sigma_{m,AI2}$	[(1/OD)·h^{-1}]	coefficient uptake C_{AI2}
U	[L/h]	feed rate			
V	[L]	volume			

Table 1 Semi-structured macroscopic model for growth and AI-2 of *S.* Typhimurium SL1344.

Incorporating mechanistic knowledge obtained from high throughput experiments in a macroscopic model is not straightforward. High-throughput analyses, such as microarray and proteomics analyses, result in non quantitative data and are used to compare different situations rather than to follow the evolution in time. For the model presented in this report, mechanistic knowledge concerning the *lsr*-operon (results from WP 4 and Taga *et al.* (2003)) is incorporated in the model to describe AI-2 uptake.

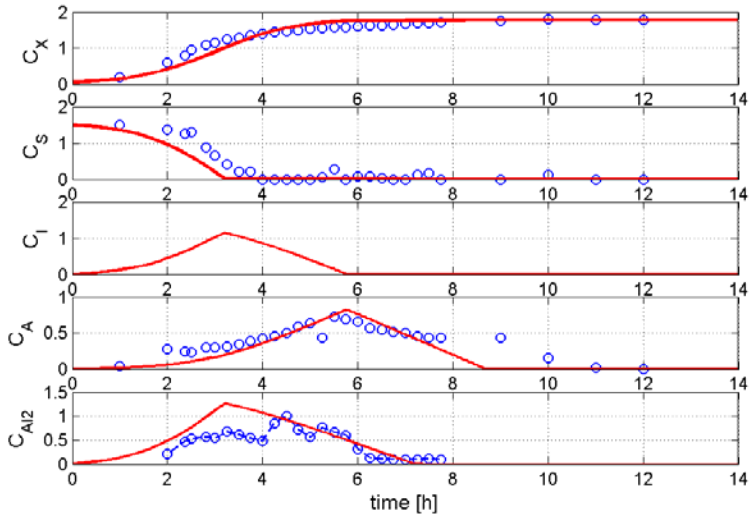


Fig. 10 Experimental data and model simulations for a batch experiment with an initial substrate concentration of 1.5 g/L.

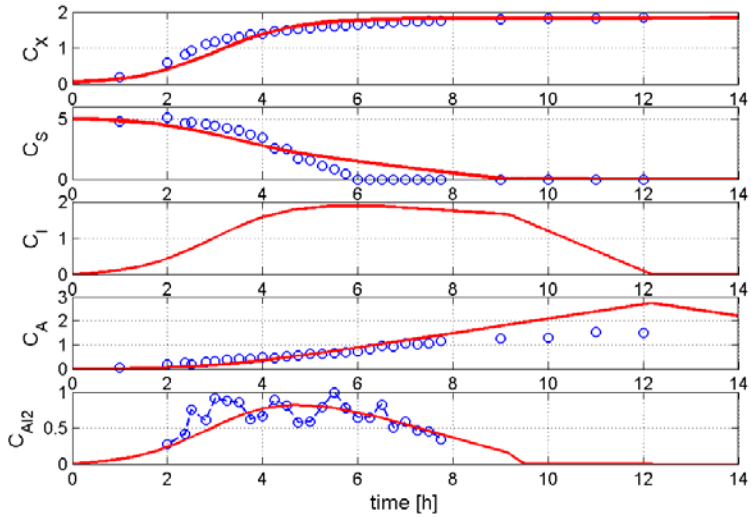


Fig. 11 Experimental data and model simulations for a batch experiment with an initial substrate concentration of 5 g/L.

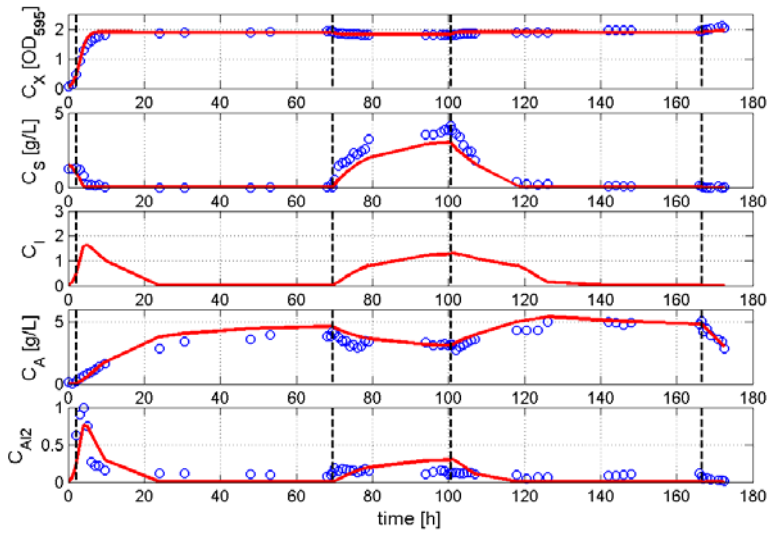


Fig. 12 Experimental data and model simulations for a continuous experiment with shifting dilution rates.

Task 3.2 Microscopic individual based modeling (IBM)

IBM considers the individual microbial cell as the basic unit of activity, reflecting reality in the model structure. Each cell obeys a number of elementary rules, comprising the cell model. By simulating a large number of such individual cells, it is possible to derive the dynamics of the heterogeneous microbial population on a macroscopic scale.

Knowledge generated within this project has been insufficient in order to set up a microscopic or individual based model for AI-2 related quorum sensing by *S.* Typhimurium within the course of this project. Therefore, Task 3.2 (and related deliverables, tasks and milestones) has not been a prime objective after the second year of the project. Notwithstanding, some methodological developments of IBM (possibly related to other application areas) have been made (see, e.g., Standaert, 2007).

Amongst the developments on IBM is the study on individual parameter identification from macroscopic data. IBMs can only be simulated when individual model parameters and their stochastic properties are known. Opposed to macroscopic variables, cell-based characteristics are much more difficult to determine experimentally. Estimation of individual parameters from macroscopic data stands for an attractive supplement to the time-consuming, labor-intensive, and error-prone individual measurements. Moreover, parameter estimation of IBMs is still a relatively unexplored area but becomes more feasible given incessant progress in computational capacity. As a case study, the individual growth rate and the associated variance have been estimated from macroscopic cell density measurements via least squares estimation. Parameter estimates are obtained based on a grid-search approach. Practical issues such as simulation upscaling, minimizing Monte Carlo variance, choice of the objective function and others, have been resolved. Results show that the average individual growth rate and its associated variance can be estimated uniquely from exponential growth data. The mean individual growth rate is smaller than the population (specific) growth rate. The fastest growing cells determine the macroscopic dynamics. The effect of number of data points and the sample spacing has been investigated. Opposed to population models, parameter estimation uncertainty is less influenced by the number of data points (only large datasets can lower the variance significantly) and grouping of samples (or replaced sampling at the same time instants) lowers the variance.

Task 3.3 Optimal experiment design (OED)

To identify the structure and the parameters of the macroscopic model as well as to select optimal experimental conditions to ensure a maximum information content in the collected data, techniques of Optimal Experiment Design (OED) are exploited during the project. The application of these techniques is not at all new for BioTeC. Especially the field of accurate and efficient model parameter estimation has been a major research topic during the past years.

Given the delay in bioreactor experiments with *Salmonella*, an optimal design methodology for kinetic parameter estimation was first worked out for an alternative case study, namely, the growth and indole-3-acetic acid (IAA) production by *Azospirillum brasilense* Sp245 (in collaboration with CMPG) (Smets et al., 2004, Cappuyns et al., 2006). A prototype model has been developed. Hereto, experimental data from batch bioreactors experiments were available. Problems of model structure discrimination (i.e., selection of a suitable model structure from a set of candidate models) as well as unique parameter estimation were encountered. Optimal fed-batch experiments have been designed for the unique estimation of the growth kinetic model parameters (i.e., Monod parameters) using the methodology of optimal experiment design. Furthermore, optimal experiments have been proposed to discriminate between the two model structures that could be suggested for the IAA production process.

In a next stage, the above-outlined design framework was applied to uniquely and accurately estimate the Monod kinetics of *S.* Typhimurium grown on galactose in LB (which was at that moment expected to be attained kinetics). Based on the minimization of the objective function of the design criterion, an optimal feeding strategy was computed. Experiment designs are shown in Figure 13. The optimized experiment encompasses a batch phase followed by a feeding phase within which the substrate concentration is maintained constant (Figure 13, top). In view of the practical implementation, the optimal feeding strategy was approximated by a constant feeding phase with a feasible substrate concentration (Figure 13, top). The effect of this approximation on the optimality of the design was *a priori* evaluated. Several optimal fed-batch experiments have been realized. As an example two fed-batch experiments are depicted in Figure 13 (middle and bottom plots). Given these informative biomass and substrate data, accurate values for the Monod kinetics could be estimated. The improvement with respect to the parameter estimates obtained from (non-informative) batch experiments is remarkable (see values given in Figure 13). Optimal experiment design for non-linear models is an *iterative* design process starting from an initial guess for the unknown model parameters which is systematically improved within an iterative scheme. A first and second iteration of the design cycle have been described above. A series of optimal design fed-batch experiments have been repeated. Data processing (including parameter estimation) for these experiments confirmed that the original model structure does not fully comprise the growth dynamics.

Fed-batch experiments which were in the first place designed for model parameter estimation (in particular of the Monod kinetics) revealed a major discrepancy between the model structure and the experimental data pointing at a model function error. This clearly shows that (unique and correct) parameter estimation and model structure selection (or discrimination) are closely entwined. Optimal experiment designs are excellent tools to test model validity or check for model structure errors.

In order to elucidate AI-2 quorum sensing in free living *Salmonella* cultures and to identify an appropriate model structure for growth, continuous bioreactor experiments with changing dilution rates were designed (see also Task 4.1). Experiments were designed such that as much information as possible is embedded in one single experiment. During this informative experiment different steady states were reached. Although the experiments were designed on the basis of common sense (and not using some mathematical methodology), the experimental data from the different steady states and transition period between the different dilution rates, made it possible to discriminate between possible model structures and to find one set of model parameters to describe different modes of operation.

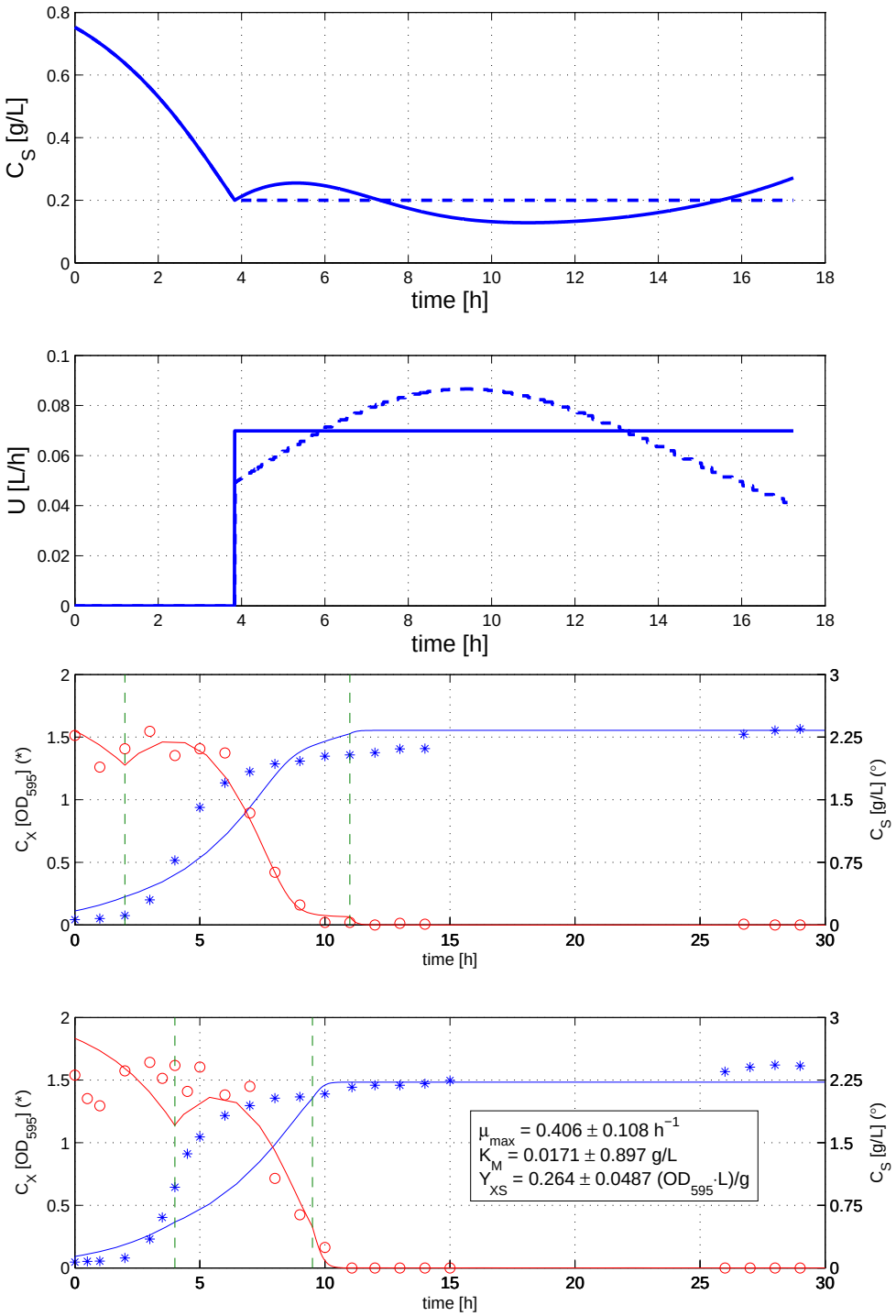


Fig. 13 Optimal experiment design for parameter estimation (top plots) (-: the original design, --: simplified design) and the implementation of the optimized fed-batch experiments with *S.* Typhimurium grown in LB medium supplemented with 0.5% galactose, at constant pH 6 and 37ºC (bottom plots) (-: identified model).

Work package 4: Studying the influence of AI-2 mediated quorum sensing in *Salmonella*

In this work package all the technical methods described in WP1-2-3 were applied to the model system *S.* Typhimurium in order to study the AI-2 type quorum sensing.

In *S.* Typhimurium, AI-2 is produced and released during exponential growth and is subsequently imported into the bacteria via the Lsr (*lux S* regulated) ATP binding cassette (ABC) transporter (Fig. 14). AI-2 induces transcription of the *lsrACDBFGE* operon, the first four genes of which encode the Lsr transport apparatus. LsrK is required for the regulation of the *lsr* operon and the AI-2 uptake process. LsrK is a kinase that phosphorylates AI-2 upon entry into the cell. Phosphorylation of AI-2 results in its sequestration in the cytoplasm. Taga *et al.* (2003) suggest that phospho-AI-2 is the inducer responsible for inactivation of LsrR, the repressor of the *lsr* operon. LsrF and LsrG are necessary for the further processing of phospho-AI-2. Transport and processing of AI-2 could be required for terminating AI-2 signaling or for interfering with AI-2 signaling of other species (Taga and Bassler, 2003).
The biosynthetic pathway leading to AI-2 production is part of the activated methyl cycle (Schauder *et al.*, 2001). AI-2 is produced from *S*-adenosylmethionine (SAM) in at least three enzymatic steps, the last step being catalyzed by LuxS, using *S*-ribosylhomocysteine (SRH, catalysed by Pfs) as a substrate and yielding 4,5-dihydroxy-2,3-pentanedione (DPD) and homocysteine. The unstable DPD spontaneously cyclizes into AI-2. The structural analysis of the *Vibrio harveyi* AI-2–binding protein LuxP in complex with the active AI-2 ligand (Chen *et al.*, 2002) revealed the chemical identity of the *V. harveyi* AI-2 as a furanosyl borate diester. Recently, it was shown that the *S.* Typhimurium AI-2 binding protein recognizes another DPD derivative, i.e. 2-methyl-2,3,3,4-tetrahydroxy-tetrahydrofuran, lacking boron (Miller *et al.*, 2004).

The original project proposal outlined bioreactor experiments to study AI-2 mediated quorum sensing in *Salmonella*. However, in their natural niche, a given bacterial species is often dominated by a consortium of other bacteria. Interactions between bacterial species occur at many levels including competitive/antagonistic interactions as well as cooperative interactions through physical associations in aggregates and biofilms, and intra- and interspecies cell-cell communication through e.g. AI-2. If this is the case, the role of AI-2 must be examined under conditions more closely mimicking those experienced by the bacterium in natural environments. To this end, we decided to split work package 4 into 2 parts, i.e. one dealing with AI-2 mediated quorum sensing in free-living conditions, i.e. bioreactor experiments (A), and one focusing on AI-2 mediated quorum sensing in *Salmonella* biofilms (B).

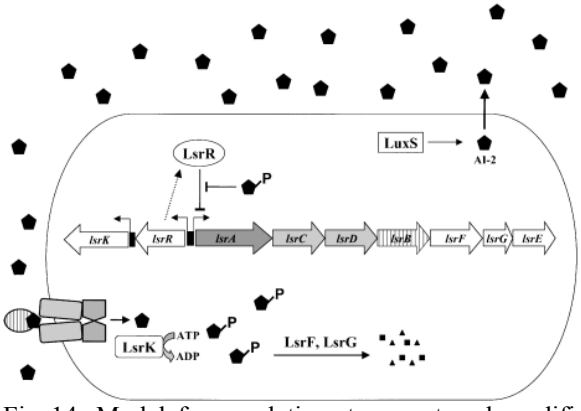


Fig 14. Model for regulation, transport and modification of AI-2 by the Lsr proteins in *S.* Typhimurium. AI-2 (pentagons) is synthesized by LuxS and accumulates extracellularly. Basal expression of the Lsr transporter allows some AI-2 to enter the cytoplasm, where it is phosphorylated by LsrK. Phospho-AI-2 induces transcription of the *lsr* operon by inactivation of LsrR. Production of the Lsr transport apparatus results in further Lsr-dependent internalization of AI-2 and subsequent phosphorylation by LsrK. Phospho-AI-2 is processed by LsrF and LsrG. Taken from Taga *et al.*, 2003.

Task 4.1 Optimizing the growth conditions for maximal expression of AI-2 synthesis and AI-2 dependent pathways in relevant environment

A. *Salmonella* in free-living conditions

Initially, it was planned to perform the bioreactor experiments in LB supplemented with glucose. Indeed, when AI-2 was first described to be produced by *S.* Typhimurium, glucose or other preferred carbohydrates were postulated to be a prerequisite for its production (Surette and Bassler, 1999). Subsequently, since we found that glucose interfered with the AI-2 assay (De Keersmaecker and Vanderleyden, 2003), hampering interpretation of data, we omitted glucose from the medium. In the absence of glucose, AI-2 is still produced, albeit transiently. Additionally, expression of the *lsr* operon is subjected to catabolite repression (Taga *et al.*, 2001; Taga *et al.*, 2003; our data). Finally, since the absence of a carbon source impeded the modelling work, we decided to add galactose to the LB medium as carbon source (as already mentioned in Task 3.1). Galactose does not affect the AI-2 bioassay. The addition of this carbon source in combination with the augmented set point for dissolved oxygen (from 5% DO (initially planned) to 30% DO) has proven to guarantee reproducibility between repeated experiments. Reproducibility is probably guaranteed because of (*i*) the growth on a single growth-limiting or preferential substrate (whereas LB contains a cocktail of candidate substrates, all of which are present in small amounts), and (*ii*) the fact that deviations in the dissolved oxygen concentration do not cross the boundary between (micro-)aerobic and anaerobic conditions initiating switches between metabolic pathways. Furthermore, the set point of 30% DO can be realized more accurately within the bioreactors (data not shown).

Several batch and fed-batch experiments with galactose have been performed. These fed-batch experiments were specifically designed in the context of optimal experiment design for parameter estimation (see Task 3.1). Notwithstanding, the results give an insight in the AI-2 production by *S.* Typhimurium. As an example, Figure 15 shows the biomass, galactose, acetic acid and AI-2 data of two fed-batch experiments. On the basis of these data and all other bioreactor experiments, the following observations can be made.

- Acetic acid is produced during growth on galactose. When galactose is depleted, acetic acid production stagnates and a switch to acetic acid consumption takes place. This acetic acid profile is similar to the profile observed in the presence of glucose. The production of acetic acid during aerobic growth on glucose is well-known for *E. coli* (Kleman and Strohl, 1994). Acetic acid is produced due to the overflow of the oxidative metabolism. The same mechanistic background can be assumed for *S.* Typhimurium grown on galactose.
- Replicate batch experiments in LB with galactose exhibit similar AI-2 courses as batch experiments in LB with glucose (see Fig. 16). AI-2 is produced in the late-exponential phase and disappears when the stationary phase is reached. The AI-2 pattern in fedbatch experiments clearly differs from the batch experiments. Once AI-2 production is initiated, AI-2 production can be maintained for a relatively long period, which coincides with the time during which the population is actively growing (i.e., a persistently high metabolic activity). The disappearance (or intake) of AI-2 seems to coincide with the end of biomass growth (which is similar to the batch experiments). This observation substantiates the finding that AI-2 production is linked to the metabolic potential of *S.* Typhimurium. This was confirmed in the continous culture experiments (see also WP 3).
- An important bottleneck within the research at this moment is the inability of measuring AI-2 quantitatively, and as such to compare AI-2 profiles of different experiments (measured with different bioassays). We tried to use synthetic DPD as a control, however, this approach did not yield satisfactory results for samples taken in bioreactor experiments.

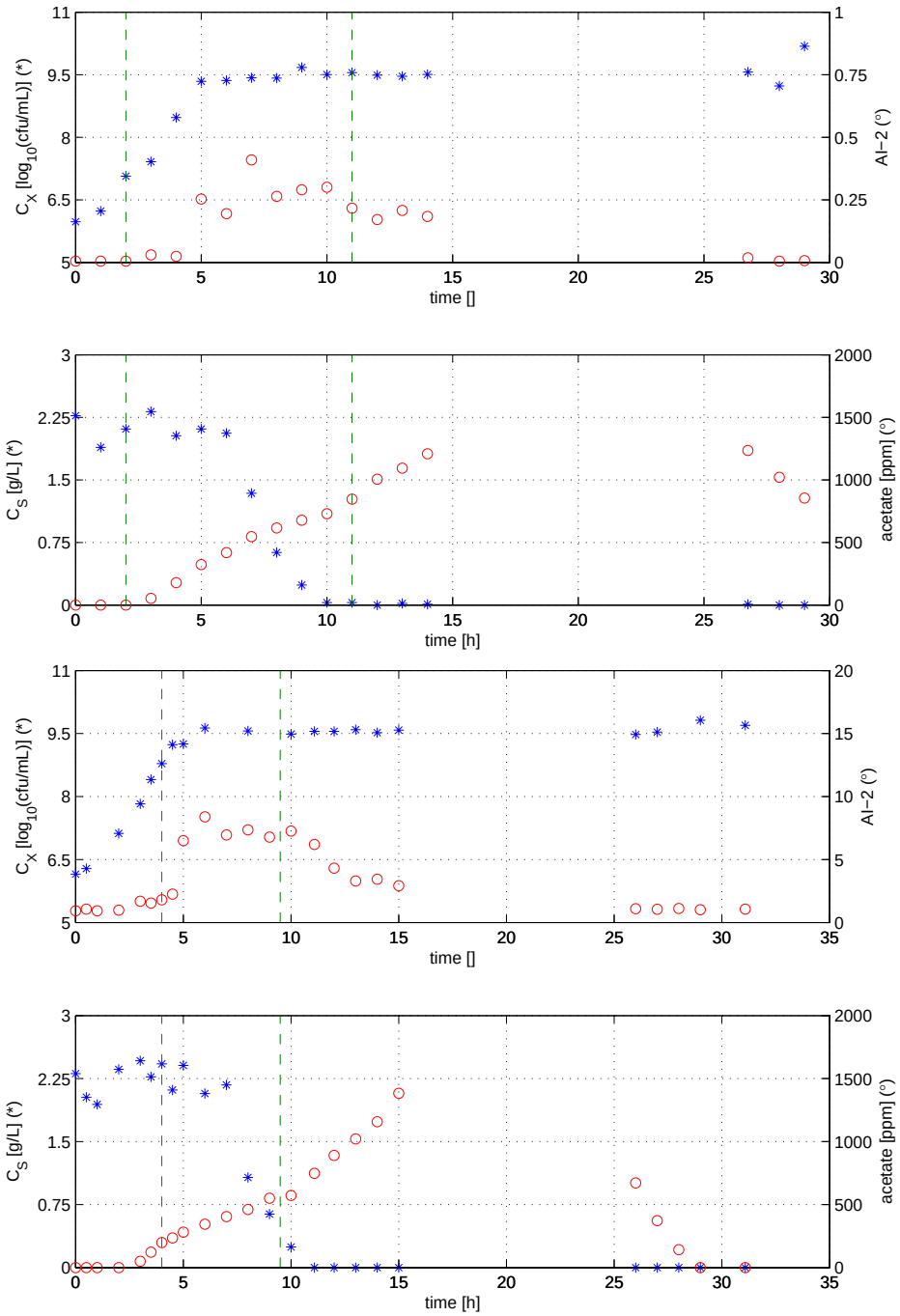


Fig. 15 Fed-batch bioreactor experiments with *S.* Typhimurium grown in LB medium supplemented with 0.5% galactose, at constant pH 6 and 37ºC. The dashed line mark the feeding period (preceded and followed by a batch phase).

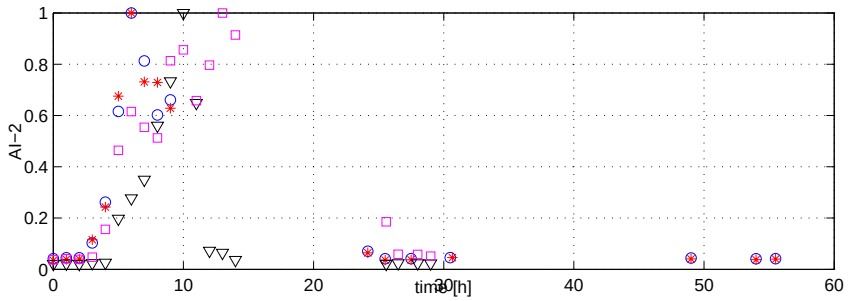


Fig. 16 AI-2 data of four replicate batch bioreactor experiments as shown in Figure 13 (depicted by the four different symbols). (AI-2 values per experiment are divided by the maximum AI-2 value of that experiment.)

B. *Salmonella* biofilms

We adapted and optimized a protocol to study biofilm formation by *Salmonella*. Biofilms are grown on polystyrene, at 16 °C in 1:20 diluted trypticase soy broth (1:20-TSB) (De Keersmaecker *et al.*, 2005). Quantification of biofilm production is performed with crystal violet staining as previously described (Stepanovic *et al.*, 2004), with minor modifications (De Keersmaecker *et al.*, 2005). We are also able to measure AI-2 production during biofilm formation.

Task 4.2 Profiling of the dynamic behavior of* Salmonella *by high-throughput measurements in conditions eliciting AI-2 dependent quorum sensing

We profiled *Salmonella* gene expression both in free-living (planktonic) conditions and in biofilms. We studied differential gene expression in different AI-2 related *Salmonella* mutants, grown in batch, and sampled at a specific time-point. These experiments highlighted the post-translation modification of LuxS (A.1). In addition, we set up a fed-batch experiment where we compared *Salmonella* wild-type with one mutant. For this experiment, we sampled at two time points of the growth phase for gene expression analysis at mRNA and protein level (A. 2). Continuous cultures of wild-type and two mutants were used to refine the AI-2 modeling work (see WP 3) and to sample mRNA and proteins (A.3 and Task 4.4). In *Salmonella* biofilms, we focused on one particular aspect of *luxS*-related biofilm regulation (B).

A. *Salmonella* in free-living conditions

1. Comparison of gene expression in AI-2 related mutants at the protein level

The SQUAD project's major goal was to investigate quorum sensing related pathways in *Salmonella*. Therefore, mutants were constructed in most of the known AI-2 synthesis and target genes: Δ*luxS, luxS::Km, pfs::Cm, lsrB::Cm, lsrK::Cm* and *lsrR::Cm*. These mutants were grown non-agitated at 37°C in erlenmeyers with LB medium supplied with 0.5% glucose and a nitrogen-flushed headspace. These particular conditions corresponded closely to the conditions that were originally chosen to perform our bio-reactor experiments with (LB, 0.5% glucose and 3% DO).

Differences between mutants in AI-2 synthesising genes (*luxS* and *pfs*) and the wild type were only restricted to the mutated genes themselves, so no downstream effect could be detected apart from the downregulation of flagellines. Most differences were detected between a wild type strain and an *lrsR::Cm* mutant. Therefore, LsrR was selected for a ChIP-chip (LsrR::Myc, see also Task 1.5) and transcriptomics analysis (WT versus *lsrR::Cm*). Data integration of the proteomics, transcriptomics and ChIP-chip data will allow determining the direct regulon of the LrsR regulator and consequently the AI-2 mediated pathways (follow-up project, PhD Inge Thijs).

The proteomics analysis of the above described mutants resulted in the identification of different proteins related to quorum sensing: LuxS, LrsF, LsrB and LrsR were retrieved. The knowledge of the position of these proteins on a 2D-gel was valuable for further experiments, since it allowed us to monitor the expression of these proteins in other experiments without the necessity to pick proteins for identification.

To our surprise, 2 protein spots for LuxS were deteceted and this led us to hypothesize that LuxS is post-translationally modified. The novelty of this finding persuaded us to examine this possible modification in more detail. *In silico* calculations showed that the observed shift in pI could possibly be caused by the addition of a phosphate group onto the LuxS protein. Western blots with anti-phosphotyrosine antibodies showed a signal in the region where LuxS was expected, and this suggested the presence of a phosphorylated tyrosine residue in LuxS (Fig. 17). However, repetitions of this experiment failed. Also Western blots of protein lysates that were treated with

YOP tyrosine phosphatase could not clearly show a tyrosine phosphorylation event on LuxS. At the moment, we are using mass spectrometry to detect and pinpoint the modification. Additionally, point mutations of all three tyrosine residues in LuxS were made. However, these mutations showed no evidence of a modified tyrosine amino acid: all mutants still showed the 2 spots for LuxS. The nature and position of this modification is being investigated further in a follow-up project (PhD Gwendoline Kint). However, this finding clearly shows that LuxS is subjected to some previously unknown regulation, which could be related to the dual function it has in the cell.

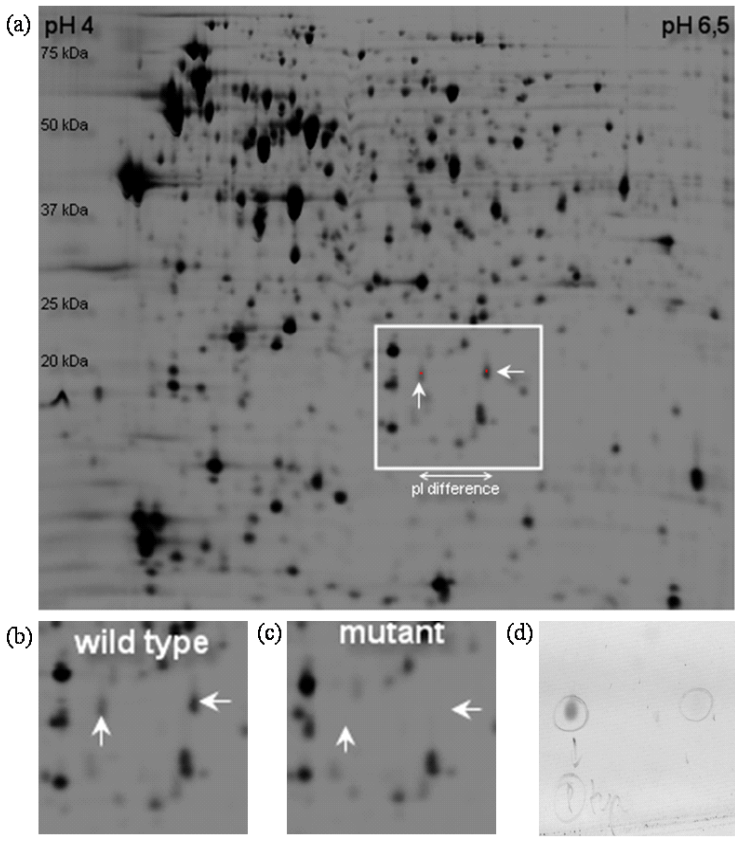


Fig. 17 Comparison of the protein expression of *S.* Typhimurium wild type and *luxS* mutant
In a 2D-DIGE proteomics experiment, protein expression patterns of wild type *S.* Typhimurium SL1344 and its *luxS* deletion mutant were compared. Proteins that were significantly (t-test P-value < 0.01) twofold up-or downregulated between the two strains were assigned for identification. (a) 2D-DIGE image of wild type *S.* Typhimurium; (b) and (c) Enlargement of a section of the 2D-DIGE images of wild type and *luxS* mutant strains; (d) corresponding Western blot section of wild type *S.* Typhimurium detected with anti-phosphotyrosine antibody.

2. Comparison of gene expression in Salmonella *wild-type and* lsrG *mutant*

In a bioreactor batch setup, using LB supplied with galactose and 30% DO at a constant pH, the effect of prolongued AI-2 activity in the cell was evaluated. To this end, a Δ*lsrG* mutant was compared to a wild type strain at two different time point, i.e. 6 hrs post inoculation (before the AI-2 peak) and at 12 hrs post inoculation (after the AI-2 peak), both at the mRNA (transcriptomics) and protein (proteomics) level. Four replicative fed-batch experiments were run (Fig. 18). As one would expect, the factor "time" had a much more pronounced effect than the factor "strain". Comparison of the two strains at 12 hrs post inoculation would theoretically show the differences induced due to a prolongued AI-2 activity. Yet, only a few proteins were differentially expressed and none of them were known AI-2 targets. This could mean that the *lsr* operon was already fully derepressed in the wild type and that the extra molecules of AI-2 had no effect. Another explanation could be that the loss of function of LsrG could have been compensated for by other AI-2 catabolic enzymes (e.g. LrsF and LrsE maybe). A recent publication substantiates this hypothesis, i.e. the reaction catalyzed by LsrG can occur

spontaneously explaining the observation that *lsrFG* double mutants can be partially complemented with either LsrF or LsrG. Indeed, the substrate for LsrF is likely present in *lsrG* mutant strains (Xavier *et al.*, 2007). LsrG catalyzes the cleavage of P-DPD producing 2-phosphoglycolic acid and an unidentified C3 compound. Among the differentially expressed genes, we found genes encoding proteins involved in propanediol utilization and phosphoglycerate transport. Currently, we are investigating if this finding could give a clue on the identity of the C3 compound. Moreover, both in the proteomics and transcriptomics analysis, phase-2 flagella were found to be differentially regulated. This points to new members of the AI-2 mediated regulon. Also some virulence related genes, i.e. belonging to SPI-1 and SPI-2 were differentially regulated. This would suggest a role for AI-2 in *Salmonella* virulence. In the follow-up project where the LrsR regulon will be further determined (see also Task 1.5 and 4.1.1), the link between AI-2 and virulence will be carefully studied. If true, these findings open up possibilities for further drug development.

A)

B)

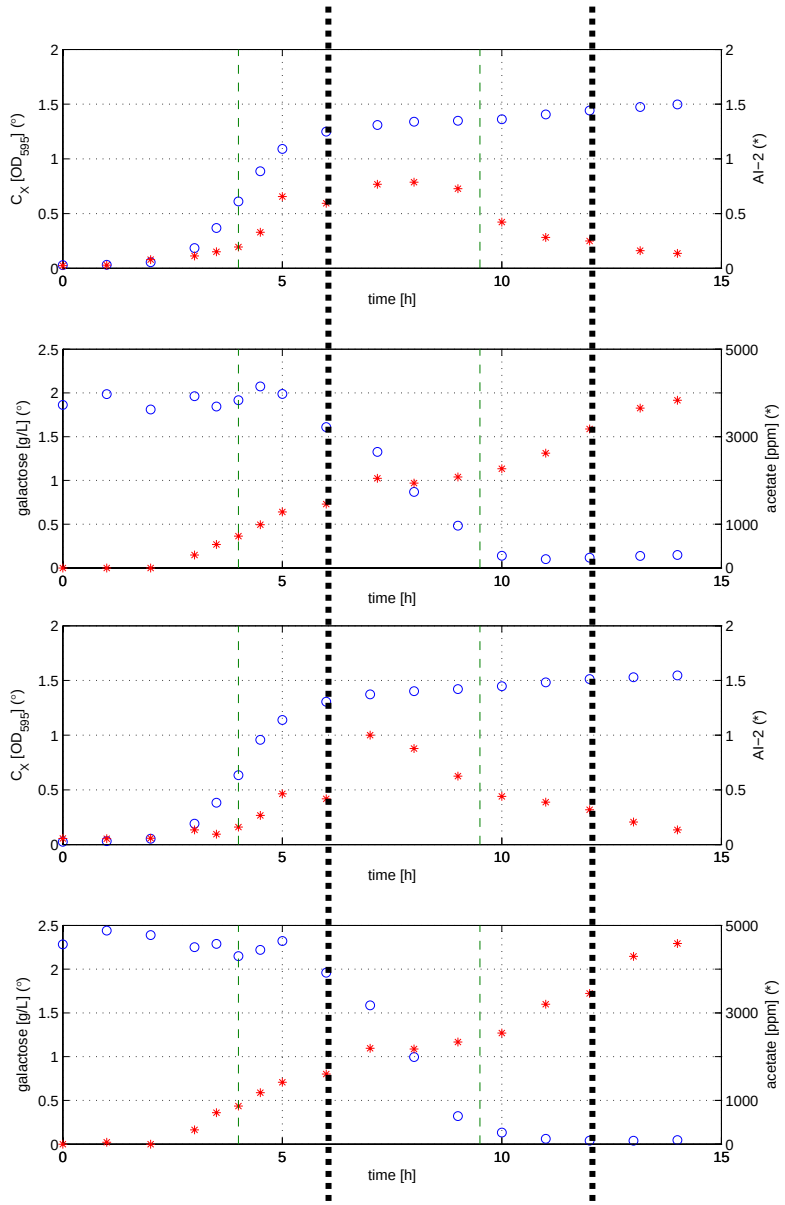


Fig. 18 Average of four replicate fed-batch experiments with *Salmonella* wild-type (A) and *lsrG* mutant (B). The fed-batch phase with constant feeding is marked by the thin dashed vertical lines. Samples for high-throughput expression profiling were taken at 6 and 12 h (bold dashed vertical lines).

3. Comparison of Salmonella *wild-type, Δ*lsrK *and Δ*lsrEFG *mutant*

As elaborated in Task 4.4, continuous culture experiments of *Salmonella* wild-type, Δ*lsrK* and Δ*lsrEFG* have been performed. Samples were taken for protein and RNA isolation. However, already during the course of the experiment, it was obvious that isolation of high-quality RNA (required for the microarray analysis) would be a challenge, since the bacterial cells showed high cell lysis. A high amount of anti-foam reagent needed to be applied. RNA was isolated, unfortunately, it did not pass the quality control at the Microarray Facility. Therefore, it was decided to set-up a batch experiment with the same strains, to perform transcriptomics analysis. The RNA isolated of the bioreactors will be used to validate target genes resulting from this analysis, by RT-PCR. Moreover, if sufficient differentially expressed genes will be found, the proteomics analysis will be performed with proteins isolated from the continuous cultures. These analyses will be continued in a follow-up project.

B. *Salmonella* biofilms

As mentioned above, to study the AI-2 mediated pathway in more *in vivo* relevant conditions, we implemented *Salmonella* biofilm formation in the SQUAD project (originally not in the project proposal).

To investigate the role of *luxS* and DPD in *Salmonella* biofilm formation, we tested *S.* Typhimurium wild type and the *luxS* mutant (CMPG6502) for their ability to form biofilms on polystyrene pegs. The *Salmonella luxS* mutant is impaired in forming mature biofilms. Introduction of a functional *luxS* gene driven by its own promoter (pCMPG5664) in the *luxS* mutant restores biofilm formation to the wild-type level (Fig. 19A).
However, synthetic DPD cannot rescue the biofilm formation defect of the *Salmonella luxS* mutant at 72 µM concentration, which is sufficient to restore *lsrA* expression in a *luxS* mutant (Fig. 19A).
Alternative ways of adding DPD to the biofilm assay were applied in an attempt to rescue biofilm formation in the *luxS* mutant. In none of these cases restoration of biofilm formation in the *luxS* mutant was observed.

To exclude an indirect effect of the activated methyl cycle, in which LuxS plays a role (Schauder *et al.*, 2001; Winzer *et al.*, 2002), as a possible cause of the lack of restoration of biofilm formation by DPD, the biofilm medium was supplemented with methionine, cysteine or *S*-adenosylmethionine. However, no rescue of biofilm formation in the *Salmonella luxS* mutant was observed. Addition of *S*-adenosylmethionine even completely abrogated biofilm formation in *Salmonella* wild type.

To explore the hypothesis that controlled synthesis of AI-2 at a given time point, and therefore tight control of LuxS activity is required for *S.* Typhimurium biofilm formation, as suggested from the previous experiments, we tested whether introducing *luxS* under the control of the strong constitutive *nptII* promoter (Michiels *et al.*, 1991) (pCMPG5643), resulted in restoration of *Salmonella* biofilm formation in the *luxS* mutant. Interestingly, introduction of *luxS* driven by this promoter (pCMPG5643) cannot rescue biofilm formation in the *luxS* mutant (Fig. 19B). Since, on the other hand, wild-type *Salmonella* containing pCMPG5643, is able to form a biofilm in the same amount as wild type (Fig. 19B), it can be further stated that the chromosomal wild-type *luxS* allele is dominant in the *luxS* merodiploid *S.* Typhimurium strain. This experiment also allows excluding possible interference of a particular antibiotic added in the biofilm assay. Chloramphenicol is added when using pCMPG5664 and ampicillin is added when using pCMPG5643.

In conclusion, our results show that *Salmonella* biofilm formation on polystyrene is dependent on an intrinsically regulated *luxS* since a *Salmonella luxS* mutant, which is impaired in biofilm formation, could be rescued by the introduction of a plasmid containing wild-type *luxS*, but not by *luxS* driven by the strong *nptII* promoter. In line with this, addition of synthetic DPD cannot rescue the biofilm formation defect of the *Salmonella luxS* mutant. Although in planktonic state, *S.* Typhimurium *luxS* expression was reported to be weak and constitutive (Beeston and Surette, 2002), during biofilm formation this might not be the case. The sessile lifestyle associated with persistence within a biofilm is distinct from the lifestyles of planktonic culture, resulting in different gene expression (Wang *et al.*, 2004; Beenken *et al.*, 2004). Our results indicate that altering the intrinsic promoter activity of *luxS* and consequently the dynamics of AI-2 synthesis, changes expression, in time and/or intensity, of *Salmonella* genes, among which genes crucial for biofilm formation. The role of LuxS in the activated methyl cycle seems not to be a major player in biofilm formation since supplementation with methionine, cysteine or *S*-adenosylmethionine could not restore biofilm formation in a *Salmonella luxS* mutant, implying the role of AI-2 as a signaling molecule in *Salmonella*. The observation that wild-type *luxS* is dominant in a merodiploid strain containing a wild-type *luxS* allele and a *luxS* allele engineered for strong constitutive expression, further suggests a complex mechanism of AI-2 mediated biofilm formation. These findings were published De Keersmaecker *et al.*, 2005.

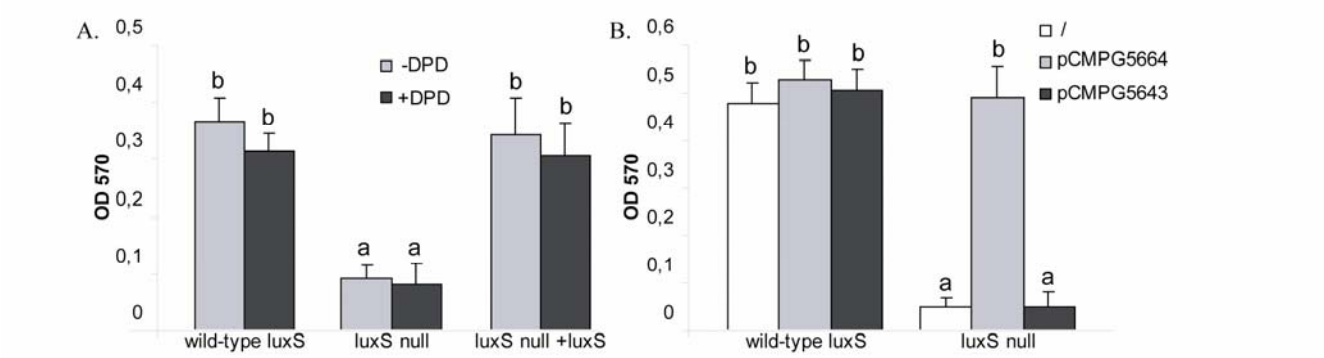


Fig. 19. Biofilm formation of *Salmonella* wild type, *luxS* mutant and *luxS* mutant complemented, in absence and presence of synthetic DPD

The extent of biofilm formation after 48 h is expressed as the absorbance at 570 nm of the crystal violet solution. A. The loss of *luxS* (denoted *luxS* null) significantly reduces biofilm formation on polystyrene pegs compared to biofilm formation by wild-type strains. Introduction of *luxS* driven by its own promoter (pCMPG5664) can rescue biofilm formation. Synthetic DPD was used at an estimated concentration of 72 µM (denoted +DPD, black bars; the negative control is denoted –DPD, grey bars). Addition of DPD cannot restore biofilm formation in the *luxS* mutant. B. Lack of complementation was observed when introducing *luxS* under the control of a strong constitutive promoter (pCMPG5643; black bars). Introduction of pCMPG5664 (grey bars) or pCMPG5643 did not affect biofilm formation by wild-type *Salmonella*. /: no plasmid present (white bars). Error bars depict standard deviations of 8 independent measurements. Means that were found to be significantly different ($P < 0.05$) by the Tukey test are indicated by different letters. Depicted results are representative for at least three independent experiments.

This complexity became even more clear when testing biofilm formation by other AI-2 related mutants (*lsrA*, *lsrR*, *lsrK*, *lsrFGE*, *pfs*, *lsrB*). Only the *lsrFGE* mutants showed a defect in biofilm formation. *lsrFGE* complemented strains are restored in biofilm formation. When we further refined the role of *lsrFGE* in biofilm formation, we could pinpoint *lsrG* as the major effector for biofilm formation. For this reason, the *lsrG* mutant was used in the bioreactor experiment explained in WP 4.2.A.2. Surprisingly, a *luxS* insertion mutant (*luxS*::Km) is still able to form biofilms.

Our results suggest that *Salmonella* biofilm formation on polystyrene is not very much dependent on the AI-2 molecule as such. Indeed, both *luxS* mutants are unable to produce AI-2, however, the insertion mutant can still form biofilms. In addition, a *pfs* mutant can still form biofilms. Pfs

is an enzyme required to synthesize AI-2 using SAM as precursor substrate. Another possibility is that alternative AI-2 producing pathways exist, LuxS or Pfs independent.

However, we found a small regulatory RNA molecule, upstream of the *luxS* gene, i.e. SraD, now renamed MicA. Most known sRNAs are encoded within intergenic regions (Wassarman *et al.*, 2001) and are highly conserved between closely related organisms (Gottesman, 2004b). sRNAs can act to regulate both the synthesis of proteins, by affecting mRNA transcription, translation and stability, and the activity of specific proteins by binding to them (Wassarman, 2002; Masse *et al.*, 2003; Gottesman, 2004a; Gottesman, 2004b). These sRNAs regulate bacterial genes involved in environmental adaptation (Masse *et al.*, 2003). As such, recording the small RNA profile will become as important as recording mRNA expression profiles to understand how cells regulate and modulate gene expression.

More than 60 small RNA-encoding genes have recently been identified in *E. coli* (Argaman *et al.*, 2001; Vogel *et al.*, 2003). The biological roles of the majority of these RNAs are as yet unknown. One of these sRNAs is encoded by the *micA* gene. This gene is well conserved in enterobacteria and encodes a ~60 nt RNA that is highly upregulated at the onset of stationary phase in *E. coli*. The gene is neighboured by *luxS* on the opposite strand. Recently, a role for sRNA in quorum sensing in *Vibrio* has been demonstrated (Lenz *et al.*, 2004).

We optimized a Northern analysis method to detect sRNAs in *Salmonella*. We investigated MicA production under different conditions (free-living vs. biofilm) and in different genetic backgrounds (WT, Δ*luxS*, *luxS*::Km etc.). It was clear that *micA* is highly expressed under biofilm conditions.

A first indication that a sRNA is involved in *Salmonella* biofilm formation was obtained by the observation that a *Salmonella hfq* mutant is impaired in biofilm formation. We could complement the phenotype by introducing a functional *hfq* gene. The RNA chaperone, Hfq, plays a diverse role in bacterial physiology beyond its original role as a host factor required for replication of Qbeta RNA bacteriophage. Recently, it was shown that Hfq is involved in the regulation of motility and outer membrane protein (OmpD) expression in addition to invasion and intracellular growth (Sittka *et al.*, 2007), which makes Hfq essential for *Salmonella* virulence. The role of Hfq in *Salmonella* biofilm formation was until now unknown. A manuscript describing these results is in preparation.

We made different complementation constructs (only sRNA, *luxS* with own promoter, i.e. with sRNA; *luxS* without sRNA etc.) to investigate the effect of these complementation vectors on *Salmonella* wild-type and the two *luxS* mutants. However, the role of MicA in *Salmonella* biofilm formation turns out not to be that straightforward. Additional experiments were needed to draw up a model. To this end, we isolated RNA from WT, Δ*luxS* and *luxS*::Km biofilms to disentangle the role of this sRNA by a genome-wide approach. The RNA samples were delivered to the MAF for labelling and hybridization. Data were preprocessed (Loess fit) and differentially expressed genes were defined using SAM, LIMMA and t-tests. Only 25 differentially expressed genes are shared between Δ*luxS* and *luxS*::Km. This already suggests major differences between genes differentially expressed between WT and *luxS*::Km and WT and Δ*luxS*, confirming that other elements than merely AI-2 (both Δ*luxS* and *luxS*::Km are deficient in AI-2 and yet their gene expression is so different) are regulating biofilm formation.
Δ*luxS* is deficient in biofilm formation and this is clearly reflected in the gene expression since genes known to be important for biofilm formation are downregulated (e.g. *csgD* (master regulator of *Salmonella* biofilm formation), *csgA*, *csgF*, *glyA*, *bcsCZBA*). The list of differentially expressed genes contains several genes with unknown function. It would be interesting to further characterize these genes, based on the biofilm phenotype. The information will be highly

valuable in the follow-up project where genes important for *Salmonella* biofilm formation will be determined.

The same biofilm samples were analysed by proteomics. Therefore some technical challenges had to be overcome. We adapted a protocol for protein-extraction of free living cultures to a protocol suitable for the extraction of proteins out of a biofilm. We used the same experimental set-up as for the microarray analysis i.e. we extracted proteins of a biofilm of 24h.
In the 2D-DIGE analysis 54 proteins were found to be significantly up or downregulated. 13 of these could be identified by MS-MS. Among these were the proteins LsrB and LsrF, which were downregulated in both *luxS*-mutants compared to the wildtype. This again indicates that as is the case for *luxS*, the *lsr* operon is not involved in the biofilm forming behavior of *Salmonella*: the expression level of the Lsr proteins is not related to the fact wether a biofilm is formed or not. Another interesting finding is the fact that OmpA was upregulated in the *ΔluxS* mutant. Very recently, it was reported that *ompA* expression is influenced by MicA (Figueroa-Bossi *et al.*, 2006; Papenfort *et al.*, 2006), however the link with *Salmonella* biofilm formation is not known yet. By constructing a *Salmonella ompA* mutant, we could demonstrate that OmpA is important for *Salmonella* biofilm formation (manuscript in preparation).

To further unravel the role of MicA in *Salmonella* biofilm formation and in *Salmonella* regulatory gene expression, we constructed additional mutants, i.e. a *micA* deletion mutant and *micA* overexpression mutant. Alternatively, we designed a *micA* antisense construct. These mutant strains were used to determine gene expression of specific genes (i.e. *luxS*, *yijC*, *metC* etc. (see below)) and to investigate biofilm formation. Recently, we started a collaboration with Dr. J. Vogel of the Max-Planck Institute for Infection Biology (Berlin, Germany) on the role of *micA* in *luxS* expression. This study will be continued in the follow-up project 'FWO-postdoctoral mandate S. De Keersmaecker'. Nevertheless, it is already clear that MicA is important in *Salmonella* biofilm formation and this partly through one of its targets, *ompA* (manuscript in preparation). However, we have strong indications that other targets of MicA are also involved (see also Task 4.3).

Task 4.3. Identification of AI-2 synthesis and AI-2 dependent pathways

From the microarray and proteomics experiments, some new putative members of the LuxS-related regulon have been identified (see Task 4.2). These will be further validated in follow-up projects. Especially the virulence related factors (e.g. phase-2 flagellin) are interesting targets for future valorization of the project results. Moreover, the involvement of a small non-coding regulatory RNA molecule in *Salmonella* biofilm formation (see Task 4.2.B) clearly illustrates the impact and importance of these previously somehow neglected regulatory factors in bacterial cells. Targetting these molecules could be an alternative way of interfering with *Salmonella* virulence and/or biofilm formation.

Since we found that a sRNA upstream of *luxS* is involved in *Salmonella* biofilm formation, it was interesting to look for additional targets of this regulatory molecule. In a preliminary *in silico* screen (see WP5) we searched for putative targets of the sRNA. One of the hits was the putative transcriptional repressor *yijC*, belonging to TetR/AcrR family. A *yijC* mutant appeared to form less biofilm than the *Salmonella* wild type. We isolated RNA from a *yijC* mutant to be compared to the WT, both grown in biofilm. The RNA was labeled and the cDNA hybridized to the *Salmonella* microarray. Data were preprocessed (Loess fit) and differentially expressed genes were defined using SAM, LIMMA and t-tests. Among the differentially expressed genes, we find virulence related genes and genes involved in biofilm formation.
Additionally, YijC was tagged with Myc, which allowed ChIP-chip experiments to define target genes of this regulator. As already mentioned for LrsR, we are using tiling arrays for this analysis (analysis still ongoing). A combination of the expression data and the location (ChIP-chip) data

will be a powerful means to gain insight in the function of YijC in *Salmonella*, which until now is still unknown.

Another potential target of MicA found in the *in silico* screen, is *metC*, encoding a cystathionine beta-lyase. The biosynthesis of methionine in bacteria requires the mobilization of sulfur from Cys by the formation and degradation of cystathionine. MetC catalyzes the breakdown of cystathionine to homocysteine, the penultimate step in methionine biosynthesis. An insertional inactivation of *metC* in *S.* Typhimurium resulted in the attenuation of virulence in a mouse model of systemic infection (Ejim *et al.*, 2004). This result confirms a previous chemical validation of the Met biosynthetic pathway as a target for the development of antibacterial agents and demonstrates that cystathionine beta-lyase is important for bacterial virulence. A *metC* mutant is Met auxotroph (Ejim *et al.*, 2004). Interestingly, we found that introduction of extra copies of *micA* (on a plasmid) in *Salmonella* WT, resulted in Met auxotrophy as well. We could confirm this using RT-PCR (Fig 20). We constructed a *metC* mutant. A *metC* mutant is affected in biofilm formation. However, the biofilm formation by a complemented mutant (introduction of *metC* gene on plasmid) was not completely restored to wild-type levels. This implies that secondary mutations could have been arisen or that the intrinsic regulation of *metC* expression is important in biofilm formation.

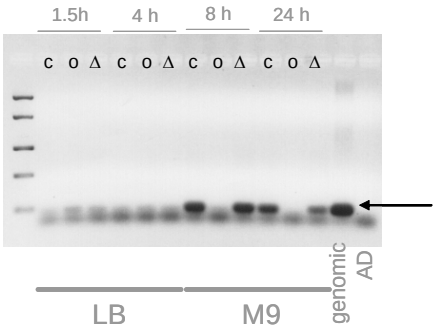


Fig. 20 RT-PCR analysis of *metC* expression in *Salmonella* strains c (control), o (overexpression of *micA*) and D (depletion of *micA*) grown in LB or M9 minimal medium, at different time points. Genomic DNA and destilled water (AD) were taken as positive and negative control, respectively.

Task 4.4. Additional experiments to validate the network structure

In order to further unravel the AI-2-dependent quorum sensing, a series of continuous bioreactor experiments was performed. Under steady state conditions, the dilution rate is equal to the specific growth rate. In contrast to batch or fed-batch experiment, the effect of the growth rate can be (more or less) studied independently of the cell concentration. AI-2 production prevailed at high specific growth rates while AI-2 uptake will dominated when the specific growth rate became (very) low. Continuous bioreactor experiments during which the dilution rate (and thus the specific growth rate) switches from *(i)* a low to a high value, and *(ii)* a high to a low value, therefore gave relevant information on the AI-2 production/uptake and AI-2 dependent pathways.
In a trial period, the above-outlined informative experiment was implemented for the wild type strain measuring biomass, galactose, acetic acid and AI-2. The results were used for constructing the macroscopic model (see also Task 3.1).
Within the experiment, four phases occur: (i) an initial batch phase, (ii) a continuous phase with low feed rate (0.25 L/h), (iii) a continuous phase with high feed rate (0.50 L/h), and (iv) a final batch phase. The initial batch phase is introduced to attain a high cell density to start the continuous phase and to start with an exponentially growing culture. The initial galactose concentration was equal to $C_S(0)$=1.5 g/L and the galactose content of the feeding solution was taken C_{SIN}=10 g/L (dissolved in fresh LB). During the final batch phase, the remaining galactose and, subsquently, acetic acid are consumed until depletion of acetic acid. As expected (based on literature reports), a steady-state is attained after one to two residence times (defined as the ratio

of the volume and the feed rate). The residence time gives an idea of the average time a particle remains in the bioreactor, being 20 and 10 hours for steady-states I and II, respectively. Under steady state conditions, all components in the system (i.e., inside and outside the cell) remain constant in time. Hence, a steady-state can be assumed when all measured variables remain constant in time (biomass, galactose, acetate, AI-2, base feeding). In that case, the specific growth rate of the culture equals the dilution rate (defined as the inverse of the residence time). For the subsequent phases, the specific growth rate of (approximately) is 0.05 h^{-1} and 0.1 h^{-1}, respectively.
Both metabolic states correspond with different levels of AI-2 where the low growth rate yield a low level of AI-2 (steady-state I) and the high growth rate corresponds with a high(er) level of AI-2 (steady-state II).

Based on this trial (and some other experiments), an *optimal* experiment was outlined with the following specifications. A transition period of at least three times residence time was adopted to guarantee steady-state. The steady-states were maintained for two days during which several samples were taken for high-throughput analysis (mRNA and proteins). Instead of two steady states, three steady states were realized, i.e, low dilution rate – high dilution rate – low dilution rate. The switch from a high to a low specific growth rate would induce an augmented AI-2 uptake.
Three informative continuous experiments were run namely, with WT, Δ*lsrFGE* and *ΔlsrK* mutants (Fig. 21). The latter was a change in initial experimental design, i.e. we planned to use a *luxS* mutant, i.e. a strain where AI-2 production would not be initiated. However, given the difference in biofilm phenotype of the two *luxS* mutants (Δ*luxS* and *luxS*::Km (see Task 4.2), we thought it would be less riskful to use a *ΔlsrK* mutant which eventually would result in the same effect as with a *luxS* mutant. Indeed, in a *lsrK* mutant no AI-2-dependent gene expression is expected since AI-2 cannot be phosphorylated. Additionally, in order to knock down all AI-2-processing, we used a *ΔlsrFGE* mutant instead of a *ΔlsrG* mutant. The Δ*lsrFGE* mutant is defective in the degradation of phosphorylated AI-2.

First a practical hurdle had to be tackled, i.e. the sterilization of 100 l of LB medium. This turned out not to be straightforward since we have suffered from contaminations in the reactor vessel. However, finally we managed to complete three continuous reactor experiments. Samples were taken for RNA and protein isolation.
We opted to start with microarray analysis, since this would allow us to have a global view on what is happening in the different steady states. Indeed, the proteomics analysis is limited by the number of spots that can be identified; this in contrast to the microarray, where the expression level of thousand genes can be determined simultaneously. We isolated RNA from the different samples and the RNA was delivered to the MAF. However, the quality of the RNA turned out to be insufficient to continue with the microarray experiment. It was already clear during the culturing that cell lysis occurred frequently which would have an effect on downstream handling. Therefore, it was decided to set-up an experiment comparing a WT, Δ*lsrFGE* and *ΔlsrK* mutant grown in erlenmeyers to perform microarray analysis. Genes differentially expressed under these conditions would be studied by low-througput techniques (RT-PCR) using the RNA of the continuous cultures. Moreveover, since we expect that the protein samples will have suffered less from the cell lysis (bacterial proteins are more stable than bacterial RNA), the proteomics analysis of the reactor samples will give additional information on the differential expression in the continous culture (see also Task 4.2.A.3).

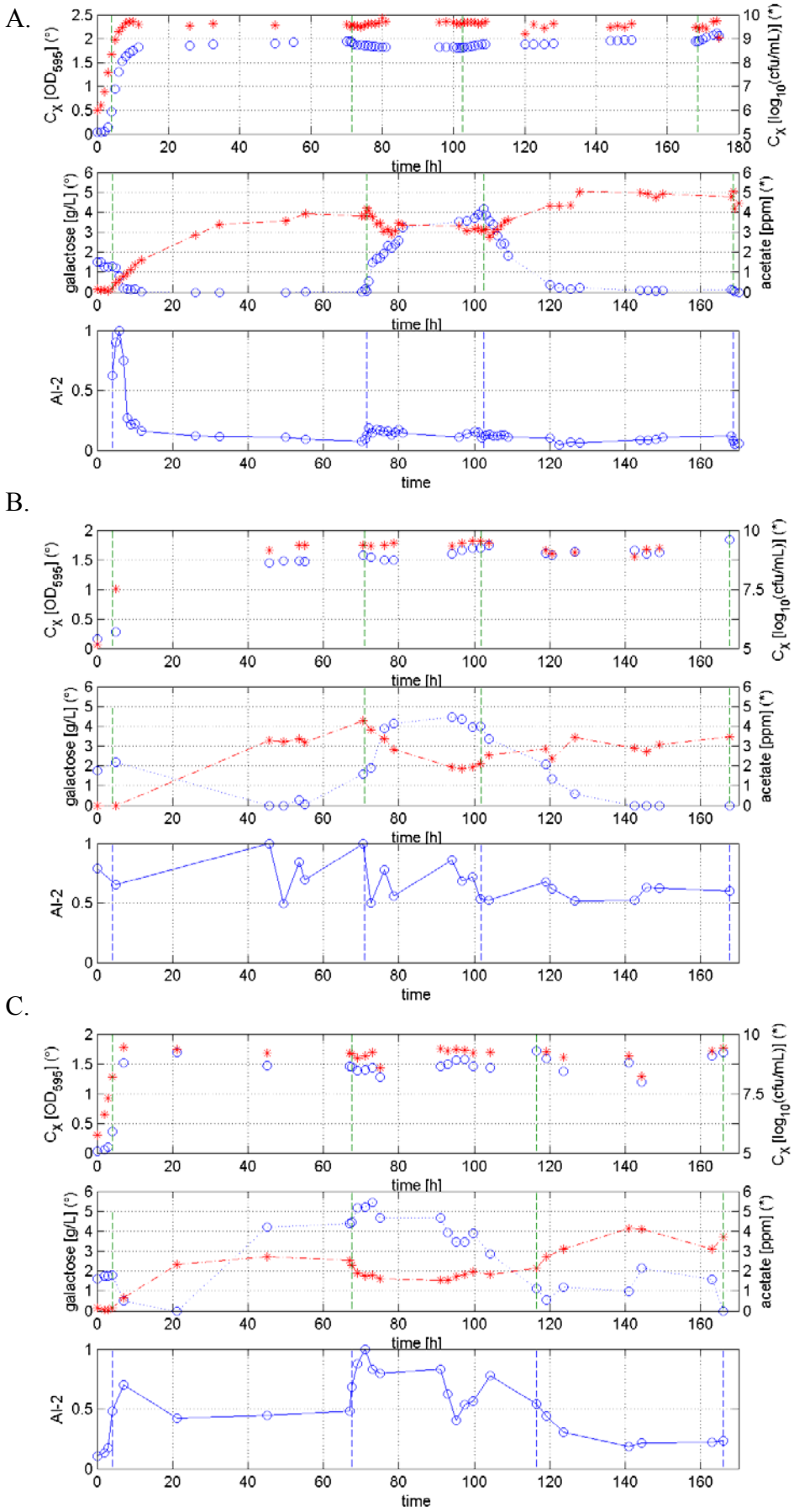


Fig. 21 Continuous culture experiment of WT (A), *ΔlsrK* (B) and *ΔlsrFGE*, with OD, CFU, galactose, acetate and AI-2 measurements.

Task 4.5. Building a single cell model of the AI-2 dependent cell behavior

Knowledge and data generated thus far are insufficient to set up a microscopic or individual based model for AI-2 related quorum sensing by *S.* Typhimurium. Besides Task 3.2, Task 4.5 can therefore not be accomplished within the timeframe of the project (deliverable is printed in italic). Note, however, that the simple structured model described under Task 3.1 already quite accurately describes the AI-2 profiles. When more results become available from microarray analyses and proteomics analyses, this knowledge can be built into the macroscopic model or can be used to develop a microscopic model.

Task 4.6. Identification of additional targets of AI-2-signaling pathway

The high-throughput experiments performed during the SQUAD project have revealed some additional targets of the AI-2/LuxS mediated pathway (virulence genes, small regulatory RNA, …). However, to determine whether these targets are directly regulated by AI-2, we should determine the LsrR regulon. Recently, *Salmonella* whole genome microarrays became available which allows this kind of analysis. ChIP-chip and transcriptomics assays, comparing gene expressing in a wildtype and *lsrR* mutant strain, have been performed to define downstream target genes of LsrR. We are currently awaiting results of the ChIP samples that are being hybridized to *Salmonella* whole genome tiling arrays at NimbleGen Systems, Inc. These results will be validated in a follow-up project. We will integrate this knowledge with the information obtained in the experiments described in Task 4.4, resulting in a global view on AI-2 related gene expression in *Salmonella*. This model will allow extracting putative drug targets to be explored and valorizised in follow-up projects.

Work package 5: Identification of *Salmonella* specific drug targets by *in silico* genomics

Parallel to the experimental approach, a second approach based on genomics, phylogenetic footprinting and phylogeny was followed.

Task 5.1. Motif finding and genome wide scanning

The optimised analysis flows of task 2.3 have been used to study

1) the PmrA regulon of *S.* Typhimurium as a proof of concept

K. Marchal, S. De Keersmaecker, P. Monsieurs, N. Van Boxel, K. Lemmens, G. Thijs, J. Vanderleyden, and B. De Moor. *In silico* identification and experimental validation of PmrAB targets in *Salmonella* Typhimurium by regulatory motif detection. Genome Biology, 2004, 5:R9.1-R9.20, Epub

2) to study the HilA regulon based on a preliminary microarray dataset

S.C. De Keersmaecker, K. Marchal, T.L. Verhoeven, K. Engelen, J. Vanderleyden, and C.S. Detweiler. Microarray analysis and motif detection reveal new targets of the *Salmonella enterica* serovar Typhimurium HilA regulatory protein, including *hilA* itself. Journal of Bacteriology, 2005, 187:4381-91

3) to study the HilA regulon based on a combination of a microarray and a ChIP-Chip dataset.

I.M. Thijs, S.C. De Keersmaecker, A. Fadda, K. Engelen, H. Zhao, M. McClelland, K. Marchal and J. Vanderleyden. Delineation of the *Salmonella* Typhimurium HilA regulon through genome-wide location and transcript analysis. Journal of Bacteriology, Accepted.

Based on literature data (Taga *et al.*, 2001; Taga *et al.*, 2003), we screened the *lsrR* (*ydeW*) intergenic region for possible motifs. These motifs were detected by screening the intergenic of *ydeW* and its orthologs in *E. coli*, *Yersinia* and *Shigella*. The screening resulted in several motifs. The two most likely candidate motifs (highest log-likelihood) were used to screen all intergenic regions of the *Salmonella* genome (Fig.22).

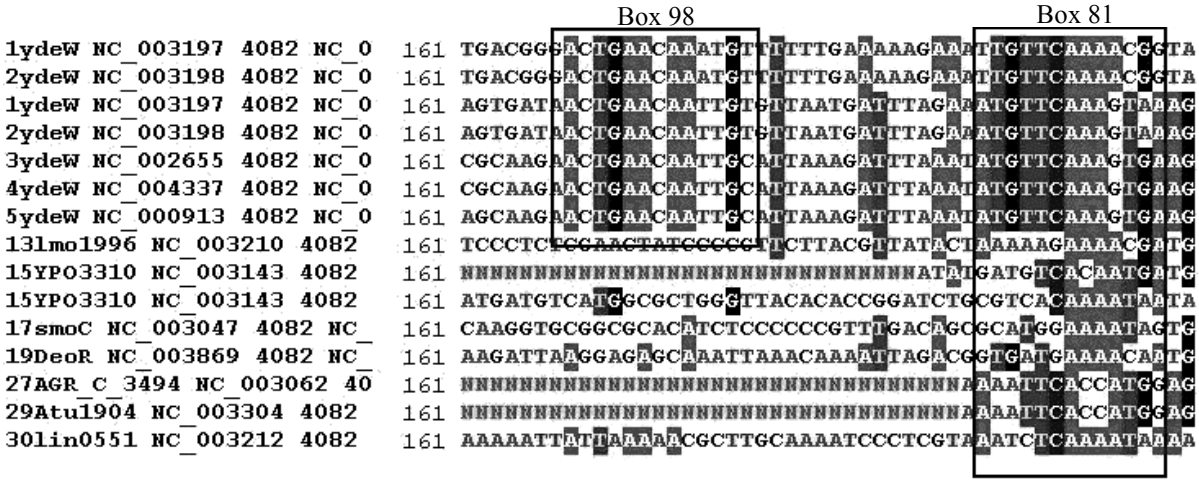


Fig. 22 Alignment intergenics of *ydeW* and orthologs. *S.* Typhimurium (NC_003197), *S. typhi* (NC_003198), *E. coli* (NC_002655, NC_000913) and *S. flexneri* (NC_004337). Not conserved in *Listeria monocytogenes* (NC_003210), *Y. pestis* (NC_003143), *Sinorhizobium meliloti* (NC_003047), *Thermoanaerobacter tengcongensis* (NC_003869), *A. tumefaciens* (NC_003062, NC_003304) and *Listeria innocua* (NC_003212). The two most likely candidate motifs are boxed (Box 98 and Box 81).

Box 81 is present once and has a consensus of ATGTTCAAAGTr, box 98 has a consensus of wCTGAACAAwTG. This motif is repeated three times in the intergenics of the *ydeW* orthologs. Although box 81 showed the best alginment, the analysis was also perfomed for box 98 since box 98 is always followed by box 81 or by a sequence with high similarity to box 81. This tandem-appearance seemed interesting to us, especially since we observed that in *E. coli* and *S. flexneri* the box 98 is present twice in the *ydeW* intergenics, but was also followed by a box 81-like motif.

Using Gibbs sampling (Thijs *et al.*, 2002), two motif models (probability representation of the consensus DNA pattern) were constructed that corresponded to box 81 and box 98. These motifs were then used to perform a genome-wide screening of the *Salmonella* intergenics. Since the motifs are based solely on the *ydeW* intergenic sequences, the screening was expected to have a low sensitivity. To counteract this, a low threshold for the minimal score was chosen. This lower threshold resulted in an increase in the number of false positives. To reduce this number of false positives, we performed phylogenetic footprinting (see Task 5.2).

Task 5.2. Phylogenetic footprinting of the identified motifs

Motif sampling with the two motif models in combination with phylogenetic footprinting resulted in 45 hits for box 81 and 24 hits for box 98. Among the interesting target genes were *uvrB* and *slyA*. *uvrB* is already shown to be LuxS regulated in *Porphyromonas gingivalis* (Chung *et al.*, 2001) and *Actinobacillus actinomycetemcomitans* (Fong *et al.*, 2001). We found a footprint (box 81) in *E. coli* and *Shigella*, but less conserved in *Yersinia*. The UvrB protein is a DNA repair enzyme that catalyzes the excision of UV-damaged nucleotide segments. Therefore, LuxS may also play a role in stress response in *S.* Typhimurium. *slyA* (salmolysin) encodes a transcriptional regulator for hemolysin. Hemolysins are bacterial toxins that function by assembling identical subunits into a membrane-spanning pore. SlyA may play a role in the regulatory system that enables *Salmonella* to adapt to the intracellular environment (Stapleton *et al.*, 2002). It has been shown that hemolysin production in *V. vulnificus* is regulated by LuxS, i.e. a *luxS* mutation

caused a two to three-fold increase of hemolysin production (Kim *et al.*, 2003). A footprint (box 98) was found in all species included in the analysis. Another interesting hit is *hfq* since both box 81 and box 98 were found in its intergenic. *hfq* encodes host factor 1 (HF-1), required for efficient expression of *rpoS* in *S.* Typhimurium. HF-I is a small, heat-stable, site-specific RNA-binding protein originally characterized for its role in replication of the RNA bacteriophage Q beta of *E. coli* (Brown and Elliott, 1996) (see also WP 4). Given the role of an sRNA in the AI-2 mediated pathway, this finding is very interesting. We detected a footprint in *Salmonella*, *E. coli* and *Shigella*, but not in the intergenic of its *Yersinia* ortholog. The same holds true for a number of other footprints we detected. This could indicate that there are significant differences between LuxS dependent regulons in *Yersinia* and other species included in this study. Although a *luxS* gene is present in the *Yersinia* genome (Bassler *et al.*, 1997; Bassler, 1999), no AI-2 production or AI-2 regulated genes of *Yersinia* have been reported yet.

We applied the phylogenetic footprinting strategy at genomic level to compile a compendium of possible motifs in *S.* Typhimurium (see application of tool developed in WP2.3). As such, a generic inventory of putative motifs in *S.* Typhimurium and closely related species was obtained. This will be very useful in the search for motifs important in AI-2-related regulatory networks: indeed using the integrative network inference algorithms developed in WP2.2 these motifs will be combined with the microarray data produced in WP 4 to infer AI-2 dependent pathways. This work will be continued in follow up projects (GOA).

Task 5.3. Studying the evolutionary aspect of the spread of quorum sensing

To study the evolutionary aspects of signal pathways the following tools were developed:

- Automatic identification of orthologs based on the definition of the reciprocal best blast hits or based on the definition of COGs
- Automatic operon prediction
- Scripts to compile a database of species specific normalised microarray expression data from the public databases SMD (Stanford microarraydatabase, ncbi omnibus (http://www.ncbi.nlm.nih.gov/geo/), EBI ArrayExpress (http://www.ebi.ac.uk/arrayexpress/).

This information and tools will be useful when comparing the degree of conservation between pathways. As a case study we compared to what extent the PhoPQ regulon was conserved between *E. coli* and *S.* Typhimurium.

P. Monsieurs, S. De Keersmaecker, W. W. Navarre, M. W. Bader, F. De Smet, M. McClelland, F. C. Fang, B. De Moor, J. Vanderleyden, and K. Marchal. Comparison of the PhoPQ regulon in *Escherichia coli* and *Salmonella* Typhimurium. Journal of Molecular Evolution, 2005, 60:462-74.

Previous cases mainly served as benchmark studies to demonstrate the power of our analysis flows developed in tasks 5.1, 5.2 and 5.3. Recently, we applied the analysis flows on a quorum sensing related task. Experiments performed in WP4 pointed towards a transeffect of the intergenic region of *luxS* in biofilm formation. Closer inspection indeed showed the presence of a sRNA *micA* located upstream of *luxS*. As further experiments clearly showed the role of *micA* in the quorum sensing related pathway, we were interested in finding the targets of this sRNA by means of our analysis flow. Indeed, *in silico* prediction of microRNA targets has successfully been applied in eukaryotes. These predictions are based on the sequence complementarity between the microRNA and its corresponding mRNA target. As a proof of concept, we developed a similar basic strategy to find additional targets of the *S.* Typhimurium micA sRNA. The used analysis flow makes use of :

1) a phylogenetic footprinting (Task 5.2) step to identify regions, conserved between closely related *micA* orthologs. These conserved regions (see Fig. 23) were assumed to be the preferential interaction regions for the interaction between the sRNA and its cognate targets.

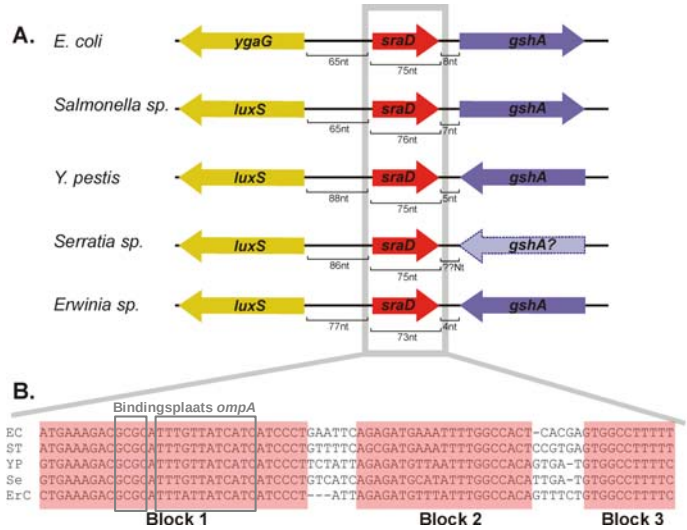


Fig. 23. Phylogenetic footprinting of *luxS-micA* (*sraD*) region in *Salmonella* and closely related species.

2) In a second step, the conserved regions were used to screen on a genome-wide level all intergenic sequences of *S.* Typhimurium (Task 5.1). A few high scoring sequences were selected as potential *micA* targets.
3) In a third step, we assessed whether the potential interaction site within these newly predicted targets could be phylogentically footprinted in the orthologs of these targets in closely related species that also contained the *micA* target (Task 5.2). If a phylogenetic footprint was detected, the confidence in our prediction was enhanced.

One promising target was selected for further validation and showed to be indeed directly dependent on *micA* (see Task 4.3). Further characterisation of the role of this sRNA in this intricate quorum sensing dependent network will be the subject of a follow up project (Postdoc FWO S. De Keersmaecker).

In the mean time, some other studies also used *in silico* approaches to identify sRNA targets (Tjaden et al., 2006; Tjaden, 2007). This, together with our preliminary results shows that sRNA targets can be identified by *in silico* approaches. In the future we will further exploit the potential of these procedures to unravel the role of sRNAs in virulence.

1.3. Overall conclusions

The achievements of the SQUAD-projects can be divided in two major parts, corresponding to technological and a scientific objectives, i.e. i) the development of tools in the different disciplines (molecular biological tools, bioinformatics tools, and bacterial population and single cell modeling) and ii) insight in the AI-2 related quorum sensing system of *S.* Typhimurium.

Molecular biological tools that have been developed during the SQUAD-project are:

- Molecular biological techniques related to *Salmonella* (mutant construction, reporter fusions…)
- *Salmonella* microarrays
- 2D-DIGE proteomics
- Tools to investigate the phosphoproteome
- ChIP-chip technology (cDNA arrays and whole-genome arrays)
- Technologies to study small regulatory RNA molecules
- Experimental set-ups to study *Salmonella* biofilm formation and gene expression in biofilms (mRNA and proteins)

Each of these tools have been implemented and validated by means of test cases, which all have been published (or submitted) to international peer-reviewed journals (see Part III of this report). Moreover, these tools are the basis of other research projects (not necessarily *Salmonella* related). Therefore, additional output is to be expected.

Concerning the bioinformatics tools, the SQUAD project offered considerable funding that enabled us to develop our expertise in the bioinformatics part of microbial systems biology. We obtained state-of-the-art expertise in microarray data processing, motif detection (including a *Salmonella* motif compendium) and set the path towards data integration for microbial network reconstruction. The goal is definitely to further valorize this algorithmic expertise (for instance, by means of service providing). This will, however, require more intensive focusing on the development of algorithms and visualization software. This will be the focus of the follow up project “BioFrame”.

The SQUAD-project also allowed the development of bioreactor-tools for *Salmonella*. Batch, fed-batch and continuous-continuous experiments of *Salmonella* grown on galactose have been run and allowed successfully the construction of a semi-structured (semi-mechanistic) macroscopic mass balance type model, containing five mass balance equations for biomass, galactose, storage carbohydrate, acetic acid and AI-2. Validation of the model on newly generated data yields promising results. However, parameter estimation remains impossible due to the lack of a quantitative measurement for AI-2. Given the present knowledge, it was unfortunately not possible to build a microscopic individual based model (IBM) for AI-2 related quorum sensing by *S.* Typhimurium within the course of this project. Nevertheless, the methodological development of a microscopic individual based model (IBM) was performed. Additionally, informative continuous culture experiments were designed and optimized. Also these technologies will be useful in other research projects.

There was a longer lag-phase in the generation of results in the SQUAD project because of some practical hurdles that had to be dealt with, e.g. move of BioTeC lab to another building, installation and implementation of autoclave at BioTeC, contamination problems of the bioreactor experiments, change in medium composition of the bioreactor experiments, no quantitative assay for AI-2 concentration determination available, problems with protein spot identification, both technical and organizational (initially planned in the VIB-Proteomics

department, however, this lab ceased to operate during the first year of the SQUAD-project) … Nevertheless, some important insights into the *Salmonella* AI-2 mediated quorum sensing have been obtained in the SQUAD-project.

Based on the batch, fed-batch and continuous culture experiments performed, it can be concluded that that AI-2 production appears to be related to the cellular metabolism rather than to population density. This makes the term "quorum sensing" less applicable to the AI-2 system of *S.* Typhimurium. Continuous-culture experiments were used to test the effect of changing metabolic states on the expression of AI-2 related genes. Based on the macroscopic data, it can be stated (and has accordingly been incorporated in the model) that AI-2 in the extracellular environment results from (*i*) production linked to the metabolic state (or growth) and (*ii*) uptake linked with biomass concentration. Details are given in Task 3.1 and 4.1.

This statement was substantiated at the molecular level, by the different high-throughput experiments we conducted (for details, we refer to 1.2), since not many differences (both at mRNA and protein level) were found between *Salmonella* wild-type and AI-2 related mutants. Nevertheless, *luxS* seems to play an important role in the metabolism of *Salmonella*, and careful regulation of LuxS activity can be anticipated. Our results indicate that LuxS is subjected to post-translational modification. However, the validation of this finding turned out to be not that straightforward (for details, see 1.2). Nevertheless, different approaches are being followed and hopefully one of them will allow demonstrating the PTM of LuxS. The importance of LuxS in *Salmonella*'s metabolism makes this enzyme an attractive potential drug target on itself. Moreover, the high-throughput analyses suggested some links with virulence. We are awaiting the ChIP-chip results to substantiate these findings. If true, the AI-2 mediated pathway could be used as starting point for anti-*Salmonella* drug development. However, the application will be more complex than merely targeting the AI-2 signal itself.

Since previous results suggested that the downstream pathways of AI-2 may be more important in other physiological states than in free-living conditions, we incorporated biofilm conditions to study the role of the *luxS* and AI-2 pathway in *Salmonella*. This implied the development of additional protocols (see above). Interestingly, we found that AI-2 is not required for biofilm formation, in contrast to what has been reported before. This finding was partly possible because of the close collaboration between our research group and the Centre for Surface Chemistry and Catalysis (COK, Prof. D. De Vos and Drs. C. Varszegi) who were among the first to chemically synthesize the AI-2 precursor DPD (for details, see 1.2). Surprisingly, we found that a small non-coding RNA molecule is involved in *Salmonella* biofilm formation. These results prompted us to invest into this cutting-edge research and to complement the wet lab experiments with innovative in silico analyses.

As elaborated in the report of the different work packages, still some experiments are ongoing and will be continued by means of follow-up projects. Therefore, many results are still in the pipeline. However, we foster high expectations, especially of the power of the data integration of transcriptomics, location (ChIP-chip) and motif data, and this on two regulators related to the AI-2 pathway (for details, see 1.2). The initial project application of the SQUAD-project might have been rather ambitious, retrospectively. Our results have shown that the pathway we aimed to study is more complex than anticipated and that many parameters need to be taken into account. However, by dissecting each of the problems in a stepwise approach, vast progress was made and the consortium has the potential to become an international player in AI-2 related research and in the technologies developed (e.g. data integration), once all experiments will be finalized.

1.4. List of deliverables

Most of the deliverables were finished (green) or will be finished in the very near future through follow-up projects (blue). Some deliverables will be subject of follow-up projects (orange). Some deliverables were not initiated (red) or were reorientated (purple) for reasons specified in the report.

WP	Task	Description of the deliverable
WP1	1.1	*Salmonella* microarrays (10 mnd)
		Optimized protocols for *Salmonella* microarrays (12 mnd)
		Development of methods to study sRNAs in *Salmonella* (new deliverable)
	1.2	Optimized protocol for 2D-DIGE in *Salmonella* (12 mnd)
	1.3	Optimized protocols for phosphoproteomics (12 mnd)
	1.4	Tools for low-throughput molecular reporter (4/26/36/46 mnd)
		Optimized methods for low-throughput screening (4 mnd)
	1.5	Optimized protocol for ChIP-chip in *Salmonella* (new deliverable)
WP2	2.1	Network generator (25 mnd)
		Tool for integration of heterogeneous data (25 mnd)
		Probabilistic relational model (PRM) (redefined deliverable): Basic implementation of PRM (48 mnd) Optimized sub-module for the introduction of prior knowledge Optimized sub-module for the inference of dynamic processes (16 mnd) Optimized sub-module to cope with hidden variables (follow up project) Optimized sub-module for experiment design (follow up project) *An initial tool is available and will be further extended in a follow up project.*
	2.3	Automated procedure for phylogenetic footprinting (6 mnd)
		Visual interface allowing online use of the developed software (postponed, related to 2.2) *replaced by downloadable compiled linux code*
		Optimized tools for comparative genomics and phylogeny (12 mnd)
		Software module for extension of the knowledge base (postponed, related to 2.2)
		Procedure for construction of a genome-wide motif compendium (48 mnd)
WP3	3.1	Widely applicable bioprocess model reduction tool (15 mnd)
	3.2	General procedure to implement mechanistic knowledge at both cell level and population level into single cell models (30 mnd)
	3.3	Optimal experimental design methodology for combined model structure discrimination and parameter estimation (15 mnd)
WP4	4.1	Low-throughput data on AI-2 activity, LuxS protein and RNA level (4 mnd) in free-living conditions in *Salmonella* biofilm
		Macroscopic mass balance type model (12 mnd)

WP	Task	Description of the deliverable
		Optimally designed experiments based on mass balance type model (14 mnd)
		Optimized model to study *Salmonella* biofilm formation (new deliverable)
		Optimal continuous culture with shifting dilution rates (36 mnd)
	4.2	High-throughput data on the behavior of *Salmonella* at mRNA level related to AI-2 mediated quorum sensing in free-living conditions (36-42 mnd) in *Salmonella* biofilm (36-42 mnd)
		High-throughput data on the behavior of *Salmonella* at protein level related to AI-2 mediated quorum sensing in free-living conditions (30-36-42 mnd) in *Salmonella* biofims (36-42 mnd)
		High-throughput data on the phosphorylation levels of *Salmonella* proteins related to AI-2 mediated quorum sensing in free-living conditions (42 mnd) in *Salmonella* biofilm (42 mnd)
	4.3	Enumeration of AI-2 dependent pathway modules in Salmonella (20 mnd) in free-living conditions in *Salmonella* biofilm
		Design of the most informative additional experiments (21 mnd) in free-living conditions in *Salmonella* biofilm
	4.4	Data generated by performing the most informative additional experiments (40 mnd)
		Small preliminary network structure (46 mnd)
	4.5	Single cell model for *Salmonella* Typhimurium implemented in a user-friendly and computationally efficient environment (44 mnd)
		Validated single cell model for *Salmonella* Typhimurium able to predict intraspecies quorum sensing (48 mnd)
	4.6	Additional triggers for AI-2-signaling pathways in *Salmonella* (48 mnd)
		Data generated by performing experiments including the additional triggers (40 mnd)
		Pathways elicited by the additional triggers (48 mnd)
WP5	5.1	Potential regulatory motifs involved in AI-2-mediated quorum sensing (18/28/38 mnd)
		Additional targets in *Salmonella* of the AI-2-mediated quorum sensing (18/28/38 mnd)
	5.2	Phylogenetic footprints of the identified motifs of the distinct target genes (22/32/46 mnd)
	5.3	Hypothesis on the evolutionary model of regulatory pathways in pathogenic prokaryotes based on the example of quorum sensing (48 mnd)

1.5. References

Ahmer BM (2004) Cell-to-cell signalling in *Escherichia coli* and *Salmonella enterica*. *Mol Microbiol* **52**: 933-945

Argaman L, Hershberg R, Vogel J, Bejerano G, Wagner EG, Margalit H, and Altuvia S (2001) Novel small RNA-encoding genes in the intergenic regions of *Escherichia coli*. *Curr Biol* **11**: 941-950

Banerjee N, Zhang MQ (2002) Functional genomics as applied to mapping transcription regulatory networks. *Curr Opin Microbiol* **5**: 313-317

Bassler BL (1999) How bacteria talk to each other: regulation of gene expression by quorum sensing. *Curr Opin Microbiol* **2**: 582-587

Bassler BL, Greenberg EP, and Stevens AM (1997) Cross-species induction of luminescence in the quorum-sensing bacterium *Vibrio harveyi*. *J Bacteriol* **179**: 4043-4045

Beenken KE, Dunman PM, McAleese F, Macapagal D, Murphy E, Projan SJ, Blevins JS, and Smeltzer MS (2004) Global gene expression in *Staphylococcus aureus* biofilms. *J Bacteriol* **186**: 4665-4684

Beeston AL, Surette MG (2002) *pfs*-dependent regulation of autoinducer 2 production in *Salmonella enterica* serovar Typhimurium. *J Bacteriol* **184**: 3450-3456

Brown L, Elliott T (1996) Efficient translation of the RpoS sigma factor in *Salmonella typhimurium* requires host factor I, an RNA-binding protein encoded by the *hfq* gene. *J Bacteriol* **178**: 3763-3770

Cappuyns AM, Bernaerts K, Smets IY, Ona O, Prinsen E, Vanderleyden J, Van Impe JF (2005) Optimal experiment design in bioprocess modelling: from theory to practice. In: F.J. Doyle III, J.O. Trierweiler and A.R. Secchi (Eds.) Proceedings of ADCHEM 2006, IFAC International Symposium on Advanced Control of Chemical Processes, 535-540. [*IFAC International Symposium on Advanced Control of Chemical Processes (ADCHEM 2006)*, Gramado (Brazil), April 2 - 5 , 2006]

Chen X, Schauder S, Potier N, Van Dorsselaer A, Pelczer I, Bassler BL, and Hughson FM (2002) Structural identification of a bacterial quorum-sensing signal containing boron. *Nature* **415**: 545-549

Chung WSO, Park Y, Lamont RJ, McNab R, Barbieri B, and Demuth DR (2001) Signaling system in *Porphyromonas gingivalis* based on a LuxS protein. *J Bacteriol* **183**: 3903-3909

Datsenko KA, Wanner BL (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci USA* **97**: 6640-6645

De Keersmaecker S, Varszegi C, van Boxel N, Habel LW, Metzger K, Daniels R, Marchal K, De Vos DE, and Vanderleyden J (2005) Chemical synthesis of DPD, a bacterial signal molecule precursor, and validation of its activity in *Salmonella typhimurium*. *J Biol Chem* **280**: 19563-19568

De Keersmaecker SC, Vanderleyden J (2003) Constraints on detection of autoinducer-2 (AI-2) signalling molecules using *Vibrio harveyi* as a reporter. *Microbiology* **149**: 1953-1956

Ejim LJ, D'Costa VM, Elowe NH, Loredo-Osti JC, Malo D, and Wright GD (2004) Cystathionine beta-lyase is important for virulence of *Salmonella enterica* serovar Typhimurium. *Infect Immun* **72**: 3310-3314

Figueroa-Bossi N, Lemire S, Maloriol D, Balbontin R, Casadesus J, and Bossi L (2006) Loss of Hfq activates the sigmaE-dependent envelope stress response in *Salmonella enterica*. *Mol Microbiol* **62**: 838-852

Fong KP, Chung WSO, Lamont RJ, and Demuth DR (2001) Intra- and interspecies regulation of gene expression by *Actinobacillus actinomycetemcomitans* LuxS. *Infect Immun* **69**: 7625-7634

Gottesman S (2004a) Small RNAs shed some light. *Cell* **118**: 1-2

Gottesman S (2004b) The small RNA regulators of *Escherichia coli*: roles and mechanisms. *Annu Rev Microbiol* **58**: 303-328

Kim SY, Lee SE, Kim YR, Kim CM, Ryu PY, Choy HE, Chung SS, and Rhee JH (2003) Regulation of *Vibrio vulnificus* virulence by the LuxS quorum-sensing system. *Mol Microbiol* **48**: 1647-1664

Kleman GL, Strohl WR (1994) Acetate metabolism by *Escherichia coli* in high-cell-density fermentation. *Appl Environ Microbiol* **60**(11):3952-8

Koller D, Pfeffer A (1998) Probabilistic Frame-Based Systems. In Proceedings of the 15th National Conference on Artificial Intelligence (AAAI) (pp. 580-587).

Laub MT, Chen SL, Shapiro L, and McAdams HH (2002) Genes directly controlled by CtrA, a master regulator of the *Caulobacter* cell cycle. *Proc Natl Acad Sci USA* **99**: 4632-4637

Laub MT, McAdams HH, Feldblyum T, Fraser CM, and Shapiro L (2000) Global analysis of the genetic network controlling a bacterial cell cycle. *Science* **290**: 2144-2148

Lee TI, Rinaldi NJ, Robert F, Odom DT, Bar-Joseph Z, Gerber GK, Hannett NM, Harbison CT, Thompson CM, Simon I, Zeitlinger J, Jennings EG, Murray HL, Gordon DB, Ren B, Wyrick JJ, Tagne JB, Volkert TL, Fraenkel E, Gifford DK, and Young RA (2002) Transcriptional regulatory networks in Saccharomyces cerevisiae. *Science* **298**: 799-804

Lenz DH, Mok KC, Lilley BN, Kulkarni RV, Wingreen NS, and Bassler BL (2004) The small RNA chaperone Hfq and multiple small RNAs control quorum sensing in *Vibrio harveyi* and *Vibrio cholerae*. *Cell* **118**: 69-82

Li H, Wang W (2003) Dissecting the transcription networks of a cell using computational genomics. *Curr Opin Genet Dev* **13**: 611-616

Marchal K, Thijs G, De Keersmaecker S, Monsieurs P, De Moor B, and Vanderleyden J (2003) Genome-specific higher-order background models to improve motif detection. *Trends Microbiol* **11**: 61-66

Marchal K, De Keersmaecker S, Monsieurs P, van Boxel N, Lemmens K, Thijs G, Vanderleyden J, and De Moor B (2004) *In silico* identification and experimental validation of PmrAB targets in *Salmonella typhimurium* by regulatory motif detection. *Genome Biology* **5**: R9 Epub

Masse E, Majdalani N, and Gottesman S (2003) Regulatory roles for small RNAs in bacteria. *Curr Opin Microbiol* **6**: 120-124

McClelland M, Sanderson KE, Spieth J, Clifton SW, Latreille P, Courtney L, Porwollik S, Ali J, Dante M, Du FY, Hou SF, Layman D, Leonard S, Nguyen C, Scott K, Holmes A, Grewal N, Mulvaney E, Ryan E, Sun H, Florea L, Miller W, Stoneking T, Nhan M, Waterston R, and Wilson RK (2001) Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2. *Nature* **413**: 852-856

McNab R, Ford SK, El Sabaeny A, Barbieri B, Cook GS, and Lamont RJ (2003) LuxS-Based Signaling in *Streptococcus gordonii*: Autoinducer 2 Controls Carbohydrate Metabolism and Biofilm Formation with *Porphyromonas gingivalis*. *J Bacteriol* **185**: 274-284

Michiels J, Vande Broek A, and Vanderleyden J (1991) Molecular cloning and nucleotide sequence of the *Rhizobium phaseoli recA* gene. *Mol Gen Genet* **228**: 486-490

Miller ST, Xavier KB, Campagna SR, Taga ME, Semmelhack MF, Bassler BL, and Hughson FM (2004) *Salmonella typhimurium* recognizes a chemically distinct form of the bacterial quorum-sensing signal AI-2. *Mol Cell* **15**: 677-687

Monsieurs P, De Keersmaecker S, Navarre WW, Bader MW, De Smet F, McClelland M, Fang FC, De Moor B, Vanderleyden J, and Marchal K (2005) Comparison of the PhoPQ regulon in *Escherichia coli* and *Salmonella typhimurium*. *J Mol Evol* **60**: 462-472

Papenfort K, Pfeiffer V, Mika F, Lucchini S, Hinton JC, and Vogel J (2006) SigmaE-dependent small Rnas of *Salmonella* respond to membrane stress by accelerating global omp mRNA decay. *Mol Microbiol* **62**: 1674-1688

Pollack JR, Perou CM, Alizadeh AA, Eisen MB, Pergamenschikov A, Williams CF, Jeffrey SS, Botstein D, and Brown PO (1999) Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat Genet* **23**: 41-46

Schauder S, Shokat K, Surette MG, and Bassler BL (2001) The LuxS family of bacterial autoinducers: biosynthesis of a novel quorum-sensing signal molecule. *Mol Microbiol* **41**: 463-476

Sittka A, Pfeiffer V, Tedin K, and Vogel J (2007) The RNA chaperone Hfq is essential for the virulence of *Salmonella typhimurium*. *Mol Microbiol* **63**: 193-217

Smets I, Bernaerts K, Sun J, Marchal K, Vanderleyden J, Van Impe J (2002) Sensitivity function-based model reduction: A bacterial gene expression case study. *Biotechnol Bioeng* **80**(2):195-200

Smets I, Bernaerts K, Cappuyns A, Ona O, Vanderleyden J, Prinsen E and Van Impe JF (2004) A prototype model for indole-3-acetic acid (IAA) production by *Azospirillum brasilense* Sp 245. In: S.L. Shah and J.F. MacGregor (Eds.) Proceedings of the 7th International Conference on DYnamics and COntrol of Process Systems, Dycops 7, CDROM, 6p.[*7th International Symposium on DYnamics and COntrol of Process Systems*, Cambridge (Massachusetts, USA), July 5-7, 2004]

Standaert AR, Francois K, Devlieghere F, Debevere J, Van Impe JF, Geeraerd AH (2007) Modeling individual cell lag time distributions for *Listeria monocytogenes. Risk Anal* **27**(1):241-54

Stapleton MR, Norte VA, Read RC, and Green J (2002) Interaction of the *Salmonella typhimurium* transcription and virulence factor SlyA with target DNA and identification of members of the SlyA regulon. *J Biol Chem* **277**: 17630-17637

Stepanovic S, Cirkovic I, Ranin L, and Svabic-Vlahovic M (2004) Biofilm formation by *Salmonella* spp. and *Listeria monocytogenes* on plastic surface. *Lett Appl Microbiol* **38**: 428-432

Sturme MH, Kleerebezem M, Nakayama J, Akkermans AD, Vaughan EE, and de Vos WM (2002) Cell to cell communication by autoinducing peptides in Gram-positive bacteria. *Antonie Van Leeuwenhoek* **81**: 233-243

Surette MG, Bassler BL (1998) Quorum sensing in *Escherichia coli* and *Salmonella typhimurium*. *Proc Natl Acad Sci USA* **95**: 7046-7050

Surette MG, Bassler BL (1999) Regulation of autoinducer production in *Salmonella typhimurium*. *Mol Microbiol* **31**: 585-595

Surette MG, Miller MB, and Bassler BL (1999) Quorum sensing in *Escherichia coli, Salmonella typhimurium*, and *Vibrio harveyi:* a new family of genes responsible for autoinducer production. *Proc Natl Acad Sci USA* **96**: 1639-1644

Taga ME, Bassler BL (2003) Chemical communication among bacteria. *Proc Natl Acad Sci USA* **100 Suppl 2**: 14549-14554

Taga ME, Miller ST, and Bassler BL (2003) Lsr-mediated transport and processing of AI-2 in *Salmonella typhimurium*. *Mol Microbiol* **50**: 1411-1427

Taga ME, Semmelhack JL, and Bassler BL (2001) The LuxS-dependent autoinducer Al-2 controls the expression of an ABC transporter that functions in Al-2 uptake in *Salmonella typhimurium*. *Mol Microbiol* **42**: 777-793

Thijs I, De Keersmaecker S, Fadda A, Engelen K, Zhao H, McClelland M, Marchal K and Vanderleyden J. Delineation of the *Salmonella* Typhimurium HilA regulon through genome-wide location and transcript analysis. *J Bacteriol* **Accepted** for publication

Thijs G, Lescot M, Marchal K, Rombauts S, De Moor B, Rouze P, and Moreau Y (2001) A higher-order background model improves the detection of promoter regulatory elements by Gibbs sampling. *Bioinformatics* **17**: 1113-1122

Thijs G, Marchal K, Lescot M, Rombauts S, De Moor B, Rouze P, and Moreau Y (2002) A Gibbs sampling method to detect overrepresented motifs in the upstream regions of coexpressed genes. *J Comput Biol* **9**: 447-464

Tjaden B, Goodwin SS, Opdyke JA, Guillier M, Fu DX, Gottesman S, Storz G (2006) Target prediction for small, noncoding RNAs in bacteria. *Nucleic Acids Res* **34**(9):2791-802.

Tjaden B (2007) Prediction of small, noncoding RNAs in bacteria using heterogeneous data. J Math Biol. Mar 13; [Epub ahead of print]

Turovskiy Y, and Chikindas ML (2006) Autoinducer-2 bioassay is a qualitative, not quantitative method influenced by glucose. *J Microbiol Methods* **66**(3), 497-503

Van Hellemont R, Monsieurs P, Thijs G, de Moor B, Van de Peer Y, Marchal K. (2005) A novel approach to identifying regulatory motifs in distantly related genomes. *Genome Biol* **6**(13): R113

Vogel J, Bartels V, Tang TH, Churakov G, Slagter-Jager JG, Huttenhofer A, and Wagner EG (2003) RNomics in *Escherichia coli* detects new sRNA species and indicates parallel transcriptional output in bacteria. *Nucleic Acids Res* **31**: 6435-6443

Wang Q, Frye JG, McClelland M, and Harshey RM (2004) Gene expression patterns during swarming in *Salmonella typhimurium*: genes specific to surface growth and putative new motility and pathogenicity genes. *Mol Microbiol* **52**: 169-187

Wassarman KM (2002) Small RNAs in bacteria: diverse regulators of gene expression in response to environmental changes. *Cell* **109**: 141-144

Wassarman KM, Repoila F, Rosenow C, Storz G, and Gottesman S (2001) Identification of novel small RNAs using comparative genomics and microarrays. *Genes Dev* **15**: 1637-1651

Whitehead NA, Barnard AML, Slater H, Simpson NJL, and Salmond GPC (2001) Quorum-sensing in gram-negative bacteria. *FEMS Microbiol Rev* **25**: 365-404

Winzer K, Hardie KR, Burgess N, Doherty N, Kirke D, Holden MTG, Linforth R, Cornell KA, Taylor AJ, Hill PJ, and Williams P (2002) LuxS: its role in central metabolism and the *in vitro* synthesis of 4-hydroxy-5-methyl-3(2*H*)-furanone. *Microbiology* **148**: 909-922

Xavier KB, Bassler BL (2003) LuxS quorum sensing: more than just a numbers game. *Curr Opin Microbiol* **6**: 191-197

Xavier KB, Miller ST, Lu W, Kim JH, Rabinowitz J, Pelczer I, Semmelhack MF, and Bassler BL (2007) Phosphorylation and processing of the quorum-sensing molecule autoinducer-2 in enteric bacteria. *ACS Chem Biol* **2**: 128-136

Zhang Z, Gerstein M (2003) Reconstructing genetic networks in yeast. *Nat Biotechnol* **21**: 1295-1297

2. Interaction with user committee

2.1. Representative of IWT

Dr. Peter Coucke

2.2. Members of user committee

Company	Representative(s) during project
AVEVE*	E. Vanderbeke
FEVIA*	S. Coghe; C. Bosch; E. Lamot
Friesland Foods	A. Nauta; J. Geurts
Janssen Pharmaceutica	S. Van Reet; A. Valcke
PharmaDM	E. Vanheusden; E. Van Craenenbroeck; L. Dehaspe
Provinciaal Instituut voor Hygiëne*	R. Calders
Scientific Institute of Public Health, division Bacteriology*	J.-M. Collard; S. Bertrand
Vitamex	K. Molly ; P. De Backer; K. De Schepper; G. Bruggeman

** added October 2004.*

2.3. Subcontractors

- VIB-MAF: Dr. P. Van Hummelen
- VIB-Proteomics Facility: Dr. B. Hoorelbeke (ceased operation); replaced by:
- Laboratory for Plant Biochemistry and Plant Physiology, University of Antwerp (Dr. E. Witters).

2.4. Consortium members attending the user committee meetings

CMPG: D. De Coster; S. De Keersmaecker; G. Kint; K. Sonck; I. Thijs; N. van Boxel; J. Vanderleyden

ESAT-SCD: B. De Moor; K. Engelen; K. Marchal; K. Lemmens; P. Monsieurs; T. Van den Bulcke

BioTeC: K. Bernaerts; A. Cappuyns; K. Gysemans; I. Smets; A. Standaert; J. Van Impe

2.5. Overview of user committee meetings organized

Project year	Calendar year	Date
1	2003	April 7
1	2003	October 17
2	2004	April 5
2	2004	October 29
3	2005	May 27
3	2005	December 2
4	2006	June 9
4	2007	March 12

2.6. General overview of user committee meetings

Dr. Bruno Lambrecht of Leuven Research and Development (LRD) made up a consortium agreement between KUL LRD and the members of the user's committee in order to protect future results on the project. The contract was signed by all of the members and was valid for the complete duration of the project.
Contacts with the members of the user committee have been positive. At the user committee meetings, more than 75% of the members were present (except for the last user meeting, however, appointments will be made to have bilateral contacts with the respective user to discuss possible valorization tracts of the obtained results) and the results were evaluated and discussed to orientate future research directions. As such, the members of the user committee were active players. Different points of view helped us to redirect some of the work packages of the GBOU project (e.g. including the biofilm work).
Thanks to the critical questions and remarks made by the members of the user committee, we were forced to exceed the semi-artificial environment of laboratory conditions and to project our results to the 'real world'. They also provided us with up-to-date information from the field.
Some members of the GBOU user committee are presently member of the user committee of other research projects.

2.7. Perspectives for further utilization of results

Most of the present results are addressing fundamental issues, and might have little direct value for application. Some results need follow-up projects to be further developed into economical or societal applications. However, the following results have potential for further applications.

During the project, many methodologies in the three different disciplines have been developed and implemented. These can be applied in bilateral projects with members of the user committee to address a specific product-related question (e.g. effect of certain compound on *Salmonella*, biofilm formation). Moreover, these technologies can be readily used in other research projects, reducing the lag-time for the valorization of those results; e.g.:

- the 2D-DIGE set-up is used in another SBO project with high valorisation potential;
- the biofilm research found its application in another research project (IOF), wherein two members of the SQUAD user committee are actively involved
- the bioinformatics tool analysis provided the base required to acquire another SBO project were data integration will be a central topic; valorization of these results will flow through this channel
- The microarray preprocessing algorithms developed in this project could be of great interest to companies involved in the development of spikes or of software tools for bioinformatics. We will evaluate these possibilities.
- methodological developments, e.g., at the level of model identification and optimal experiment design, will be further exploited in other research projects.
- experience with *Salmonella* bioreactor experiments could be useful in the production of vaccines

The finding that AI-2 is rather a reflection of the metabolic status of *Salmonella* than a quorum sensing signal, complicated the execution of the anticipated project outline. However, the fact that LuxS plays such a pivotal role in *Salmonella*'s metabolism, makes it an attractive potential target for anti-*Salmonella* compounds. A follow-up project is needed to unravel the role of the post-translational modification of LuxS in its activity. This information could provide an additional incentive in the development of specific anti-*Salmonella* therapeutics.

Our finding that a small regulatory RNA is involved in *Salmonella* biofilm formation, opens up perspectives for the development of drug targets. Given the hype on the application of siRNA in eukaryotic systems, we believe that insight into the role of this small RNA molecule in *Salmonella* biofilm formation, can contribute to the development of anti-(*Salmonella*)-biofilm therapies based on similar technologies, adapted towards prokaryotes. Moreover, our insight into the targets of this small regulatory RNA molecule (by the combination of wet lab and in silico approaches) could provide an alternative way of interfering with *Salmonella* biofilm formation.

As outlined in the scientific report, some experiments are not yet finished. By combining the transcriptomics, proteomics, location and motif data, we hope to extract some more potential drug targets for further development. Since it is the policy to report SQUAD-related results and their potential validation over a period of 5 years after end of the project, we hope to be able to valorize some of the results in the years to come. Bilateral contacts with the members of the user committee meeting could provide these opportunities.

3. Valorization report

3.1. Dissemination of results

We are disseminating the results achieved during the GBOU project through journal and conference publications, through oral and poster presentations at scientific meetings, through dissertations, master theses, and by organizing workshops.

International peer-reviewed papers

**joint publications*

S. De Keersmaecker and J. Vanderleyden. Constraints on detection of autoinducer-2 (AI-2) signalling molecules using *Vibrio harveyi* as a reporter. 2003. Microbiology. 149:1953-6. SCI 2.617

K. Engelen, B. Coessens, K. Marchal, and B. De Moor. MARAN: a web-based application for normalizing micro-array data. Bioinformatics, 2003, 19:893-894

*K. Marchal, G. Thijs, S. De Keersmaecker, P. Monsieurs, B. De Moor, and J. Vanderleyden. Genome-specific higher-order background models to improve motif detection. Trends in Microbiology, 2003, 11: 61-66

S. Rombauts, K. Florquin, M. Lescot, K. Marchal, P. Rouzé, and Y. Van de Peer. Computational approaches to identify promoters and cis-regulatory elements in plant genomes. Plant Physiology, 2003, 132:1162-1176

B. Coessens, G. Thijs, S. Aerts, J. Mathys, Y. Moreau, K. Marchal, F. De Smet, K. Engelen, P. Glenisson, and B. De Moor. INCLUSive-A web portal and service registry for microarray and regulatory sequence analysis. Nucleic Acids Research, 2003, 31:3468-3470

B. De Moor, K. Marchal, J. Mathys, and Y. Moreau, ``Bioinformatics: Organisms from Venus, technology from Jupiter, algorithms from Mars'. European Journal of Control, 2003, 9:237-278

*G. Dirix, P. Monsieurs, B. Dombrecht, R. Daniels, K. Marchal, J. Vanderleyden, and J. Michiels. Peptide signal molecules and bacteriocins in Gram-negative bacteria: a genome-wide *in silico* screening for peptides containing a double-glycine leader sequence and their cognate transporters. Peptides, 2004, 25:1425-1440

*K. Marchal, S. De Keersmaecker, P. Monsieurs, N. Van Boxel, K. Lemmens, G. Thijs, J. Vanderleyden, and B. De Moor. *In silico* identification and experimental validation of PmrAB targets in *Salmonella typhimurium* by regulatory motif detection. Genome Biology, 2004, 5:R9.1-R9.20, Epub

*P. Monsieurs, G. Dirix, K. Marchal, J. Vanderleyden, and J. Michiels. Screening genomes of Gram-positive bacteria for double-glycine motif containing peptides. Microbiology, 2004, 150:1121-6

K. Bernaerts, E. Dens, K. Vereecken, A. Geeraerd, A.R. Standaert, F. Devlieghere, J. Debevere and J.F. Van Impe 2004. Concepts and tools for predictive modeling of microbial dynamics. Journal of Food Protection, 67(9), 2041-2052

F. Poschet, K.M. Vereecken, A.H. Geeraerd, B.M. Nicolaï and J.F. Van Impe 2004. Analysis of a novel class of predictive microbial growth models and application to co-culture growth. International Journal of Food Microbiology, 100, 107-124.

*P. Monsieurs, S. De Keersmaecker, W. W. Navarre, M. W. Bader, F. De Smet, M. McClelland, F. C. Fang, B. De Moor, J. Vanderleyden, and K. Marchal. Comparison of the PhoPQ regulon in *Escherichia coli* and *Salmonella typhimurium*. Journal of Molecular Evolution, 2005, 60:462-74

*S.C. De Keersmaecker, C. Varszegi, N. Van Boxel, L.W. Habel, K. Metzger, R. Daniels, K. Marchal, D. De Vos, and J. Vanderleyden. Chemical synthesis of (*S*)-4,5-dihydroxy-2,3-pentanedione, a bacterial signal molecule precursor, and validation of its activity in *Salmonella typhimurium*. Journal of Biological Chemistry, 2005, 280:19563-8

*S.C. De Keersmaecker, K. Marchal, T.L. Verhoeven, K. Engelen, J. Vanderleyden, and C.S. Detweiler. Microarray analysis and motif detection reveal new targets of the *Salmonella enterica* serovar Typhimurium HilA regulatory protein, including *hilA* itself. Journal of Bacteriology, 2005, 187:4381-91

T. Van den Bulcke, K. Lemmens, Y. Van de Peer, K. Marchal. Inferring transcriptional networks by mining 'omics' data. Current Bioinformatics, 2006, 1: 301-313

R. Van Hellemont, P. Monsieurs, G. Thijs, B. De Moor, Y. Van de Peer, K. Marchal. A novel approach to identify regulatory motifs in distantly related genomes. Genome Biology, 2005, 6:R113.1-R113.17, Epub

K. Bernaerts, K.P.M. Gysemans, T. Nhan Minh, J.F. Van Impe 2005. Optimal experiment design for cardinal values estimation: guidelines for data collection. International Journal of Food Microbiology, 100, 153-165.

K. Bernaerts and J. F. Van Impe 2005. Optimal dynamic experiment design for estimation of microbial kinetics at sub-optimal temperatures: modes of implementation. Simulation Modelling Practice and Theory, 13, 129-138.

J.F. Van Impe, F. Poschet, A.H. Geeraerd and K.M. Vereecken 2005. Towards a novel class of predictive microbial growth models. International Journal of Food Microbiology, 100, 97-105.

K. Engelen, B. Naudts, B. De Moor, K. Marchal. A calibration method for estimating absolute expression levels from microarray data. Bioinformatics, 2006, 22:1251-8

T. G. Van den Bulcke, K. M.F. Van Leemput, B. Naudts, P. van Remortel, H. Ma, A. Verschoren, B. De Moor, and K. Marchal. SynTReN: A generator of synthetic gene expression data for design and analysis of structure learning algorithms. BMC Bioinformatics, 2006, 26:7-43

*P. Monsieurs, G. Thijs, A. A. Fadda, S. C.J. De Keersmaecker, J. Vanderleyden, B. De Moor and K. Marchal. More robust detection of motifs in coexpressed genes by using phylogenetic information. BMC Bioinformatics, 2006, 7:160

K. Lemmens, T. Dhollander, T. De Bie, P. Monsieurs, K. Engelen, B. Smets, J. Winderickx, B. De Moor, and K. Marchal. Inferring transcriptional module networks from ChIP-chip-, motif- and microarray data. Genome Biology, 2006, 7:R37

*S.C. De Keersmaecker, T.L. Verhoeven, J. Desair, K. Marchal, J. Vanderleyden, and I. Nagy. Strong antimicrobial activity of *Lactobacillus rhamnosus* GG against *Salmonella typhimurium* is due to accumulation of lactic acid. FEMS Microbiological Letters, 2006, 259:89-96

S. De Keersmaecker, K. Sonck and J. Vanderleyden. 2006. Let LuxS speak up in AI-2 signaling. Trends Microbiol. Trends Microbiol. 14:114-9.

K. Van Leemput, T. Van den Bulcke, T. Dhollander, B. De Moor, K. Marchal and P. van Remortel. Exploring the operational characteristics of inference algorithms for transcriptional networks by means of synthetic data. In press, Artificial Life

*S. C. De Keersmaecker, I. M. Thijs, J. Vanderleyden, and K. Marchal. Integration of omics data: how well does it work for bacteria? Molecular Microbiology, 2006, 62:1239-50.

*S. Lebeer, S. C. De Keersmaecker, T. L. Verhoeven, A. A. Fadda, K. Marchal, and J. Vanderleyden. Functional analysis of *luxS* in the probiotic strain *Lactobacillus rhamnosus* GG reveals a central metabolic role important for growth and biofilm formation. Journal of Bacteriology, 2007, 189:860-71

Janssens J., K. Metzger, R. Daniels, D. Ptacek, T. Verhoeven, L.W. Habel, J. Vanderleyden, D. De Vos, S. De Keersmaecker. 2007. Synthesis of *N*-acyl homoserine lactone analogues reveals strong activators of SdiA, the *Salmonella typhimurium* LuxR homologue. Appl Environ Microbiol. 73:535-44.

*I.M.V. Thijs, S.C.J. De Keersmaecker, K. Marchal, K. Engelen, A. Fadda, M. McClelland and J. Vanderleyden. Delineation of the *Salmonella* Typhimurium HilA regulon through genome-wide transcript and location analysis. Journal of Bacteriology. Accepted.

H. Zhao, K. Engelen, B. De Moor, K. Marchal. CALIB: a Bioconductor package for estimating absolute expression levels from two-color microarray data. Bioinformatics, accepted for publication

K.A.J. Sonck, G. Schoofs, S. C. De Keersmaecker, J. Vanderleyden. The proteome of *Salmonella* Typhimurium grown under *in vivo*-mimicking conditions. Submitted.

*A.M. Cappuyns, K. Bernaerts, I.Y. Smets, O. Ona, J. Vanderleyden, E. Prinsen and J.F. Van Impe 2007. Optimal fed batch experiment design for estimation of Monod kinetics of *Azospirillum brasilense*: from theory to practice. Submitted.

Manuscripts in preparation are not included.

Conference proceedings

A.R. Standaert, F. Poschet, F.V. Uylbak, J.-U. Kreft and J.F. Van Impe 2003. Evaluating a novel class of predictive microbial growth models in an individual-based framework. In: *Proceedings of the 9th International Symposium on Computer Applications in Biotechnology*, CDROM, 6p.
[*9th International Symposium on Computer Applications in Biotechnology (CAB9)*, Nancy (France), March 28-31, 2004]

K.P.M. Gysemans, K. Bernaerts and J.F. Van Impe 2003. Constrained input optimization for optimal parameter estimation of a predictive biokinetic model In: *Proceedings of the 9th International Symposium on Computer Applications in Biotechnology*, CDROM, 6p.
[*9th International Symposium on Computer Applications in Biotechnology (CAB9)*, Nancy (France), March 28-31, 2004]

K. Bernaerts, T. Nhan Minh and J.F. Van Impe 2003. Critical evaluation of a nonlinear model from predictive microbiology using sensitivity analysis and optimal experimental design. In: I. Troch and F. Breitenecker (eds.) Proceedings 4th MathMod Vienna, 1274-1280, ARGESIM Report No. 24, Volume 2,Technical University Vienna (ISBN 3-901608-24-9) [IMACS Symposium on Mathematical Modeling, Vienna (Austria), February 5-7, 2003]

K. Bernaerts, K. Gysemans, T. Nhan Minh and J.F. Van Impe 2003. Optimal experiment design for cardinal values estimation: instructions for data collection. In: J.F.M. Van Impe, A.H. Geeraerd, I. Leguérinel and P. Mafart (Eds.), Predictive Modelling in Foods - Conference Proceedings, 111-113, KULeuven/BioTeC, Belgium(ISBN 90-5682-400-7) [Fourth International Conference on Predictive Modelling in Foods, Quimper (France), June 15-19, 2003]

*I. Smets, K. Bernaerts, A. Cappuyns, O. Ona, J. Vanderleyden, E. Prinsen and J.F. Van Impe 2004. A prototype model for indole-3-acetic acid (IAA) production by *Azospirillum brasilense* Sp 245. In: S.L. Shah and J.F. MacGregor (Eds.) Proceedings of the 7th International Conference on DYnamics and COntrol of Process Systems, Dycops 7, CDROM, 6p.[*7th International Symposium on DYnamics and COntrol of Process Systems*, Cambridge (Massachusetts, USA), July 5-7, 2004]

A.R. Standaert, F. Poschet, A.H. Geeraerd, F. Van Uylbak, J.-U. Kreft and J.F. Van Impe 2004. Evaluating a novel class of predictive microbial growth models in an individual-based framework. In: *Proceedings of the 9th International Symposium on Computer Applications in Biotechnology*, CDROM, 6p. [*9th International Symposium on Computer Applications in Biotechnology (CAB9)*, Nancy (France), March 28-31, 2004]

K.P.M. Gysemans, K. Bernaerts and J.F. Van Impe 2004. Constrained input optimization for optimal parameter estimation of a predictive biokinetic model. In: *Proceedings of the 9th International Symposium on Computer Applications in Biotechnology*, CDROM, 6p. [*9th International Symposium on Computer Applications in Biotechnology (CAB9)*, Nancy (France), March 28-31, 2004]

T. Debie, P. Monsieurs, K. Engelen, B. De Moor, N. Christianini, and K. Marchal. Discovering regulatory modules from heterogenous information sources. Proceedings in Pacific Symposium on Biocomputing, Hawai 4-8 jan, 2005, http://helix-web.stanford.edu/psb05/. Internal Report, ESAT-SCD 04-142, K.U.Leuven (Leuven, Belgium).

A.R. Standaert, E.J. Dens and J.F. Van Impe 2003. Individual-based modeling: a mechanistic complement underpinning macroscopic models in predictive microbiology. In: *Proceedings of the 17th Forum for Applied Biotechnology, Part II, Communications in Agricultural and Applied Biological Sciences,* 68/2(b), 529-533 [*17th Forum for Applied Biotechnology (FAB)*, Ghent (Belgium), September 18-19, 2003]

*I. Smets, A. Cappuyns, K. Bernaerts, N. Van Boxel, K. Sonck, S. De Keersmaecker, P. Monsieurs, T. Van den Bulcke, K. Marchal, J. Mathys, B. De Moor, J. Vanderleyden, and J.F. Van Impe 2005. Identifying new drug targets to combat pathogenic infections: an interdisciplinary approach. *Proceedings of the 16th IFAC World Congress*, DVD-ROM, 6p. [*16th IFAC World Congress (IFAC2005)*, Prague (Czech Republic), July 4-8, 2005]

*A.M. Cappuyns, K. Bernaerts, I.Y. Smets, O. Ona, E. Prinsen, J. Vanderleyden, J.F. Van Impe. 2005. Optimal experiment design in bioprocess modelling: from theory to practice. In: F.J. Doyle III, J.O. Trierweiler and A.R. Secchi (Eds.) Proceedings of ADCHEM 2006, IFAC International Symposium on Advanced Control of Chemical Processes, 535-540. [*IFAC International Symposium on Advanced Control of Chemical Processes (ADCHEM 2006)*, Gramado (Brazil), April 2 - 5 , 2006]

K. Bernaerts, E. Van Derlinden, J.F. Van Impe 2005. Estimation of cardinal temperature parameters from dynamic microbial growth experiments: a comparison of different approaches In: M. Vivien and C. Reynès (Eds.) Proceeding of the 9th European conference Food Industry and Statistics, 201-208. [*9th European conference Food Industry and Statistics*, Faculté de Pharmacie, Montpellier (France), January 25 - 27, 2006]

*A. M. Cappuyns, K. Bernaerts, S. C. De Keersmaecker, J. Vanderleyden and J. F. Van Impe 2007. Model for growth and AI-2 type quorum sensing of Salmonella Typhimurium SL1344. accepted for oral presentation at 10th International Symposium on Computer Applications in Biotechnology (CAB 2007), Cancun (Mexico), June 4-6, 2007)

*A. M. Cappuyns, K. Bernaerts, S. C. De Keersmaecker, J. Vanderleyden and J. F. Van Impe 2007. AI-2 type quorum sensing of *Salmonella Typhimurium*: hypotheses testing on the basis of macroscopic data. (submitted)

Abstracts and presentations

K. Sonck, S. De Keersmaecker, G. Schoofs, J. Vanderleyden and I. Nagy. Differential proteome analysis of *Salmonella typhimurium* SL1344 using 2D-DIGE technology. ASM conference on *Salmonella*, September 20-24, 2003, Sardinia, Italy, Book of abstracts p. 82

K. Gysemans, K. Bernaerts, T.N. Minh and J. Van Impe 2003. Critical evaluation of a nonlinear model from predictive microbiology using sensitivity analysis and optimal experimental design. 22nd BeNeLux meeting on Systems and Control, March 19-21, 2003, Lommel (Belgium), ISBN 90-73802-81-4, p. 90

*A. Cappuyns, I. Smets, K. Bernaerts, O. Ona, J. Vanderleyden, E. Prinsen and J.F. Van Impe 2004. Towards a model for Indole-3-Acetic Acid (IAA) production by *Azospirillum brasilense* Sp245. *23rd Benelux Meeting on Systems and Control*, March 17-19, 2004, Helvoirt (The Netherlands), ISBN 90-386-2955-9, p. 23.

*A.M. Cappuyns, K. Bernaerts, I.Y. Smets, S. De Keersmaecker, J. Vanderleyden, J.F. Van Impe 2005. Growth and AI-2 production by the food pathogen *Salmonella typhimurium*. Book of abstracts of 4th NIZO Dairy Conference: prospects for health, well-being and safety, P47. [*4th NIZO Dairy Conference: prospects for health, well-being and safety*, Papendal (the Netherlands), 15-17 june, 2005]

*A.M. Cappuyns, K. Bernaerts, I.Y. Smets, S. De Keersmaecker, J. Vanderleyden, J.F. Van Impe 2006. Growth and AI-2 production by the food pathogen *Salmonella* Typhimurium. Proceedings of the Eleventh Conference on Food Microbiology, University of Liege and Gent, September 14-15, 2006, p.161 [*11th Conference on Food Microbiology*, Gent (Belgium), September 14-15, 2006]

*A.M. Cappuyns, K. Bernaerts, S.C. De Keersmaecker, J. Vanderleyden and J.F. Van Impe 2007. Modeling AI-2 type quorum sensing in *Salmonella* Typhimurium. Accepted for oral presentation at the 26nd BeNeLux meeting on Systems and Control, March 13-15, 2007, Lommel (Belgium)

*K. Marchal, S. De Keersmaecker, P. Monsieurs, N. van Boxel, K. Lemmens, G. Thijs, J. Vanderleyden, and Bart De Moor. *In silico* identification and experimental validation of novel PmrAB targets in *Salmonella typhimurium* by regulatory motif detection. ASM conference on *Salmonella*, September 20-24, 2003, Sardinia, Italy, Book of Abstracts p 70.

*S. De Keersmaecker, K. Marchal, T. L. Verhoeven, C.S. Detweiler, S. Falkow, and J. Vanderleyden. Cluster analysis and motif detection reveal possible new targets of the HilA regulatory protein in *Salmonella typhimurium*. ASM conference on *Salmonella*, September 20-24, 2003, Sardinia, Italy, Book of Abstracts p 83.

R. Van Hellemont, K. Marchal, P. Monsieurs, G. Thijs, Y. Van de Peer, and B. De Moor. Development of a strategy to identify regulatory elements in eukaryotic promoters based on phylogenetic footprinting. SMBE conference (Society for Molecular Biology and Evolution), June 26-29, 2003, Newport Beach, California.

K. Engelen, K. Marchal, J. De Brabanter, T. Ayoubi, P. Van Hummelen, and B. De Moor. Comparison of different methodologies to identify differentially expressed genes in two-sample

cDNA microarrays European conference on computational biology, September 27-30, 2003, Paris, France, Book of Abstracts p 359.

*K. Marchal, P. Monsieurs, K. Lemmens, N. van Boxel, S. De Keersmaecker, G. Thijs, and J. Vanderleyden, B. De Moor. *In silico* identification and experimental validation of novel PmrAB targets in *Salmonella typhimurium* by regulatory motif detection. European conference on computational biology, September 27-30, 2003, Paris, France, Book of Abstracts p. 425.

*P. Monsieurs, K. Marchal, G. Thijs, S. De Keersmaecker, J. Vanderleyden, and B. De Moor. Characterization of the PhoPQ-regulon by *in silico* motif detection. European conference on computational biology, September 27-30, 2003, Paris, France, Book of Abstracts p. 443.

*K. Marchal, S. De Keersmaecker, P. Monsieurs, N. van Boxel, K. Lemmens, G. Thijs, B. De Moor1, and J. Vanderleyden. Reconstruction of the PmrAB dependent transcriptional network of *Salmonella typhimurium* by in silico regulatory motif detection, ESF: Bacterial Neural Networks: EuroConference on Inter- and Intracellular Signalling, and Global Regulation in Bacteria, May 8-13, 2004, San Feliu de Guixols, Spain (selected for oral presentation).

*P. Monsieurs, K. Marchal, S. De Keersmaecker, G. Thijs, B. De Moor, J. Vanderleyden. Comparison of the PhoPQ regulon in *Escherichia coli* and *Salmonella typhimurium,* ESF: Bacterial Neural Networks: EuroConference on Inter- and Intracellular Signalling, and Global Regulation in Bacteria, May 8-13, 2004, San Feliu de Guixols, Spain.

*I. Thijs, Marchal, K., van Boxel, N., Vanderleyden, J., S. De Keersmaecker. (2004). ChIP-CHIP: a genome-wide location analysis to identify transcription factor binding sites in the *Salmonella typhimurium* SL1344 genome. ESF: Bacterial Neural Networks: EuroConference on Inter- and Intracellular Signalling, and Global Regulation in Bacteria, May 8-13, San Feliu de Guixols, Spain. Book of abstracts

S. De Keersmaecker, L. Habel, C. Varszegi, N. van Boxel, R. Daniels, D. De Vos and J. Vanderleyden. Chemical synthesis of the quorum sensing molecule AI-2. 2nd ASM conference on Cell-cell communication in bacteria, Bannf, Canada, July 23-27, 2004, Book of abstracts p. 61.

K.A. Sonck, G. Schoofs, S. De Keersmaecker and J. Vanderleyden. Differential proteome analysis of *Salmonella typhimurium* SL1344 using 2D-DIGE technology. Belgian symposium on *Salmonella* research and control in poultry. Ghent, Belgium, September 9, 2004. Book of abstracts p. 71

I. Thijs, Marchal, K., van Boxel, N., Vanderleyden, J., S. De Keersmaecker. (2004). ChIP-CHIP: a genome-wide location analysis to identify transcription factor binding sites in the *Salmonella typhimurium* SL1344 genome. Belgian symposium on *Salmonella* research and control in poultry. Ghent, Belgium, September 9, 2004. Book of abstracts p. 73

K. Marchal, B. Naudts, K. Van Leemput, K. Engelen, A. Verschoren and B. De Moor. Is it realistic to infer a gene network from a small set of microarray experiments? 5th international conference on systems biology, 9-13 oktober 2004.

K. Engelen, K. Marchal, J. De Brabanter and B. De Moor. Comparison of different methodologies to identify differentially expressed genes in two-sample cDNA microarrays. 5th international conference on systems biology, 9-13 oktober 2004.

K. Lemmens, T. Debie, P. Monsieurs, K. Engelen, B. De Moor, N. Christianini and K. Marchal. Discovering regulatory modules from heterogeneous information sources. ESF workshop on transcription networks: a global view, 26-28 May, Madrid 2005, Book of Abstracts p. .37.

*P. Monsieurs, S. De Keersmaecker, W. W. Navarre, M. W. Bader, F. De Smet, M. McClelland, F. C. Fang, B. De Moor, J. Vanderleyden, and K. Marchal. Comparison of the PhoPQ regulon in

Escherichia coli and *Salmonella typhimurium*. International conference on Bacterial Genomes. April 12-16, 2005, Halifax, Canada. Book of abstracts p.54

T. Van den Bulcke, K. Van Leemput, B. Naudts, P. van Remortel, A. Verschoren, B. De Moor and K. Marchal. A generator of biologically plausible synthetic gene expression data for design and analysis of structure learning algorithms, ECAL Systems biology workshop, 5-9 September, 2005, oral presentation.

K. Engelen, B. Naudts, K. Van Leemput, B. De Moor and K. Marchal. Normalization of cDNA microarrays using external control spikes. ECCB, 2005. Madrid, 28 september-1 October.

K. Lemmens, T. De Bie, P. Monsieurs, K. Engelen, B. De Moor, N. Cristianini , and K. Marchal. Discovering regulatory modules from heterogeneous information sources. ECCB, 2005. 28 september-1 October. Oral presentation student symposium.

T. Van den Bulcke, K. Van Leemput, K. Marchal, P. van Remortel, A. Verschoren, B. Naudts, and B. De Moor A synthetic gene expression data generator for validation of reverse engineering methods. ECCB, 2005. 28 september-1 October.

P. Monsieurs, G. Thijs, B. De Moor, K. Marchal. Motif Detection combining Gibbs Sampling and Comparative Genomics . September 28 – October 1, 2005, Madrid, Spain. Book of abstracts p.94.

Sonck, K., De Keersmaecker, S. & Vanderleyden J. (2005) DIGE protein labeling: How low can you go? Poster presented at the 4th HUPO annual world congress in München (August 28th 2005 – September 1st 2005).

K. Engelen, B. De Moor, K. Marchal. A calibration method for estimating abolute expression levels from microarray data. 29-31 May, 2006, Amsterdam, International Symposium on networks in bioinformatics. Oral presentation.

K. Lemmens, T. Dhollander, T. De Bie, P. Monsieurs, K. Engelen, B. Smets, J. Winderickx, B. De Moor, and K. Marchal. Inferring transcriptional module networks from ChIP-chip-, motif- and microarray data, 29-31 May, 2006, Amsterdam, International Symposium on networks in bioinformatics. Oral presentation.

T. Van den Bulcke, K. Van Leemput, P. van Remortel, B. De Moor, K. Marchal. SynTreN: a synthetic gene expression data generator for validation of reverse engineering methods. 29-31 May, 2006, Amsterdam, International Symposium on networks in bioinformatics. Oral presentation.

T. Van den Bulcke, K. Van Leemput, P. van Remortel, B. Naudts, B. De Moor, and K. Marchal. Benchmarking gene network inference algorithms using synthetic gene expression data. Benelearn-Workshop on bioinformatics & nature inspired computing, May 10 - Ghent – Belgium. Oral presentation.

T. Van den Bulcke, K. Van Leemput, T. Dhollander, B. De Moor, P. van Remortel and K. Marchal. Assessing characteristics of gene network inference algorithms using synthetic gene expression data. Workshop: Probabilistic Modeling and Machine Learning in Structural and Systems Biology June 17-18. Finland.

K. Lemmens, T. Dhollander, T. De Bie, P. Monsieurs, K. Engelen, B. De Moor, K. Marchal. Inferring transcriptional modules from ChIP-chip-, motif- and microarray data. ISMB 2006, Aug 6-10, Fortaleza, Brasil.

K. Engelen, B. De Moor, K. Marchal. A calibration method for estimating absolute expression levels from microarray data. ISMB 2006, Aug 6-10, Fortaleza, Brasil.

T. Van den Bulcke, K. Van Leemput, T. Dhollander, B. De Moor, P. van Remortel, K. Marchal. Using synthetic gene expression data to assess characteristics of gene network inference algorithms. ISMB 2006, Aug 6-10, Fortaleza, Brasil.

*S. De Keersmaecker, D. De Coster, N. van Boxel, K. Sonck, I. Thijs, H. Zhao, K. Engelen, P. Monsieurs, K. Marchal and J. Vanderleyden. In search for the role of LuxS in *Salmonella enterica* sv. Typhimurium biofilm formation. ASM 2006, Sept 9-13, Victoria, Canada.

*P. Monsieurs, G. Thijs, A. A Fadda, S. C. J. De Keersmaecker, J. Vanderleyden. B. De Moor, K. Marchal. More robust detection of motifs in coexpressed genes by using phylogenetic information. ASM 2006, Sept 9-13, Victoria, Canada.

*I. M. Thijs, S. C. J. De Keersmaecker, K. Marchal, K. Engelen, M. McClelland, S. Porwollik, J. Vanderleyden. Genomewide location analysis in *Salmonella enterica* SV. Typhimurium: expanding the HilA regulon, ASM 2006, Sept 9-13, Victoria, Canada.

A. Fadda, P. Monsieurs, K. Marchal. A compendium of predicted motifs for *Bacillus subtilis*. ESF 2006, San Feliu, Spain.

*P. Monsieurs, G. Thijs, A.A. Fadda, S.C. De Keersmaecker, J. Vanderleyden, B. De Moor, K. Marchal. More robust detection of motifs in coexpressed genes by using phylogenetic information, Benelux Bioinformatics Conference 2006, Wageningen, The Netherlands.

H. Zhao, K. Engelen, B. De Moor, and K. Marchal. CALIB: a BioConductor package for estimating absolute expression levels from cDNA microarray data. Benelux Bioinformatics Conference 2006, Wageningen, The Netherlands.

Dissertations (finished)

- Kristof Engelen: Evaluatie van algoritmische methoden voor de analyse van microroostergegevens en de inferentie van genetische netwerken. K.U. Leuven. 2005.
- Pieter Monsieurs: Motif Detection in Prokaryotes Based on Comparative Genomics. K.U. Leuven 2006.

In progress:
- Astrid Cappuyns
- Kristel Gysemans
- Him Hermans
- Joost Janssens
- Gwendoline Kint
- Karen Lemmens
- Kathleen Sonck
- Arnout Standaert
- Inge Thijs
- Tim Van den Bulcke

MSc Theses

Karen Lemmens. 2003. In silico detectie van AI-2-afhankelijke motieven in Salmonella typhimurium aan de hand van phylogenetic footprinting. ESAT-SCD. KU.Leuven

Kristof Vrancken & Karlijn Huys. 2003. Phylogenetic footprinting: *ydeW*, a quorum sensing related gene. (masters in bioinformatics)

S. Yellajosyula. 2003. Phosphorylated proteins in *Salmonella* Typhimurium (masters in bioinformatics)

Csaba Varszegi. 2004. Chemische synthese en validatie van de biologische activiteit van een universeel signaalmolecule. CMPG – COK. K.U. Leuven.

D. Valkenborg and T. D'Hollander. 2004. Ontwikkeling van een dynamische netwerk generator. Ingenieursthesis. Faculteit Toegepaste Wetenschappen.

Nele Ooms. 2004. Een geïntegreerde benadering voor de studie van het effect van de probioticumstam *Lactobacillus rhamnosus* GG op de darmpathogeen *Salmonella typhimurium* SL1344: I. transcriptoomanalyse. Bio-ingenieursthesis. Faculteit Bio-ingenieurswetenschappen.

Mieke Steensels. 2004. Een geïntegreerde benadering voor de studie van het effect van de probioticumstam *Lactobacillus rhamnosus* GG op de darmpathogeen *Salmonella typhimurium* SL1344: II. proteoomanalyse. Bio-ingenieursthesis. Faculteit Bio-ingenieurswetenschappen.

S. Qureshi. 2004. Development of an automated analysis flow for blocksampler. (masters in bioinformatics)

Lisbet De Groote 2005. Grondige studie van groei en AI-2 productie door *Salmonella typhimurium*. Departement Chemische Ingenieurstechnieken; Departement M2S; K.U. Leuven. Promotors: Prof. J. Vanderleyden/Prof. J. Van Impe.

Ying Zhang, Xiaowei Jiang. 2005. Genetic algorithm with bayesian networks for gene structure learning. (masters in bioinformatics)

Gwendoline Kint. 2006. Onderzoek naar 'quorum sensing' gerelateerde pathways in de darmpathogeen *Salmonella typhimurium*. CMPG. K.U. Leuven.

Nick Schorpion. 2006. Implementatie van een genoom-wijde analyse van DNA-eiwitcomplexen in de darmpathogeen Salmonella enterica sv. Typhimurium. CMPG. K.U.Leuven

Koen Verhelst 2006. Evaluatie van technieken voor modeldiscriminatie voor biotechnologische processen. Departement Chemische Ingenieurstechnieken; K.U. Leuven. Promotors: Prof. J. Van Impe/Dr K. Bernaerts

Jan Bollen. 2006. Layout and visualisation of gene module networks. (masters in bioinformatics)

Hong sun. 2006. Glimpse of motif finding algorithms. (masters in bioinformatics)

Zhao Xi. 2006.Assessing global normalization assumption in preprocessing analysis oftwo-color microarray expression data. (masters in statistics, U. Hasselt)

Workshop organization

Internal speakers
On a regular basis, lab seminars in the different research groups by members of the consortium.

External speakers
- Prof. M. McClelland. Sidney Kimmel Cancer Center, San Diego, USA. "Microarray analysis of gene distribution and gene expression in the *Salmonella*." Sept 18, 2003.

- Prof. J. Oberto. Institut de Biologie Physico-Chimique CNRS, Laboratoire de Physiologie Bactérienne, Paris, France. Transcripto-genomics of *E. coli* HU. April 20, 2005.

- Prof. D. Bumann. Max-Planck Institute for Infection Biology, Berlin, Germany; Hannover Medical School, Hannover, Germany. "System-level analysis of *Salmonella*-host interactions in vivo. March 12, 2007.

Organized workshop
We participated in a course on 'bioinformatics" on request of ICES (Institute for Continuing Education in Science), Universiteit Gent.
Belgian Bioinformatics Conference 2007, Leuven

3.2. Contact with industries

Contact with industries have been mainly limited to contacts with the members of our user committee (see II for further details).

The bioinformatics tools and concepts developed in this project need to be more elaborated before they can be valorized as such. However, by interacting with members of the users' committee and pharmaceutical companies we noticed that there seems to be a large need for (more advanced) bioinformatics expertise and that industry is keen to discuss this issue. That is why we would like to further extend and eventually valorize the expertise we have gained in the framework of this project. Two steps were made into that direction:
* ESAT/SCD became member of Bioscope-IT (http://www.bioscope-it.be/index.php?option=com_content&task=view&id=2&Itemid=15). BioScope-IT is a Service Project to help lower the threshold for Flemish biotech companies to apply bioinformatics. As such there will be a better dissemination of the academic research towards industry on the one hand and on the other hand, it would facilitate anticipation of the academic research on the industrial needs.
* we developed a new initiative to set up a systems biology platform in Flanders. This initiative led to a follow up SBO proposal. In this project we would like to further elaborate on developing advanced data integration tools for systems biology.

The same holds true for the technology developed to study biofilm formation. We initiated the preparation of a new SBO project application where other industrial partners will be involved.

3.3. SQUAD as inspiration for product development

Most of the present results are addressing fundamental issues, and might have little direct value for application at this stage. Some results need follow-up projects to be further developed into economical or social applications, as we elaborated in 2.6.

3.4. SQUAD as seed for new research initiatives

SQUAD catalyzed national and international collaborations, e.g.
- At the national level, we are collaborating with U. Antwerp and U. Ghent.
- For the ChIP-chip experiments, we are collaborating with Prof. M. McClelland (SKCC, San Diego, USA).
- For the benchmarking of bioinformatic tools, we collaborate with Prof. Collado-Vides (Center for Genomic Sciences, Mexico; http://www.cifn.unam.mx/Computational_Genomics/computational_genomics.html). Besides the public data, he has access to many private data sources (both microarray data, ChIP-chip data and motif data of *E. coli*). We organised a short visit to the lab of Prof. Collado-Vides to set up a long term collaboration. K. Marchal and K. Lemmens visited the lab of Prof. Collado-Vides for an internship.

- For work on the sRNA molecule, we are collaborating with Prof. J. Vogel, Max Planck Institute for Infection Biology, RNA Biology Group, Berlin, Germany. S. De Keersmaecker spent one month in his lab to perform additional experiments.

SQUAD was also the incentive to apply for new research projects and grants, e.g.

Personal grants

- I. Thijs (FWO-aspirant)
- G. Kint (FWO-aspirant)
- J. Janssens (IWT grant)
- T. Theys (IWT grant)
- R. De Smet (IWT grant)
- S. De Keersmaecker (FWO postdoctoral researcher)
- K. Engelen (PDM-K.U.Leuven)
- P. Monsieurs (PDM-K.U.Leuven)

Projects

- IMPAXOS-SBO-40073
- FWO-project "Identification of genes and their regulatory motifs that are involved in antiproliferative signaling induced by 1,25-dihydroxyvitamin D3 or in the tissue-selective activity of vitamin D analogs"
- Centre of Excellence (K.U.Leuven) "SymBioSys", the K.U. Leuven Center for Computational Systems Biology.
- IOF AIQSI
- GOA CMPG
- SBO BioFrame
- IUAP BioMaGNet

Projects in the running or being applied

- CREA
- SBO on biofilms
- SBO on methodological developments at the level of model identification and optimal experiment design