



HAL
open science

An expanded protein-protein interaction network in *Bacillus subtilis* reveals a group of hubs: Exploration by an integrative approach

Elodie E. Marchadier, Rut R. Carballido y Lopez, Sophie S. Brinster, Celine C. Fabret, Peggy Mervelet, Philippe Bessieres Bessières, Marie-Françoise Noirot-Gros, Vincent V. Fromion, Philippe P. Noirot

► To cite this version:

Elodie E. Marchadier, Rut R. Carballido y Lopez, Sophie S. Brinster, Celine C. Fabret, Peggy Mervelet, et al.. An expanded protein-protein interaction network in *Bacillus subtilis* reveals a group of hubs: Exploration by an integrative approach. *Proteomics*, 2011, 11 (15), pp.2981-2991. <10.1002/pmic.201000791>. <hal-01000282>

HAL Id: hal-01000282

<https://hal.science/hal-01000282v1>

Submitted on 28 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

RESEARCH ARTICLE

An expanded protein–protein interaction network in *Bacillus subtilis* reveals a group of hubs: Exploration by an integrative approach

Elodie Marchadier^{1,2}, Rut Carballido-López¹, Sophie Brinster³, Céline Fabret⁴, Peggy Mervelet¹, Philippe Bessières², Marie-Françoise Noirot-Gros¹, Vincent Fromion^{2*} and Philippe Noirot¹

¹ INRA, UMR1319 Micalis, Jouy-en-Josas, France

² INRA, UR1077 Mathématique Informatique et Génome, Jouy-en-Josas, France

³ INSERM, U1016, Institut Cochin, Paris, France

⁴ CNRS, UMR 8621 IGM, Orsay, France

We have generated a protein–protein interaction network in *Bacillus subtilis* focused on several essential cellular processes such as cell division, cell responses to various stresses, the bacterial cytoskeleton, DNA replication and chromosome maintenance by careful application of the yeast two-hybrid approach. This network, composed of 793 interactions linking 287 proteins with an average connectivity of five interactions per protein, represents a valuable resource for future functional analyses. A striking feature of the network is a group of highly connected hubs (GoH) linking many different cellular processes. Most of the proteins of the GoH have unknown functions and are associated to the membrane. By the integration of available knowledge, in particular of transcriptome data sets, the GoH was decomposed into subgroups of party hubs corresponding to protein complexes or regulatory pathways expressed under different conditions. At a global level, the GoH might function as a very robust group of date hubs having partially redundant functions to integrate information from the different cellular pathways. Our analyses also provide a rational way to study the highly redundant functions of the GoH by a genetic approach.

Received: December 15, 2010

Revised: February 28, 2011

Accepted: March 22, 2011



Keywords:

Cell division / Data integration / DNA replication / Interactome / Microbiology / Stress responses / Transcriptome

1 Introduction

In vivo protein–protein interactions (PPI) direct the assembly of protein complexes, which function as molecular machines in cellular pathways. Hence, the identification of PPIs at proteome scale, the interactome, is crucial to

understand the cellular functions in a given organism. In bacteria, large-scale PPI networks have been established using high-throughput yeast two-hybrid (Y2H) approaches, which identify direct binary interactions, in *Helicobacter pylori* [1], *Synechocystis* sp. [2] and *Campylobacter jejuni* [3]. A high-throughput bacterial two-hybrid was also used to map the binary interactome of *Mycobacterium tuberculosis* [4]. In addition, large-scale purification of affinity-tagged proteins followed by mass spectrometry (AP/MS) identified the protein co-complexes in *Escherichia coli* [5, 6] and in *Mycoplasma pneumoniae* [7].

Correspondence: Dr. Philippe Noirot, INRA, UMR1319 Micalis, F-78350 Jouy-en-Josas, France

E-mail: philippe.noirot@jouy.inra.fr

Fax: +33-1-34652521

Abbreviations: AD, activation domain; GoH, group of highly connected hubs; PCC, pearson correlation coefficient; PPI, protein-protein interaction; TM, transmembrane; Y2H, yeast two-hybrid

*Additional corresponding author: Dr. Vincent Fromion

E-mail: vincent.fromion@jouy.inra.fr

A static view of such PPI networks is a graph with proteins depicted as nodes and interactions as edges. PPI networks, as well as metabolic networks, are characterized by a few highly connected nodes (hubs) among numerous less connected nodes (scale-free topology) [8, 9]. Highly connected hubs tend to correspond to proteins that are essential for cell survival, whereas the scale-free topology confers tolerance to random node removal but a high sensitivity to targeted removal of hubs [10]. In the PPI network of *Saccharomyces cerevisiae* [11], the most highly connected proteins tend to be enriched in duplicate pairs, suggesting that potentially vulnerable nodes are protected by partial redundancy. In yeast, proteins contributing to phenotypic robustness against environmental and genetic variations tend to be highly connected in the PPI and genetic networks and, when duplicated, the different paralogs show considerable divergence in expression [12]. Thus, many of the hubs potentially sensitive to undergo mutations appear to be protected by partial functional redundancy at multiple levels in the genetic architecture. With respect to hub functions, the integration of data from different ‘omics’ approaches has been shown to improve functional annotations and to help to formulate biological hypotheses [13, 14]. The combination of yeast PPI and mRNA expression profiling data helped to distinguish two types of hubs: ‘party hubs’, which interact with most of their partners simultaneously, and ‘date hubs’, which bind their different partners at different times or subcellular locations [15]. Date hubs tend to represent global or ‘high level’ connectors between functional modules, whereas party hubs seem to function inside modules, at a ‘lower level’ of the organization of the cell.

In the Gram-positive bacterium *Bacillus subtilis*, we previously generated a high-quality small PPI network using the Y2H approach [16] and characterized the biological function of some PPIs in various cellular processes [16–22], demonstrating the biological relevance of such high-quality PPI networks. Here, we used the same approach to extend our existing PPI network to other selected cellular processes such as cell division, the bacterial actin-like cytoskeleton, adaptive responses to stress and to proteins of unknown function. The resulting network, composed of 287 proteins linked by 793 specific interactions, connects in a single graph many different cellular functions. Remarkably, a group of highly connected proteins that link the functional processes together was identified in the network. We investigated the biological function of this group of hubs (GoH) by integrating existing knowledge and available expression data. The GoH can be decomposed into subgroups of party hubs corresponding to protein complexes and regulatory pathways that are expressed under different conditions. Our analysis shows that at a global level the GoH might function as a very robust group of date hubs integrating information from/to the different functional modules of the cell. These findings provide new leads to test genetically the function of the GoH.

2 Materials and methods

2.1 Plasmids, strains and media

E. coli strains were grown in LB medium supplemented with ampicillin 100 µg/mL [23]. *B. subtilis* strain 168 was grown in LB medium with the appropriate antibiotics at the following concentrations: erythromycin, 5 µg/mL; chloramphenicol, 6 µg/mL; spectinomycin, 100 µg/mL; kanamycin, 5 µg/mL [24]. *S. cerevisiae* strains PJ69-4a and PJ69-4α [25] were grown in rich YEPD medium or in synthetic complete medium (SC) lacking the appropriate amino acids (Leu, Trp and His) or nucleotides (Ade and Ura) [26]. Vectors pGBDU-Cx and pGAD-Cx [25] were used to fuse proteins of interest with the Gal4p DNA-binding domain (baits) and Gal4p activation domain (AD) (preys), respectively.

2.2 Y2H library screening and specificity assays

The fragmented *B. subtilis* genomic library of preys (Supporting Information Table S1) as well as the bait and prey vectors were constructed as described in the Supporting Information Materials and Methods Section. The genomic library was screened using a mating strategy with bait-containing cells as previously described [16] (see also Supporting Information for details). False-positive interactions generated by the Y2H system were eliminated experimentally in the following way: a selected number of DNA inserts representing all the potentially interacting proteins identified in the screen were PCR amplified from the prey plasmids using primers derived from the pGAD sequence. PCR products were mixed with pGAD vector linearized at the multicloning site and used to transform haploid PJ69-4α cells (see the Supporting Information). Transformants harbored a reconstituted prey plasmid generated by homologous gap repair [27]. Two pools of 3–4 independent transformants each were made and the integrity of the junction between the Gal4p AD domain and the insert-encoded peptide was verified by sequencing each pool. The two pools of haploid cells containing the rescued preys were subjected to mating with cells containing: (i) an empty bait vector, (ii) the initial bait used in the screen and (iii) a variety of unrelated baits. The diploid cells were screened for the expression of the interaction phenotypes (His⁺ and Ade⁺). Specific interactions were reproduced at least two times independently with the initial bait and not associated with self-activation or stickiness of the prey protein. The interactions not fulfilling these criteria corresponded to false positives and were discarded.

2.3 Transcriptome data analysis

All *B. subtilis* transcriptome data sets available at the time of this study were retrieved from the literature and were from

three databases: KEGG (Kyoto Encyclopedia for Genes and Genomes – <http://www.genome.ad.jp/kegg/expression>); GEO (Gene Expression Omnibus – <http://www.ncbi.nlm.nih.gov/geo>) and SMD (Stanford Microarray Database – <http://genome-www5.stanford.edu>). In total, 815 hybridization data were collected and are referenced in Supporting Information Table S2.

All data treatments were done using the R software. Lowess normalization of the individual arrays was done when the channels for Cy3 and Cy5 were available. Inter-arrays normalization was performed by applying the Lenth method [28] to each array: data have been median-centered and reduced using the following formulas:

$$\widehat{T}_{ij} = \frac{T_{ij} - \overline{T}_j}{1.5|\overline{T}_{S_j}|}, \quad (1)$$

where i is the index of the i th gene; j the j th condition; T_{ij} the log-ratio of the i th gene in the j th condition. For a given j , \overline{T}_j is the median of T_{ij} ; $|\overline{\lambda}_j|$ the median of the absolute value of $T_{ij} - \overline{T}_j$; T_{S_j} the values of T_{ij} belonging to $[-3.75|\overline{\lambda}_j|, 3.75|\overline{\lambda}_j|]$ and, $|\overline{T}_{S_j}|$ the median of the absolute value of T_{S_j} ; \widehat{T}_{ij} the normalized value

As many different synonyms can be used to identify a gene, the normalized data sets from different sources were merged together using a correspondence table (Supporting Information Table S3). For each gene, an expression profile over each condition in the data set was obtained, and pairwise Pearson correlation coefficients (PCCs) were calculated with the complete gene set. Associated p -values were determined for each gene pair using the Fisher's Z transformation:

$$(Z(r) = 0.5 \log((1+r)/(1-r))), \quad (2)$$

where r is the PCC) followed by a t -test applied on $\sqrt{N-3} Z(r)$, where N is the number of conditions. We chose to consider the significantly correlated and anti-correlated genes expression profiles according to a stringent multiple-test adjustment: the Bonferroni correction with a threshold of 1%.

Available knowledge on genes was extracted from *B. subtilis* genome annotation [29, 30] and systematic requests of GenBank, and Prosite databases were performed to enrich this functional annotation. Finally, systematic requests of the DBTBS database (<http://dbtbs.hgc.jp/>) were performed to include knowledge on the transcriptional regulations.

In the initial data set of 815 hybridizations, many experiments could be considered as replicates and were averaged to form a final data set of 573 conditions (Supporting Information Table S2, sheet 'non-redundant'). An inter-conditions Lenth normalization [28] was performed, and for each gene a robust variance estimation of the distribution of differential expression was obtained by the method of Lenth, and a p -value was associated with each log ratio by a Student test. Finally, genes exhibiting a significant differential expression were determined by the

multiple-test adjustment of Benjamini–Hochberg with a threshold of 5%. Data were discretized by assigning a value of 1 and -1 to significant differential expression values, whereas all the other expression values were assigned a value of 0. Note that a less stringent multiple-test adjustment was chosen here to obtain a good representation of 1 and -1 in the discrete matrix. From the resulting binary matrix, the *Linear time Closed item set Mine* bi-clustering algorithm [31] (<http://research.nii.ac.jp/~uno/works.htm>) was performed to extract all the maximum bi-clusters. Finally, an R script was developed to identify non-recovering bi-clusters involving GoH genes.

3 Results

3.1 *Bacillus subtilis* protein interaction network

Previously, we constructed a PPI network mainly focused on DNA replication and chromosome maintenance [32]. As this Y2H PPI network provided invaluable functional insights (see Section 1), we used the same careful application of the approach (see Section 2) to construct PPI networks focused on other cellular processes of interest such as cell division, the bacterial actin-like (MreB) cytoskeleton and cell responses to various stresses (involving proteases, sensor and regulatory proteins) thereby further expanding the existing networks. A total of 70 proteins were expressed as full length and also as sub-domains in several instances in fusion with the Gal4p DNA binding domain (BD), representing 98 different baits (Supporting Information Table S4). All the baits were used to screen the *B. subtilis* genomic library composed of prey fragments fused to the Gal4p AD as described in Section 2. Overall, the systematic specificity assays (see Section 2 for details) eliminated $\sim 80\%$ of the candidates isolated from the library screens, yielding protein pairs with a verified capacity to interact specifically in yeast. These interactions are therefore potential PPIs in *B. subtilis*. At an early stage of construction, the PPI network formed a single graph in which seven hub proteins (FruA, Xh1A, SwrC, YhaP, YhgE, YkcC and YqfF) appeared densely connected with proteins involved in cell division, adaptive responses to stress, cytoskeleton and DNA replication as well as with each other, forming a GoH. To identify more systematically the proteins composing the GoH, the seven hub proteins as well as the proteins most highly connected to them (RacA, YclI, YdgH, YqBD, YueB, YwqJ, YkoT) were used as baits to screen the *B. subtilis* library. In addition, for nine proteins with transmembrane (TM) domains (FruA, SwrC, YclI, YdgH, YhaP, YhgE, YkoT, YqfF, YueB), various baits were designed by combining a soluble domain with a varying number of TM domains (Supporting Information Fig. S1 and Supporting Information Table S4). In total, screens were performed with 28 baits corresponding to 14 different proteins. These screens yielded 408 preys, which were rescued in haploid yeast cells, arrayed and subjected to

specificity assays. All preys were also tested in specific Y2H assays for interaction with all GoH proteins as well as with proteins not connected to the GoH in the network. The resulting PPI network was composed of 287 proteins linked by 793 specific interactions (Fig. 1 and Supporting Information Table S6) with an average connectivity of ~ 5 interactions per protein. This network contains 39 hubs defined as having > 10 interaction partners. As some of these hubs

were highly interconnected, we searched the maximal cluster in which each hub had at least 11 interaction partners (see Supporting Information Results Section). This resulted in a mathematically defined and finite GoH composed of 26 proteins within the PPI network (Fig. 1B).

Because such a GoH had not been previously described in bacterial PPI networks, we investigated further the PPIs forming the GoH (see Supporting Information Results

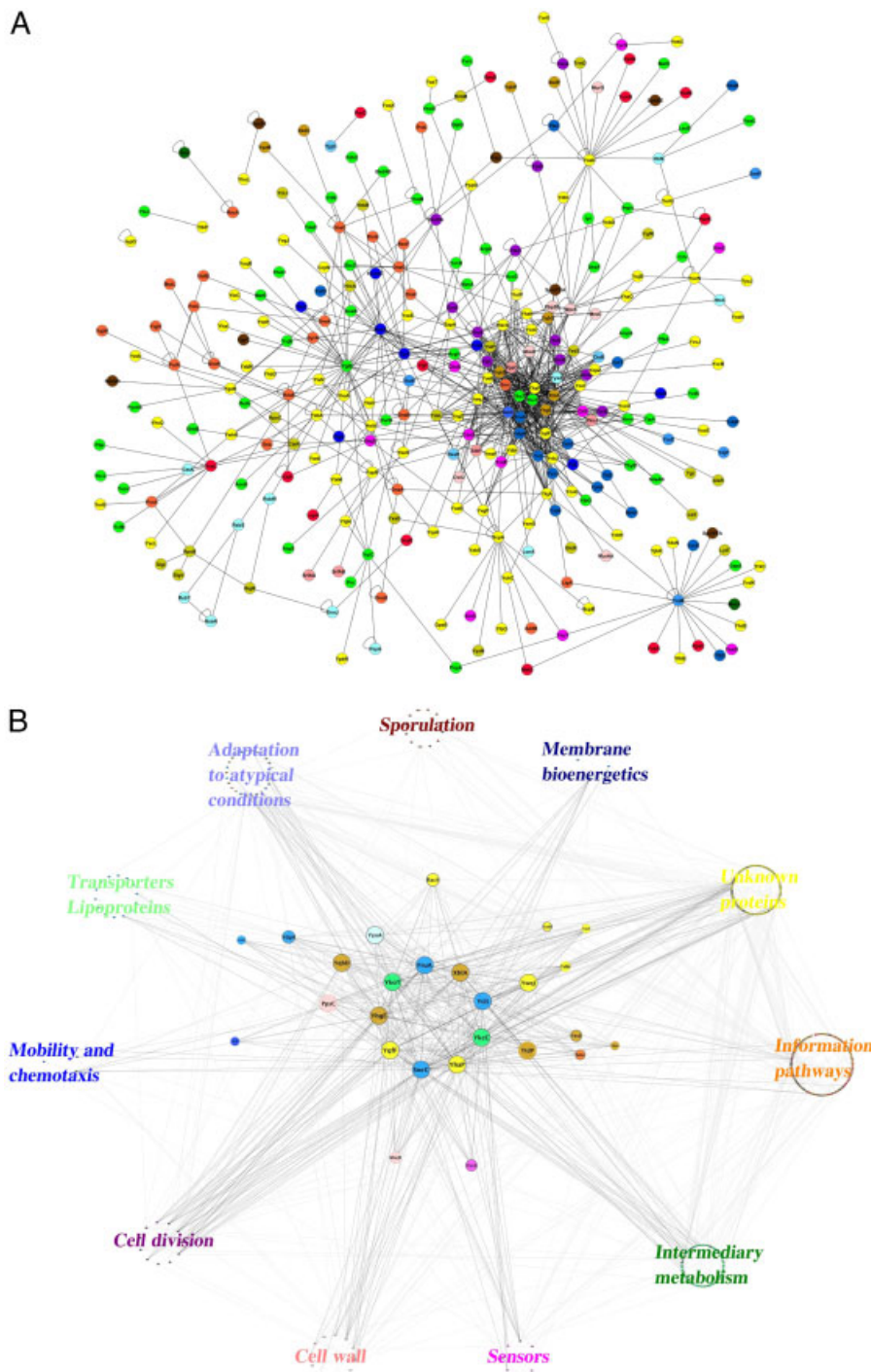


Figure 1. Global representation of the *B. subtilis* protein interaction network. (A) The network is composed of 287 proteins linked by 793 specific interactions. Proteins are represented by nodes (circles) colored according to their functional category (<http://genolist.pasteur.fr/SubtiList/> - data release R16.1). Interactions are represented by edges (lines) connecting nodes. This picture was generated with Cytoscape v2.6.1 [43] using defined visual styles. (B) Visualization of the group of hubs (GoH) and connected cellular processes. The complete protein interaction network is shown organized around the GoH. The 14 hubs most densely connected (large circles) are positioned in the middle of the figure. From the mathematical definition of the GoH (see Section 3), 13 additional hubs with decreasing connectivity are represented (smaller circles at larger distances from the figure center).

Section). We found that: (i) GoH proteins are highly enriched in TM and coil-coiled domains; (ii) specific interaction domains of the hub proteins always overlap protein domains predicted to be soluble; (iii) GoH proteins do not interact with all protein domains containing coiled-coils and (iv) the combination of coil-coiled and TM domains does not suffice to be part of the GoH. Altogether, these findings strongly support that PPIs within the GoH are specific in nature.

3.2 Effects of membrane-spanning domains on PPI

It is generally thought that a Y2H system based on the reconstitution of a functional transcription factor in the yeast nucleus cannot be used to analyze interactions between membrane-associated proteins [33]. To test for a potential bias caused by TM domains in the detection of interactions, we selected a PPI subset comprising various bait fragments composed of a soluble domain associated with varying numbers of TM domains (Supporting Information Fig. S1 and Supporting Information Table S4). The number of PPIs did not correlate with the number of TM domains in the bait protein (Fig. 2). This held true when 13 full-length proteins (4 from the GoH and 9 from outside the GoH) having 1–5 TM domains were included in the analysis. Thus, lack of correlation between the number of PPI and of TM domains is a general rule. This also indicated that our Y2H assay did not introduce a bias towards the detection of TM-containing interacting proteins. Detailed analysis of the PPIs for the nine proteins of the GoH revealed a complex effect of TM domains on the PPI profiles (see Supporting Information Results Section and Supporting Information Fig. S2): TM domains can be neutral, promote or hinder

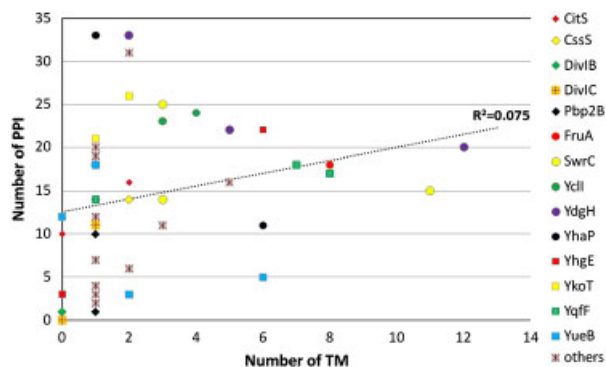


Figure 2. The number of PPIs detected by Y2H is independent of the number of membrane-spanning domains. Different baits containing a soluble domain associated with varying numbers of TM were designed for 14 membrane-associated proteins. A plot of the number of PPIs versus numbers of TM domains is represented. Different baits corresponding to one same protein are represented by one symbol. In addition, 13 full-length proteins (MreD, Xh1A, YkcC, YyxA from the GoH; EzrA, FtsH, FtsL, HtrA, HtrB, MreC, WprA, YkjA, YojM outside the GoH) were added in the plot (others).

PPIs depending on the protein considered. This finding strongly argues that the specific PPI detected in the Y2H assays are not predominantly due to hydrophobic interactions between TM domains. Instead, the complexity of the PPI profiles may be a consequence of the folding adopted by various protein domains expressed as baits. The partnership of each bait fragments was assigned to the full-length protein in the PPI network (Fig. 1).

3.3 Integration of transcriptomic data

Most proteins of the GoH (18/26) have unknown functions, and the remaining eight proteins with a functional annotation are involved in distinct cellular processes (Supporting Information Table S5). The functional information available on the 26 proteins of the GoH was aggregated as exhaustively as possible using motif search, genome-context methods and literature co-citations (see Supporting Information Results Section). Altogether, the knowledge about GoH proteins was too sparse to provide any insight into the functional role of the GoH. The inactivation of individual GoH genes of unknown function did not reveal any significant phenotypes associated with connected processes such as cell division, cell morphogenesis and chromosome segregation (see Supporting Information Results Section). To obtain functional information on the whole PPI network and in particular on the proteins of the GoH, *B. subtilis* transcriptome data sets representing 815 experimental hybridizations were collected from the literature and from the KEGG, GEO and SMD databases (Supporting Information Table S2). Integration of mRNA co-expression data has been shown to increase the confidence of a subset of PPIs [34–36]. The lists of gene names used in the different experiments were unified through a correspondence table (Supporting Information Table S3), and inter-array normalization was performed as described in Section 2. Within this data set, replicates for each condition were averaged to form a non-redundant data set containing 573 experimental conditions (Supporting Information Table S2). Co-expression was determined by calculating the pairwise PCCs for each expression profile in the non-redundant data set. Significantly co-expressed gene pairs were identified using the stringent Bonferroni adjustment with a threshold of 1%. As expected, most of the co-expressed genes belonged to known transcription units (sampled from DBTBS) and 16 of 23 GoH genes were co-expressed with genes of the same putative operons, confirming the validity of our statistical method as well as the quality of the normalized data sets. Interacting protein pairs were not enriched among co-expressed genes as opposed to the validated set of PPIs (Fig. 3A). Several reasons could explain this: (i) the expression of interacting proteins might be correlated only over a small number of biologically relevant conditions; (ii) the expression data set and the PPI data sets have distinct biases towards cellular processes; (iii) the PPIs detected by Y2H

correspond to direct physical interactions that can be transient whereas PPIs validated by other methods (mostly pull-downs) likely reflect more stable interactions. The latter is likely the most important reason because the subset of 122 protein pairs with significant co-expression ($PCC > 0.5$ corresponding to p -values < 0.001 , self-interactions excluded) was found to contain many potentially stable interactions that are known to be part of molecular machines involved in DNA replication, chromosome maintenance, cell

division, chemotaxis and transcription (Supporting Information Table S6).

Interestingly, some pairs of GoH proteins exhibited significant co-expression, suggesting that they could form more stable complexes. A correlation matrix including all the GoH genes revealed four groups of correlated expression profiles (Fig. 3B): (i) *yueB* and *ywqJ*; (ii) *xhlA*, *yqbD* and *yopZ*, three genes from the *B. subtilis* integrated prophages PBSX, SKIN and SP β , respectively; (iii) *fruA*, *mreD*, *smc*, *swrC* and

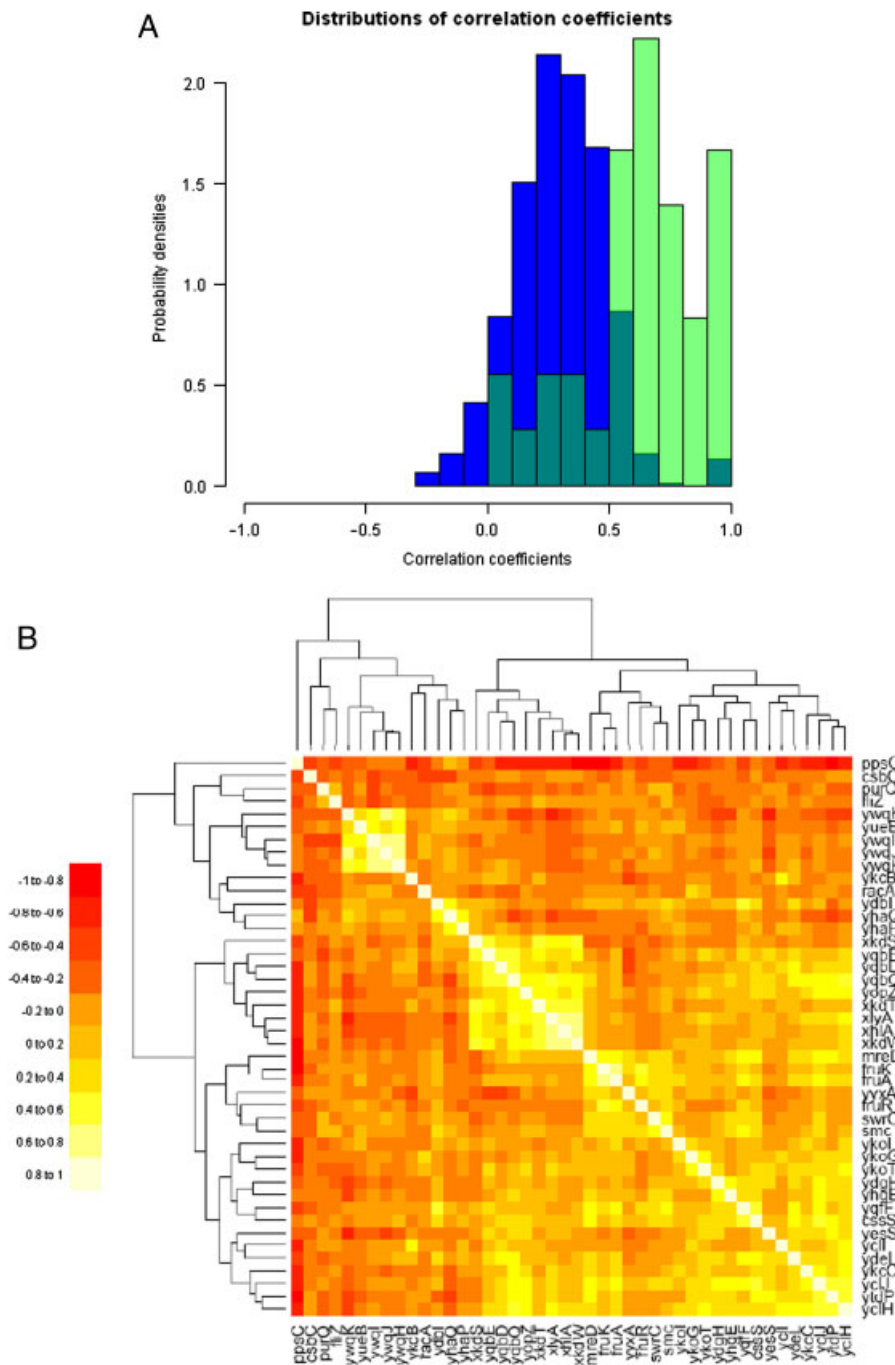


Figure 3. Analysis of co-expressed protein pairs of the GoH. (A) Distributions of Pearson correlation coefficients (PCC) between interacting pairs. Distribution of PCC associated with the whole PPI network is represented in blue (median = 0.294). Distribution of PCC associated with a subset of validated PPI is represented in green (median = 0.658). (B) Hierarchical classification of expression profiles of the GoH genes. The ordered correlation matrix was generated using the Ward clustering method and its associated dendrogram. Genes encoding the GoH proteins are indicated as well as the genes part of the same operonic structure. Correlation coefficients were calculated on the complete data set and are color-coded from white (tending to 1) to red (tending to -1).

yyxA and (iv) *yhgE*, *ydgH*, *yqfF* and *cssS*. A more detailed analysis identified the genes co-expressed with both *yueB* and *ywqJ* (Supporting Information Table S7), potentially providing insight into the biological role of these GoH proteins. Altogether, this analysis of the expression profiles of the genes encoding the GoH proteins revealed four groups of genes exhibiting correlated expression within the groups and uncorrelated expression with the other groups. Figure 4 shows the integration of this information in a graphical representation of the GoH. We concluded that proteins of the GoH are not co-expressed in the cell, indicating that the GoH likely does not form a large multi-protein complex but rather corresponds to smaller complexes whose composition varies in different growth conditions and/or during the cell cycle.

3.4 Identification of conditions for the differential co-expression of groups of GoH genes

For the four co-expression groups identified in the GoH (Fig. 4), the biological conditions in which the groups' components were significantly differentially expressed were investigated. To this end, the matrix corresponding to the non-redundant data set comprising 573 different conditions was discretized and maximum bi-clusters including all the genes displaying the same expression profiles were detected as described in Section 2. Two types of bi-clusters were obtained as illustrated in Table 1. One bi-cluster included GoH genes with similar expression profiles in a large number of conditions specific to this group (*ywqJ-yueB*, *yhgE-ydgH* and *mreD-smc*). Another bi-cluster included GoH genes with similar expression profiles or in a limited number of conditions in which many other genes are also differentially expressed (*yqfF-cssS* and *xhlA-yqbD-yopZ*). For example, *mreD* and *smc* were co-repressed upon cell wall stresses, overproduction of the non-phosphorylated response regulators YclJ and YrkP and nitrate starvation, and they appeared co-induced in knockout mutants of

extracytoplasmic function sigma factors, in particular σ^M , σ^W and σ^X , which mediate responses to cell envelope-active antibiotics [37]. The genes *yueB* and *ywqJ* were co-induced upon *sigD* knockout and DegU overexpression, and they were co-repressed upon *degU* knockout. As expected, the phage-encoded genes *xhlA*, *yqbD* and *yopZ* were co-induced with many other phage genes under conditions of replication arrest in a *recA+* but not in a *recA*-deficient background. Also, the *yqfF* and *cssS* genes were co-repressed upon treatment by cefotaxime and phosphomycin, two antibiotics inhibiting the cell wall biosynthesis. Altogether, these results provide valuable biological information about the specific conditions in which GoH genes are significantly co-expressed. These conditions could be exploited in future experiments to investigate the functions of the GoH.

4 Discussion

A *B. subtilis* PPI network was generated using a careful application of the Y2H approach in which all candidates from genome-wide screens were experimentally verified (see Section 2). About 20% of the PPIs were true bait-prey interactions in yeast, indicating a potential to interact in *B. subtilis* too. Overall, the PPI network described here is composed of 793 interactions linking 287 proteins with an average connectivity of 5 interactions per protein. This network includes our previously reported network of 112 interactions linking 78 proteins involved in DNA replication and chromosome maintenance [32]. Previous functional studies have shown that PPIs obtained by our Y2H approach have a high biological significance [16–22, 38]. This is consistent with the demonstration for the yeast interactome that high-throughput Y2H data contain direct binary interactions of high quality when compared against a gold standard of directly interacting proteins [39]. Thus, the extended high-quality *B. subtilis* PPI network reported here represents a valuable resource for the discovery and functional analysis of new players in many cellular processes.

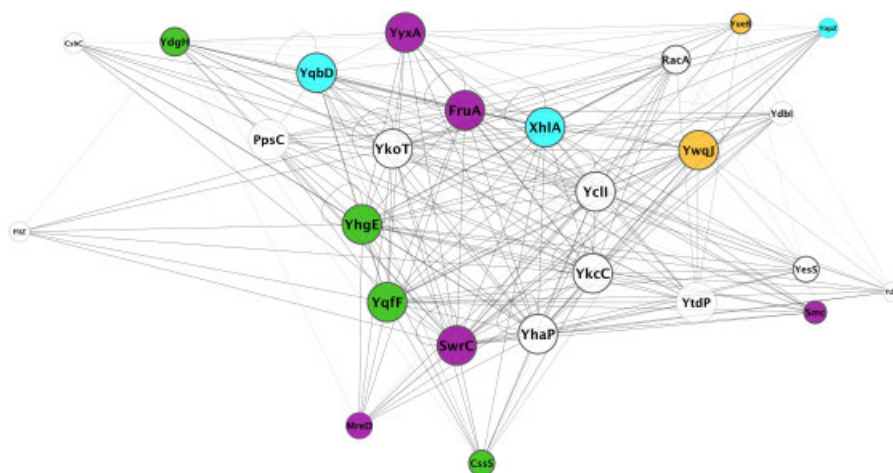


Figure 4. Integration of transcriptome data on the PPI network. The GoH is represented as in Fig. 1B, with proteins (nodes) connected by interactions (edges) to form a network, except that all the edges connecting to cellular processes have been removed. The different co-expression groups identified by the Benjamini–Hochberg correction (threshold 5%, Fig. 3B) are represented by nodes of different colors.

Table 1. Examples of similar expression profiles identified by bi-clustering

Co-expressed genes	Conditions for repression	Conditions for induction
<i>acoA, alr1, araR, cotSA, cssS, gerKA, lytH, manA, msmR, narG, nasC, rbsA, rocC, sacP, sleB, spollIAH, tuaH, uxaB, xpf, yaaH, ycgE, ykcB, yclC, ycll*, yddC, ydfB, ydjB, yesV, yhdB, yhdW, yitF, yitK, yjdG, yjlB, ykcB, ykoQ, ykoU, ykuG, ykvU, yngA, yodH, yomD, yomM, yomO, yonD, yonF, yoql, yoqW, yosF, yosl, yotB, yotE, ypbD, yqal, yqaL, yqcA, yqeD, yqfF, yqxG, yrdF, yrhC, yrkE, ysfD, yteA, ytfJ, ytlC, ytxM, yuxG, yvaD, yvaE, yvdK, yvfJ, yvrB, ywdJ, ywgB, ywmD, yxjC, yxjO, yxIF</i>	ANTIBIO2_CW.788 ANTIBIO2_CW.782	
<i>xepA, xkdD, xkdl, xkdO, xkdR, xkdV, xlyB, xpf, ydcP, ydcRS, yddB, yddC, yddE, yddG, yomD, yomF, yomMN, yomS, yomV,W, yomZ, yonD, yonF, yonH, yonJ, yonK, yopA, yopX, yopZ, yoqE, yoqLJ, yoqL, yorA, yorEF, yorlJK, yorMNO, yorS, yorV, yosF, yosHl, yosK, yosS, yosV, yosZ, yotB, yotl, yqaT, yqbDE, yqbN, yqbO, yqbP, yqcA, yqcC, xhIA</i>	ANTIBIO2_CW.782 STRESS_dna.recomb.481 STRESS_dna.recomb.424	STRESS_dna.recomb.460 STRESS_dna.recomb.454
yueB, ywqJ	STRESS_thioredoxinA.117 ANTIBIO1_CW.272 ANTIBIO1_DNA.307 ANTIBIO1_DNA.309 ANTIBIO1_DNA.310 PHYS_temp.321 MUT_rap.phr.397 MUT_rap.phr.400 STRESS_dna.recomb.436 MUT_degU.638 MUT_degU.639 MUT_degU.641	STRESS_dna.recomb.430 MUT_degU.648 MUT_sigD.677 MUT_sigD.699 MUT_sigD.701
yhgE, ydgH	STRESS_Sodium.21 MUT_codY.681 STRESS_Nstarvation.695	STRESS_NaCl.26 MUT_sigB_STRESS.29 ANTIBIO1_298 ANTIBIO2_PROT.767 ANTIBIO2_PROT.832 ANTIBIO1_RNA.279 ANTIBIO1_RNA.281
smc, mreD	STRESS_Sodium.21 STRESS_Ag.23 ANTIBIO1_CW.271 STRESS_Van.336 MUT_twocompo.658 MUT_twocompo.665 STRESS_Nstarvation.693 STRESS_Nstarvation.694 STRESS_Nstarvation.695 STRESS_Nstarvation.696	MUT_degU.648 MUT_sigM.686 MUT_sigV.689 MUT_sigX.690 MUT_sigW.691

Groups of genes exhibiting similar expression profiles in the discrete matrix are indicated as well as the corresponding experimental conditions in which these genes are co-repressed (discrete values are equal to -1) or co-induced (discrete values are equal to 1). The experimental conditions are labeled as in Supporting Information Table S2. GoH genes belonging to the same co-expression group are indicated in bold. GoH genes within the same bi-cluster are labeled with a star.

The network contains 39 hubs interacting with more than 10 partners. Remarkably, the most highly connected hubs interact with each other to form a GoH. By iteratively searching for proteins having at least 11 connections with the GoH and adding them to the GoH, 26 proteins were defined as components of the GoH in the current network. To the best of our knowledge, such a GoH has not been described in previously reported Y2H large-scale bacterial networks [1–4].

Most proteins of the GoH (22/26) have TM domains, which raised the possibility that GoH interactions are

mediated by non-specific hydrophobic interactions between TM domains. However, several lines of evidence argued against this hypothesis. First, the preys isolated from the library of fragmented proteins revealed that interacting fragments always included a soluble protein domain. Second, the *B. subtilis* proteome contains 1133 TM proteins, and bait proteins with TM domains did not interact randomly with any TM-containing fragment in the library but displayed a limited and specific partnership. Furthermore, varying the number of TM domains in the bait protein had complex effects on the partnership. Indeed, TM

domains could promote, hinder or not affect the PPIs depending upon the protein considered. Altogether, these findings indicate that the specific PPIs detected in the Y2H assays cannot be explained solely by non-specific hydrophobic interactions between TM domains. Instead, TM domains likely contribute to specific PPIs by promoting the proper folding of the soluble domains and by directly mediating specific hydrophobic interactions.

The GoH proteins are densely connected with several major cellular processes (Fig. 1B). Considering that the PPI network was constructed from proteins having known functions in cell division, cell responses to various stresses and the bacterial cytoskeleton, and that it included proteins associated with the DNA replication and chromosome maintenance networks, connections with these processes were expected. However, the connections with proteins classified in the functional groups of transporters, sporulation, membrane bioenergetics and intermediary metabolism were not expected, indicating that the GoH occupies a very crucial position in the network, linking many cellular processes together. Most of the GoH proteins have uncharacterized functions and except *mreD* and *smc*, none of the GoH genes are essential for viability [40]. Thus, although the role of an individual hub may not appear critical for the cell, collectively, the GoH could form a very robust system that could connect the cellular processes in all conditions encountered by the cell.

To gain insight on the organization of the GoH, transcriptome data sets were collected from heterogeneous sources, aligned with respect to gene identifiers and names, and normalized between experiments to define for each *B. subtilis* gene a transcription profile under 573 experimental conditions. For each gene pair, transcription profiles were compared using a PCC associated with a confidence score. The profiles from genes in the same operon were well correlated, thus validating the approach and the quality of the normalized data. Although individual experiments did not reveal changes in the expression of GoH genes, probably because of background noise, the combination of many data sets allowed to distinguish the correlated expression profiles from the noise, and to extract valuable information. Indeed, four subgroups of genes encoding proteins from the GoH that displayed correlated expression profiles could be defined: (i) *ywqJ* and *yueB*; (ii) *xhlA*, *yqbD* and *yopZ*; (iii) *fruA*, *mreD*, *smc*, *swrC* and *yyxA* and (iv) *yhgE*, *ydgH*, *yqfF* and *cssS* (Fig. 4). These findings indicate that the GoH proteins are not all expressed under the same conditions, and that likely they do not form a large multiprotein complex in the cell. Rather, the GoH appears composed of several subgroups of hubs co-expressed under most conditions, which could form multiprotein complexes by interacting simultaneously with most of their partners. Interestingly, *M. pneumoniae* homologues of FruA (Mpn078), Smc (Mpn426) and YkcC (Mpn483) were found to have multiple partners and were classified as multifunctional proteins [7]. Proteins from these subgroups can be classified as 'party

hubs' [15]. The complexes formed by party hubs could be stable molecular machines or less stable interactions involved in signaling and regulatory pathways. Indeed, the subgroups of GoH proteins have a wide range of average value of the PCCs between the protein pairs involved (0.33–0.68 with threshold 1%). Although each party hub binds preferentially to a subgroup of co-expressed hubs, it can also bind to other hubs with uncorrelated expression under different conditions and/or at different times. This suggests that at a global level, the GoH functions as a group of date hubs which could connect and integrate information from at least ten different cellular processes (Fig. 1). The GoH could also participate in the global organization of the *B. subtilis* proteome.

The GoH is composed of several pairs of paralogous proteins (YesS-YtdP, YkcC-YkoT and YueB-YhgE). Interestingly, paralogs were found to be preferentially distributed to the most highly connected proteins in the yeast interactome [11]. The authors proposed that this redundancy evolved to help protect important proteins of the interactome from deleterious mutations. In *B. subtilis*, the presence of paralogs in the GoH could also contribute to its robustness against mutation. However, the main contribution to robustness likely comes from the high redundancy of interaction paths provided by the GoH to link essential cellular processes such as cell division, cell wall biogenesis, chromosome dynamics and others. The organization of the GoH as a group of date and party hubs would make it extremely robust against genes knockouts and would ensure that essential processes remain connected under many environmental conditions. Because of its strong association with the membrane, in particular with two-component systems, it is tempting to speculate that the GoH could integrate external signals. A network of multiple interactions, probably transient or of low affinity, between membrane-associated proteins could also facilitate the formation of dynamic multiprotein structures in the membrane. Such networks have been proposed to promote the stable assembly of the division proteins at midcell in *E. coli* [41] and in *B. subtilis* [42]. The robustness of the GoH against mutations and perturbations represents, however, an important limitation to the analysis of its biological functions by a classical genetic approach. Indeed, inactivation of individual GoH genes did not reveal phenotypes associated with the connected processes (see Supporting Information). Simultaneous inactivation of multiple genes is likely necessary to detect any phenotype. Our analysis suggests that GoH genes within the four co-expression subgroups (see Fig. 4) should be targeted for multiple knockouts. Additionally, the subset of conditions identified by bi-clustering methods, and in which some GoH genes have highly correlated expression profiles, should be selected to perform the genetic analyses of the multiple knockout mutants. Overall, the analyses described here provide rational ways to tackle experimentally the highly redundant functions of the GoH.

We are grateful to Richard Daniel, Ling J. Wu, Jeff Errington, Slade Jensen, Lyndal Thompson, Liz Harry and Etienne Dervyn for the construction of some of the baits and preys related to cell division and for stimulating discussions about the experiments. We also thank Sylvie Huet, Anne Goelzer, Christelle Hennequet-Antier and Leslie Aïchaoui-Denève for constructive discussions about statistics and computational approaches. This work was financially supported in part by an INRA-AgroBI grant to V.F. and P.N., and in part by the BaSysBio project (LSHG-CT-2006-037469) funded by the European Commission.

The authors have declared no conflict of interest.

5 References

- [1] Rain, J. C., Selig, L., De Reuse, H., Battaglia, V. et al., The protein–protein interaction map of *Helicobacter pylori*. *Nature* 2001, **409**, 211–215.
- [2] Sato, S., Shimoda, Y., Muraki, A., Kohara, M. et al., A large-scale protein–protein interaction analysis in *Synechocystis* sp. PCC6803. *DNA Res.* 2007, **14**, 207–216.
- [3] Parrish, J. R., Yu, J., Liu, G., Hines, J. A. et al., A proteome-wide protein interaction map for *Campylobacter jejuni*. *Genome Biol.* 2007, **8**, R130.
- [4] Wang, Y., Cui, T., Zhang, C., Yang, M. et al., Global protein–protein interaction network in the human pathogen *Mycobacterium tuberculosis* H37Rv. *J. Proteome Res.* 2010. DOI: 10.1021/pr100808n.
- [5] Arifuzzaman, M., Maeda, M., Itoh, A., Nishikata, K. et al., Large-scale identification of protein–protein interaction of *Escherichia coli* K-12. *Genome Res.* 2006, **16**, 686–691.
- [6] Butland, G., Peregrin-Alvarez, J. M., Li, J., Yang, W. et al., Interaction network containing conserved and essential protein complexes in *Escherichia coli*. *Nature* 2005, **433**, 531–537.
- [7] Kuhner, S., van Noort, V., Betts, M. J., Leo-Macias, A. et al., Proteome organization in a genome-reduced bacterium. *Science* 2009, **326**, 1235–1240.
- [8] Barabasi, A. L., Albert, R., Emergence of scaling in random networks. *Science* 1999, **286**, 509–512.
- [9] Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N., Barabasi, A. L., The large-scale organization of metabolic networks. *Nature* 2000, **407**, 651–654.
- [10] Jeong, H., Mason, S. P., Barabasi, A. L., Oltvai, Z. N., Lethality and centrality in protein networks. *Nature* 2001, **411**, 41–42.
- [11] Kafri, R., Dahan, O., Levy, J., Pilpel, Y., Preferential protection of protein interaction network hubs in yeast: evolved functionality of genetic redundancy. *Proc. Natl. Acad. Sci. USA* 2008, **105**, 1243–1248.
- [12] Levy, S. F., Siegal, M. L., Network hubs buffer environmental variation in *Saccharomyces cerevisiae*. *PLoS Biol.* 2008, **6**, e264.
- [13] Ge, H., Walkout, A. J., Vidal, M., Integrating ‘omic’ information: a bridge between genomics and systems biology. *Trends Genet.* 2003, **19**, 551–560.
- [14] Hazbun, T. R., Malmstrom, L., Anderson, S., Graczyk, B. J. et al., Assigning function to yeast proteins by integration of technologies. *Mol. Cell* 2003, **12**, 1353–1365.
- [15] Han, J. D., Bertin, N., Hao, T., Goldberg, D. S. et al., Evidence for dynamically organized modularity in the yeast protein–protein interaction network. *Nature* 2004, **430**, 88–93.
- [16] Noirot-Gros, M. F., Dervyn, E., Wu, L. J., Mervelet, P. et al., An expanded view of bacterial DNA replication. *Proc. Natl. Acad. Sci. USA* 2002, **99**, 8342–8347.
- [17] Dervyn, E., Noirot-Gros, M. F., Mervelet, P., McGovern, S. et al., The bacterial condensin/cohesin-like protein complex acts in DNA repair and regulation of gene expression. *Mol. Microbiol.* 2004, **51**, 1629–1640.
- [18] Duigou, S., Ehrlich, S. D., Noirot, P., Noirot-Gros, M. F., DNA polymerase I acts in translesion synthesis mediated by the Y-polymerases in *Bacillus subtilis*. *Mol. Microbiol.* 2005, **57**, 678–690.
- [19] Noirot-Gros, M. F., Velten, M., Yoshimura, M., McGovern, S. et al., Functional dissection of YabA, a negative regulator of DNA replication initiation in *Bacillus subtilis*. *Proc. Natl. Acad. Sci. USA* 2006, **103**, 2368–2373.
- [20] Carballido-Lopez, R., Formstone, A., Li, Y., Ehrlich, S. D. et al., Actin homolog MreBH governs cell morphogenesis by localization of the cell wall hydrolase LytE. *Dev. Cell* 2006, **11**, 399–409.
- [21] Hamoen, L. W., Meile, J. C., de Jong, W., Noirot, P., Errington, J., SepF, a novel FtsZ-interacting protein required for a late step in cell division. *Mol. Microbiol.* 2006, **59**, 989–999.
- [22] Soufo, C. D., Soufo, H. J., Noirot-Gros, M. F., Steindorf, A. et al., Cell-cycle-dependent spatial sequestration of the DnaA replication initiator protein in *Bacillus subtilis*. *Dev. Cell* 2008, **15**, 935–941.
- [23] Sambrook, J., Fritsch, E. F., Maniatis, T., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor 1989.
- [24] Harwood, C. R., Cutting, S. M., *Molecular Biology Methods for Bacillus*, Wiley, New York 1990.
- [25] James, P., Halladay, J., Craig, E. A., Genomic libraries and a host strain designed for highly efficient two-hybrid selection in yeast. *Genetics* 1996, **144**, 1425–1436.
- [26] Guthrie, C., Fink, G. R., *Guide to Yeast Genetics and Molecular Biology*, Academic Press, Inc., San Diego 1991.
- [27] Orr-Weaver, T. L., Szostak, J. W., Yeast recombination: the association between double-strand gap repair and crossing-over. *Proc. Natl. Acad. Sci. USA* 1983, **80**, 4417–4421.
- [28] Lenth, R. V., Quick and easy analysis of unreplicated factorials. *Technometrics* 1989, **31**, 469–473.
- [29] Kunst, F., Ogasawara, N., Moszer, I., Albertini, A. M. et al., The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. *Nature* 1997, **390**, 249–256.
- [30] Barbe, V., Cruveiller, S., Kunst, F., Lenoble, P. et al., From a consortium sequence to a unified sequence: the *Bacillus subtilis* 168 reference genome a decade later. *Microbiology* 2009, **155**, 1758–1775.

- [31] Uno, T., Kiyomi, M., Arimura, H., *Proceedings of IEEE ICDM'04, Workshop FIMI'04* 2004.
- [32] Noirot, P., Noirot-Gros, M. F., Protein interaction networks in bacteria. *Curr. Opin. Microbiol.* 2004, 7, 505–512.
- [33] Stagljar, I., Korostensky, C., Johnsson, N., te Heesen, S., A genetic system based on split-ubiquitin for the analysis of interactions between membrane proteins in vivo. *Proc. Natl. Acad. Sci. USA* 1998, 95, 5187–5192.
- [34] Ge, H., Liu, Z., Church, G. M., Vidal, M., Correlation between transcriptome and interactome mapping data from *Saccharomyces cerevisiae*. *Nat. Genet.* 2001, 29, 482–486.
- [35] Jansen, R., Greenbaum, D., Gerstein, M., Relating whole-genome expression data with protein–protein interactions. *Genome Res.* 2002, 12, 37–46.
- [36] Kemmeren, P., van Berkum, N. L., Vilo, J., Bijma, T. et al., Protein interaction verification and functional annotation by integrated analysis of genome-scale data. *Mol. Cell* 2002, 9, 1133–1143.
- [37] Helmann, J. D., The extracytoplasmic function (ECF) sigma factors. *Adv. Microb. Physiol.* 2002, 46, 47–110.
- [38] Meile, J. C., Wu, L. J., Ehrlich, S. D., Errington, J., Noirot, P., Systematic localisation of proteins fused to the green fluorescent protein in *Bacillus subtilis*: identification of new proteins at the DNA replication factory. *Proteomics* 2006, 6, 2135–2146.
- [39] Yu, H., Braun, P., Yildirim, M. A., Lemmens, I. et al., High-quality binary protein interaction map of the yeast interactome network. *Science* 2008, 322, 104–110.
- [40] Kobayashi, K., Ehrlich, S. D., Albertini, A., Amati, G. et al., Essential *Bacillus subtilis* genes. *Proc. Natl. Acad. Sci. USA* 2003, 100, 4678–4683.
- [41] Karimova, G., Dautin, N., Ladant, D., Interaction network among *Escherichia coli* membrane proteins involved in cell division as revealed by bacterial two-hybrid analysis. *J. Bacteriol.* 2005, 187, 2233–2243.
- [42] Daniel, R. A., Noirot-Gros, M. F., Noirot, P., Errington, J., Multiple interactions between the transmembrane division proteins of *Bacillus subtilis* and the role of FtsL instability in divisome assembly. *J. Bacteriol.* 2006, 188, 7396–7404.
- [43] Shannon, P., Markiel, A., Ozier, O., Baliga, N. S. et al., Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003, 13, 2498–2504.