Chapter 10
Gene Network Holography of the Soil Bacterium *Bacillus subtilis*

Dalit Roth, Asaf Madi, Dror Y. Kenett, and Eshel Ben-Jacob

10.1 Introduction

The development of microarray chip technologies has made it possible to generate vast amounts of raw data regarding the genomic response of the soil bacteria, *Bacillus subtilis* (Caldwell et al. 2001; Fawcett et al. 2000; Hamoen et al. 2002; Molle et al. 2003; Serizawa et al. 2004) [for a review on microarray studies in *B. subtilis*, see (Kocabas et al. 2009)]. Advanced analysis methods have been devised to extract meaningful information from gene-expression data. Most of these methods focus on distinguishing between groups of subjects or identifying the most relevant genes that help to distinguish between these groups – marker genes that exhibit distinct up- or down-regulation.

Current gene-expression data analysis methodologies can be divided into two categories: supervised approaches that aim to determine genes that fit a predetermined pattern; and unsupervised approaches, which aim to characterize the components without a priori assumptions. Supervised methods are usually used to find individual genes such as the nearest neighbor approach (Golub et al. 1999) and/or multiple genes, such as decision trees (Quinlan 1992), neural networks (Rumelhart et al. 1986), and support vector machines (Furey et al. 2000; Hartigan 1975; Rumelhart et al. 1986). Unsupervised methods are usually based on cluster analysis (Everitt 1993; Hansen and Jaumard 1997; Hartigan 1975; Mirkin 1996). Several

D. Roth and A. Madi
Faculty of Medicine, Tel Aviv University, 69978 Tel Aviv, Israel

D.Y. Kenett,
School of Physics and Astronomy, Tel Aviv University, 69978 Tel Aviv, Israel

E. Ben-Jacob (✉)
School of Physics and Astronomy, Tel Aviv University, 69978 Tel Aviv, Israel
The Center for Theoretical and Biological Physics, University of California San Diego, La Jolla, CA 92093, USA
e-mail: eshel@tamar.tau.ac.il; eshelbj@gmail.com

algorithmic techniques were previously used in clustering gene-expression data, including hierarchical clustering (Eisen et al. 1998), self organizing maps (Tamayo et al. 1999), K-means (Herwig et al. 1999), simulated annealing (Alon et al. 1999), and graph theoretic approaches: HCS (Hartuv and Shami 2000), CAST (Ben-Dor et al. 1999), and CLICK (Sharan et al. 2003).

Recently, we have presented the Genome Holography (GH) approach to the investigation and analysis of gene networks (Madi et al. 2008). This method is based on the Functional Holography method (FH), which was previously used to investigate many different complex systems, such as the brain (Baruchi and Ben-Jacob 2004), the immune system (Madi et al. 2009), and the stock market (Shapira et al. 2009). The FH method includes collective normalization of the correlations (according to the correlations of each gene with all the others). The matrices of normalized correlations are then analyzed using dimension reduction algorithms (here we use the Principal Component Algorithm – PCA) to keep the most relevant information. Next, to reveal functional motifs – functional relations between genes, the genes are located in a reduced three-dimensional (3D) space whose axes are the three leading principal vectors of the PCA. We note that projection on a low dimension space of principal vectors is common practice in clustering investigations. Apart from the projection on a 3D space, to retain putative relevant information that can be embedded in the higher dimensions, the genes in the reduced space are linked with lines colored according to the values of the correlations and for selected range of correlation values. Doing so makes it possible to decipher, for example, if distinct clusters in the 3D space are linked in the higher dimensions.

Here we present the approach by analyzing a database of gene expression of *B. subtilis* exposed to sublethal levels of antibiotics (Hutter et al. 2004). The gene-expression levels were monitored in response to 37 different kinds of antibiotics and at three time points after the exposure. We also note that, although the focus in this work on analyzing the matrices of correlations between genes (the gene correlation matrices), important information can also be extracted, in principle, by analyzing the matrices of correlations between the responses to the different antibiotics.

To test the capabilities of the GH method, we investigate its ability to identify, from the gene-expression data, the organization of genes into operons. These functional units (cliques) in the genome are composed of one or more genes cotranscribed into one polycistronic mRNA – a single mRNA molecule that codes for more than one protein. At the macroscopic scale, operons are organized as a network connected by regulators that control, as many other biological networks, joint biological functions and pathways. Recent advancement of experimental methods induced a rapid increase in the available detailed information about the various genes clustered in the operon system. However, knowledge is still lacking about the functional principles that govern the relationship between operon genes and external regulators. We will demonstrate that our analysis method can extract such information from gene-expression data. For example, when projecting the genes of each operon onto the 3D space according to their correlations with the other genes, they tend to form distinct clusters.

The strengths of the GH analysis of gene data will be further demonstrated by representing two more applications. First, the application of the GH analysis to investigate time development of gene regulatory networks under a specific stress will be introduced. Second, we will combine the FH analysis with the commonly used Minimal Spanning Tree (MST) analysis (Mantegna 1999; Mantegna and Stanley 2000; Tumminello et al. 2007; Xu et al. 2001). This allows a more complex yet informative 3D representation of the constructed MST.

Out of the gene database, genes belonging to three well known regulatory networks: sporulation, cannibalism, and competence were selected. These three networks are developmental pathways that maintain the survival of the bacteria under transition state stress conditions, such as nutrient depletion and high cell density (Serror and Sonenshein 1996; Serror et al. 2001; Sonenshein 2000). Other stresses that are also known to cause changes in those pathways are oxidative, osmotic, general stress conditions (Claverys et al. 2006; Dowds et al. 1987; Ruzal and Sanchez-Rivas 1998) and antibiotics (Atsuhiro et al. 2003; Jonas et al. 1990; Prudhomme et al. 2006; Rogers et al. 2007; Vazquez-Ramos and Mandelstam 1981).

Using this novel methodology, we show for a given network the internal gene interactions, in terms of their expression similarity. Furthermore, we demonstrate that these gene interactions change with time, which enabled to identify specific regulated genes that were significantly affected by the antibiotic stress. Finally, novel intra-gene network motifs that are found for a given antibiotic stress are presented.

## 10.2   Functional Holography Analysis of Gene Expression

The FH approach was first introduced by Baruchi et al. (Baruchi and Ben-Jacob 2004) for analysis of recorded human brain activity. The term hologram stands for "whole" – holo in Greek, plus "information" or "message" – gram in Greek. Implementing this methodology on gene expression makes possible to uncover hidden gene structural and functional motifs.

10.2.1   Holographic Presentation of the Genes

In Fig. 10.1, the normalized correlation matrix is presented, which efficiently sorts the genes according to the operons they belong to. In Fig. 10.2, we show the holographic presentation of the matrix of normalized correlations using the projection on the 3D space as described by Madi et al. (2008). Genes belonging to the same operon are given the same color. Lines colored according to the correlation values link pairs of genes with correlations above 0.7. It is possible to observe that
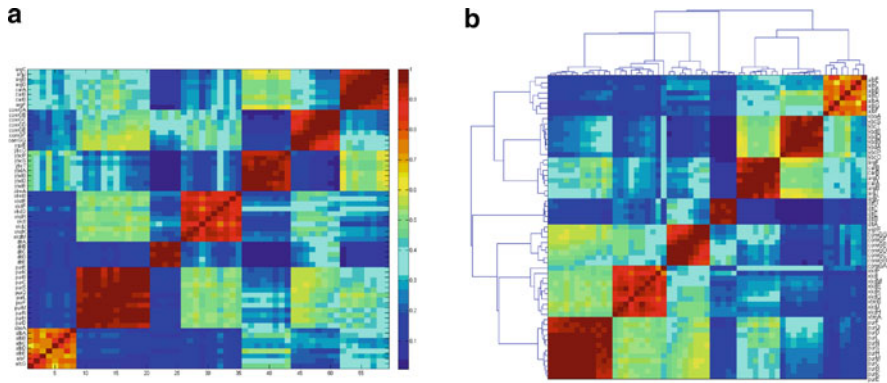
Fig. 10.1 Matrix of normalized correlations. (**a**) The supervised sorted matrix (according to the operons). (**b**) The unsupervised version, sorted by the dendrogram algorithm. Note that the dendrogram algorithm sorts the operons in a different order (7, 3, 1, 5, 2, 4, 6)
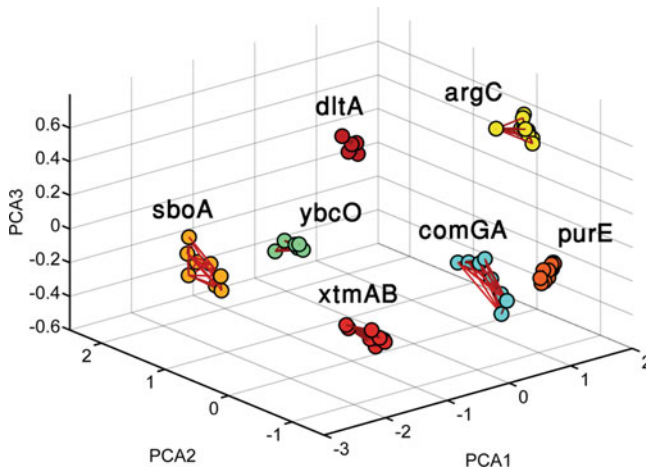
Fig. 10.2 Holographic presentation of the genes in the 3D space. The genes are located in the 3D space whose axes are the three leading principal components calculated by applying PCA to the matrix of normalized correlations. The links between pairs of genes are for correlations above 0.7

only genes belonging to the same operon are linked, implying that interoperon correlations are weaker than intra-operon correlations.

10.2.2 Internal Structure of Gene Operons

In this section, we focus on holographic zooming (Baruchi and Ben-Jacob 2004; Baruchi et al. 2006) analysis of subgroups of genes – the operons. The idea is to

separately perform the collective normalization on the correlation matrix of the subgroup of genes and to calculate a new 3D space for this specific subgroup. A clear correspondence then surfaces between the genes functional relations and the known structures of the operons. The results are illustrated for specific operon – *pyrR* (Chander et al. 2005) that has a non trivial internal organization.

The *pyrR* operon (Chander et al. 2005) has a complex structure, as shown in Fig. 10.3a. It is composed of ten genes that are organized in three subunits: (1) The gene *pyrR* that acts as self-inhibitor of the operon as a whole and also acts as an inhibitor of the three subunits. (2) The gene *pyrP* is located downstream from the *pyrR* with a terminator segment in between. (3) The third subunit is composed of eight genes downstream from the *pyrP* with terminator and promoter segments in between. This operon is regulated by SigA and PurR.

The normalized correlation matrix of the *pyrR* operon genes, *pyrR*, *pyrP*, *pyrB*, *pyrC*, *pyrAA*, *pyrAB*, *pyrK*, *pyrD*, *pyrF* and *pyrE*, and the corresponding holographic presentation are shown in Fig. 10.3b, c. In Fig. 10.3c, it can clearly be seen that *pyrR* and *pyrP* are distinct from each other and from the eight genes of the third subunit of the operon. We also note that the holographic functional organization of the operon in the 3D space corresponds to the structural organization of the operon, as *pyrR* gene is linked only to the *pyrP*, which, in turn, is linked to the rest of the genes. The genomic scheme of the operon in Fig. 10.3a is consistent with these results, as *pyrR* and *pyrP* are separated from the rest of the genes in the operon by terminators. Furthermore, it was previously shown that PyrR is a protein that
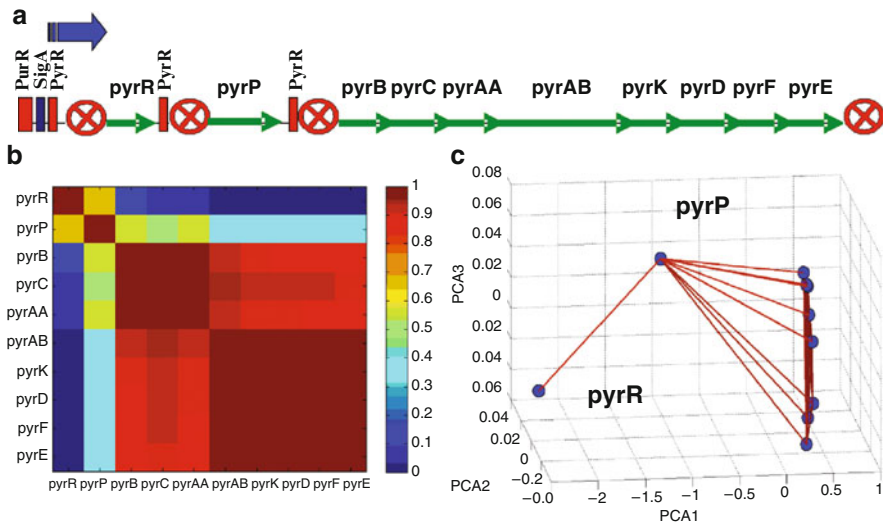
Fig. 10.3 The pyrR operon. (**a**) Schematic internal sequential structure presentation (Chander et al. 2005). Promoter represented by a *blue arrow*. Terminator represented by a *marked circle*. *Rectangles* represent regulation regions of *sigA*, *purR*, and *pyrR*. (**b**) Matrix of normalized correlation. Note the low correlations between *pyrR* to the rest of the genes, matching its function as a negative regulator. (**c**) FH holographic presentation, with correlation values above 0.7 marked by *lines*

regulates the expression of genes and operons of pyrimidine nucleotide biosynthesis (*pyr* genes) in many bacteria and specifically in B. subtilis. *pyrR* acts by binding to specific sequences on *pyr* mRNA, causing transcriptional attenuation when intracellular levels of uridine nucleotides are elevated (Chander et al. 2005).

Based on this demonstrated efficiency of the method, we proceeded to test its ability to predict/reveal unknown functional relations. To test this ability, we investigate the *spoVAA* operon (Azevedo et al. 1993). In Fig. 10.4a, we show the currently presumed internal structure of this operon. The matrix of normalized correlations and its holographic networks in the PCA space are shown in Fig. 10.4b, c, respectively. Inspecting these results, it was observed that *lysA* has weak correlations with the other genes in the operon. These results are somewhat unexpected since no terminator or regulation factors were found between *spoVAF* and *lysA* (Genbank L09228). Azevedo et al. found a 2.3 kb transcript originating about 1 kb upstream of the *lysA* start codon, suggesting that transcription of *spoVA* continues into the *lysA* gene. However, the *lysA* gene is also transcribed monocistronically as a 1.3-kb transcript. A possible explanation might be the existence of a regulation element, a terminator perhaps, between the *spoVAF* and *lysA* genes. Another possible explanation can also be the existence of an additional unknown pathway (through another gene) in which the *lysA* gene acts as a negative regulator of the spoVAA operon. LysA mediates the last step of the lysine biosynthesis (Rodionov et al. 2003). The lysine-mediated gene regulation in bacteria appears to operate via a unique RNA structural element similar to riboswitch that is involved in the regulation of purin biosynthesis (Mandal et al. 2003). The LYS element is
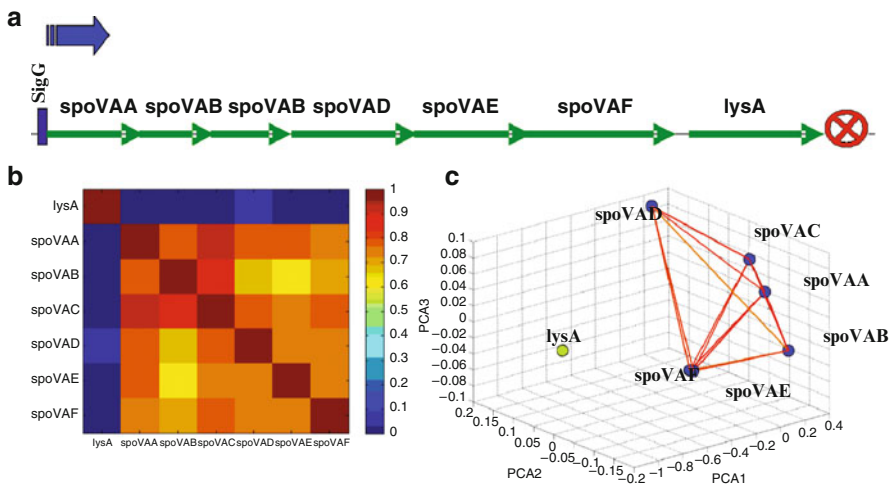
Fig. 10.4 The spoVAA operon (Azevedo et al. 1993). (**a**) Schematic internal sequential structure presentation. The promoter is represented by a *blue arrow*, the terminator by a *red crossed circle*, and the binding site of the activator by a *purple rectangle*. (**b**) Matrix of normalized correlations. (**c**) FH holographic presentation in 3D PCA space, where correlation values above 0.8 is shown in *lines*

characterized by its compact secondary structure with a number of conserved helices and extended regions of sequence conservation, which could be necessary for specific metabolite binding (Rodionov et al. 2003). Comparative genomic analysis predicted conserved RNA secondary structures in lysine metabolism genes such as *lysC* and *lysA*. Thus, our analysis supports the genomic prediction of a regulatory element adjacent to the *lysA* gene and transcription of lysA monocistronically.

10.3   Minimal Spanning Tree Analysis

The use of graph theory and network theory has become widespread over the past few years in the analysis of many physical (Donetti et al. 2005), biological (de Jong 2002; Ortega et al. 2008), and economic systems (Coronnello et al. 2005; Tumminello et al. 2007). In particular, graph theory can be useful for studying correlation-based systems.

The weighted adjacency matrix. The first step in order to utilize the analyses methods from network theory is to compute the adjacency matrix that corresponds to the correlation matrix. The adjacency matrix *A* is a binary matrix that was developed to describe the topology of the network connectivity. In the case of activity network such as correlations between stocks (Mantegna 1999) or synchronization between neuron firing (Fuchs et al. 2009), the function matrix (e.g., correlations or synchronizations) is first transformed into a distance matrix. This can be done in different ways; one of the more common ways is by using the ultrametric distance, first suggested by Mantegna et al. (Mantegna 1999; Mantegna and Stanley 2000). This is done by transforming the normalized correlation between two nodes, $Aff(i,j)$, into a distance by

$$D(i,j) = \sqrt{2[1 - Aff(i,j)]} \qquad (10.1)$$

Note that high correlations correspond to short distances and vice versa. In a network that corresponds to a weighted adjacency matrix, all the nodes are connected (with weighted links between the nodes according to the distances) and hence the topological structure is that of a complete graph.

Minimum spanning tree. Since the relevant information can be obscured in the complete graph (West 2001), the idea is then to extract a subgraph, the minimum spanning tree (MST), from the complete network, based on the weights of the links in order to extract the most relevant information contained in the complete graph. We then focus on the topological structure of the sub graph, and make use of graph theory techniques to analyze this information (Albert and Barabási 2002; Graham and Hell 1985; Mantegna and Stanley 2000; Newman 2003).

Here we have chosen to use the Kruskal algorithm (Kruskal 1956) to compute the MST. This algorithm looks for a subset of the branches that forms a tree that includes every node, where the total weight of all the branches, namely the score

that is derived from the correlation, in the tree is minimized. If the graph is not connected, then it finds a *minimum spanning forest* (a minimum spanning tree for each connected node). More specifically, the algorithm starts where every node considered a separate tree in the forest. Then, the algorithm considers each branch in turn, in order by increasing weight. If a branch connects two different trees, then it is added to the set of branches of the MST, and two trees connected by this branch are merged into a single tree on the other hand; if a branch connects two nodes in the same tree, then it is discarded. This is repeated until the forest is reduced into one tree.

To demonstrate the use of MST to analyze gene networks (see also Xu et al. 2001), we apply this method to study genes belonging to seven operons. This tree is presented in Fig. 10.5. We color-code the genes of each operon in the same color, and thus are able to observe that each operon is grouped together in the tree.

Finally, we combine the MST analysis with the FH analysis, from which we obtain the Functional Holography minimal Spanning Tree (FHMST). To achieve this, we first calculate the reduced PCA space for a given set of genes, and then project on it the gene connections, according to the results of the MST analysis.

The results of this method are demonstrated using the sporulation genes, discussed above. First, we perform the FH analysis on these genes, using the data for all three-time points for this example. Once we have created the FH space, we compute the MST for these genes, and from it obtains the connections between the genes. This is presented in Fig. 10.6.
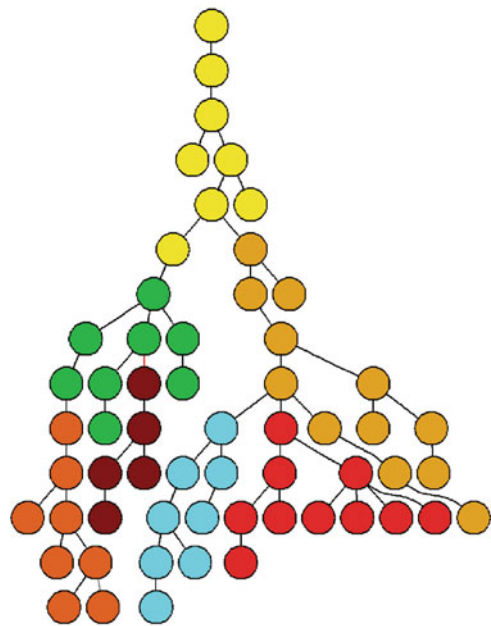
Fig. 10.5 MST analysis of gene operon data. We created the MST for genes belonging to seven operons and color-coded each operon according to Fig. 10.2. We observe that each operon is grouped in the tree
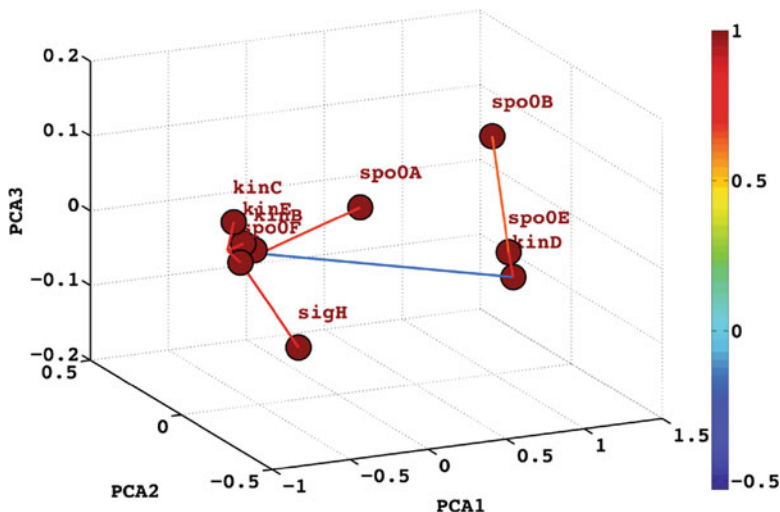
Fig. 10.6 Functional Holography Minimal Spanning Tree (FHMST) for the sporulation gene network

10.4 Time Progression of the Sporulation and Competence Networks

Sporulation is a multistage, developmental process that is responsible for the conversion of a growing cell into a dormant cell type known as the spore or endospore (Stragier and Losick 1996). Sporulation is not initiated automatically (deterministically) upon nutrient limitation, but instead it is the end result of a series of steps that might be described as cellular decisions regarding how to best cope with the stress (Veening et al. 2005).

The master regulator for entry into sporulation is the response regulator Spo0A (Fig. 10.7) (Hoch 1993). The activity of Spo0A is governed by a multicomponent phosphorelay, which consists of five histidine autokinases (KinA, KinB, KinC, KinD, and KinE), and two phosphorelay proteins (Spo0F and Spo0B) (Jiang et al. 2000), which are responsible for the phosphorelation of Spo0A (Burbulys et al. 1991). The level of phosphorylation of Spo0A~P is also influenced by dedicated phosphatases that remove phosphoryl groups from Spo0F~P and from Spo0A~P itself (Spo0E) (Grossman 1995; Parego et al. 1994).

The competence regulatory network, which acts as a stochastic switch, controls the escape rate into competence from the sporulation path and the exit time back from the competence state (Jiang et al. 2000). Competence is the ability of the bacteria to bind and take up exogenous DNA (Lorenz and Wackernagel 1994). In the center of the signal-transduction network of the competence system positioned, ComK (Fig. 10.7). *comK* is a positive auto-regulatory gene that regulate the onset of competence (Madi et al. 2008). Activation of the ComK transcription factor is
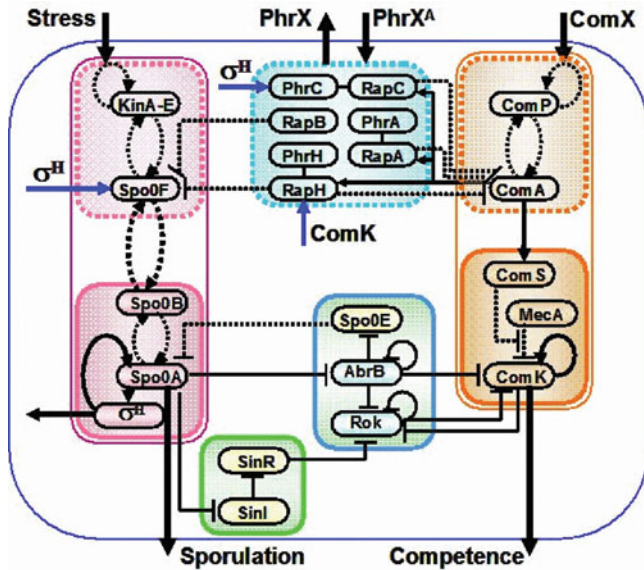
Fig. 10.7 The sporulation-competence signal transduction network. The network modules are described in details in the text. The sporulation module is shown on the *left* and the competence module is shown on the *right*. Each module is composed of a sensing unit and a regulatory unit. These are the KinA-E stress sensing and the Spo0A timer comprising the sporulation module and the Spo0P-ComA quorum sensing and the ComK switch comprising the competence module. The two main modules interact via the Rap communication and information processing module (*upper center*), the AbrB-Rok decision module (*middle center*) and the SinR-SinI commitment unit (*lower center*). The input and output signals are represented by the *wide solid lines* that cross the outer *black* (Schultz et al. 2009)

controlled by many genes and appears to be the step at which multiple physiological signals that affect competence are integrated (Grossman 1995).

10.4.1 Time Progress of Sporulation Initiation Gene Network

Here we focus on the sporulation initiation genes, which include *kinA-E*, *spo0F*, *sigH*, *spo0B*, *spo0A*, and *spo0E*. First, the normalized gene correlations were calculated for all time points (Fig. 10.8). In Fig. 10.9, the FH analysis is presented, for these genes for the three different time points, under the stress of all 37 antibiotics. The analysis of each time point is performed using the FH projection principle.

A close investigation of Fig. 10.9 leads to three important observations. First, in all three-time points we observe that the genes are separated into two main clusters that are negatively correlated. Second, we observe at the 80 min interval after exposure that the genes *spo0E* and *kinD* become more negatively correlated to the

Fig. 10.8 Normalized correlation matrix for the sporulation initiation genes, for all three time points
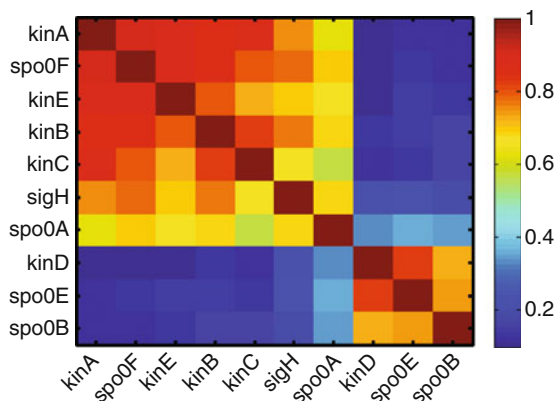

Kin genes. Third, by comparing Fig. 10.9a, b, it is possible to observe that the gene *spo0A* significantly changes its location in the reduced 3D PCA space. In the first time point, 10 min after exposure, this gene has a special relationship to the remaining genes that organize as two separate clusters.

The gene that is affected the most by the different antibiotics is *spo0A*. The correlation of *spo0A* to the other genes changes throughout the three different time points. After 10 min, *spo0A* is situated in the PCA space between the two clusters and has positive correlation with *spo0E* from one cluster and *kinB*, *kinE*, and *spo0F* from the second cluster. After 40 and 80 min, *spo0A* becomes closer to the *Kin* cluster, *spo0F* and to *sigH* (Fig. 10.9). The transition of *spo0A* toward this latter cluster might indicate an ongoing process of synchronization in the sporulation system over the exposure time to antibiotics.

The separation of the genes into two groups under all antibiotics is quite surprising. Most of the genes mentioned above should have been correlated, leading to phosphorylation of Spo0A and entrance into sporulation. Spo0E is the only negative regulator in this network which dephosphorylate Spo0A ~ P. Here it was found that *kinD* and *spo0B* are also correlated to *spo0E* and have negative correlation to the other genes, negative correlation that becomes stronger as the exposure time increased (Fig. 10.9). More work is needed to clarify the negative correlation of *spo0B* to the other sporulation genes, especially with *spo0F*.

Fujita and Losick found that KinA, KinB, and KinC behave in a significantly different manner than KinD and KinE, with the former being capable of triggering sporulation (Fujita and Losick 2005). This provides a possible partial explanation to the negative correlation of *kinD* to the other *kin* genes. Furthermore, Hoch (Stephenson and Hoch 2002) has described the different expression time of the five *Kin* genes. However, while his work showed that *kinA*, *kinD*, and *kinE* are highly expressed during growth and at equal levels, we have not observed such high correlations between these three genes, rather only for the *kinA* and *kinE* genes.

In an attempt to resolve this issue, we considered analyzing the sporulation genes for only specific antibiotics, for the different exposure times. A number of studies
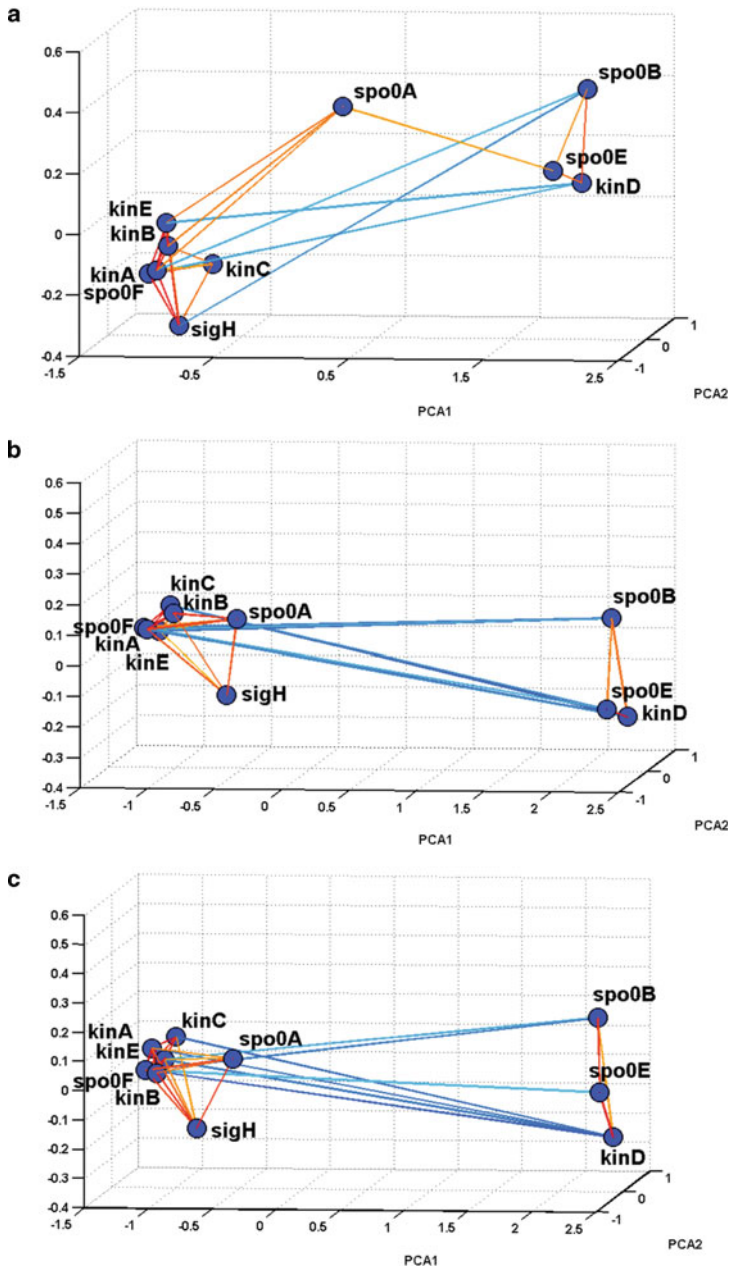
Fig. 10.9 Holographic network of sporulation initiation genes at three time points after exposure. (**a**) After 10 min of exposure, (**b**) after 40 min of exposure, and (**c**) after 80 min of exposure to all antibiotics. For visualization purposes, the correlation lines connecting the genes were set by correlation threshold greater than 0.4 and smaller than −0.4

have shown that some protein synthesis antibiotics, such as chloramphenicol, lincomycin, erythromycin, as well as antibiotics from other classes such as novobiocin, nalidixic acid, and penicillin, inhibit sporulation initiation (Atsuhiro et al. 2003; Jonas et al. 1990; Vazquez-Ramos and Mandelstam 1981). Thus, we analyzed the interaction between the genes under specific antibiotics that were predicted to have an effect on the expression of these genes and compared it to the interactions between the genes under all of the remaining antibiotics. The comparison of the FH analysis of the sporulation inhibition antibiotics to the noninhibiting antibiotics is presented in Fig. 10.11. Comparing the left panels of Fig. 10.11 to the right panels, it is possible to observe a different effect of these two groups of antibiotics. For the case of the sporulation inhibition antibiotics, the results of the FH analysis were similar to the ones obtained for all antibiotics (compare right panel of Fig. 10.11 to Fig. 10.9). In contrast, in the case of the noninhibiting antibiotics, all genes were clustered as one group with very high correlations between the genes (above 0.8).

We thus suggest that the high correlation within the sporulation network shows the synchronization of the genes under antibiotic stress. However, antibiotics that have a major effect on sporulation possibly influence this strong correlation. Since we know that most of the genes that lead to sporulation are correlated (*kinA, B, C, E, spo0A, sigH, sigF*), and that *spo0E* that has a negative regulation on sporulation has also negative correlation with those genes, this suggests that *kinD* and possibly *spo0B* are involved in negative regulation of the sporulation system under the effect of specific antibiotics.

10.4.2   Time Progress of Competence Gene Network Response

In this section, we illustrate the ability of the new approach to reveal additional dynamical motifs related to the time progress in the response of the two gene networks to the antibiotic stress. First, we examined the response of the competence network to 37 antibiotics (Fig. 10.12). The competence gene network studied here contains the genes *comK*, *Rok*, *abrB*, *degU*, and *comG* operon (Fig. 10.13). The FH analysis described above was repeated for these genes, separately for each time point following exposure (Fig. 10.14).

Studying Fig. 10.14, it is possible to see that the most important change is the formation of positive correlation between *degU* and late competence genes (*comK* and *comG* operon).

comK is an auto-regulatory gene in the center of the competence network. AbrB and Rok are repressors of *comK*, and DegU is a positive regulator. ComK is a positive regulator of the *comG* operon and a repressor of the *rok* gene (Fig. 10.13). Since *comG* operon is under a direct regulation of ComK and it triggers its transcription, the cluster of *comG* genes and *comK*, and the strong correlation between these genes at all time intervals is highly expected. Furthermore, ComK
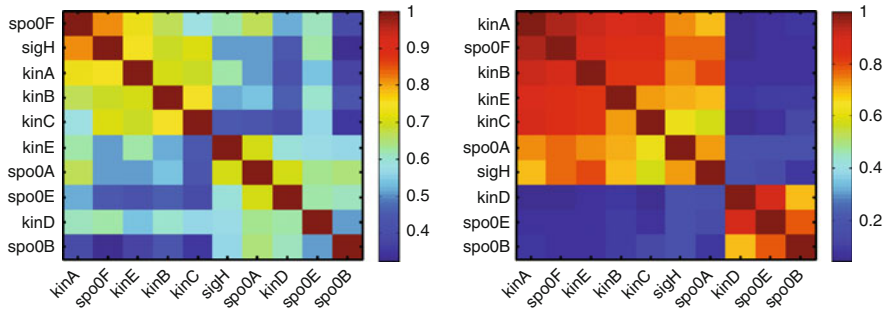
Fig. 10.10 Normalized correlation matrix for the sporulation initiation genes, for all three time points. The *right panel* represents the normalized correlations of the genes under specific antibiotics (chloramphenicol, clarithromycin, clindamycin, erythromycin, fusidic acid, neomycin, puromycin, spectinomycin, tetracyclin, novobiocin, nalidixic acid and penicillin). The *left pane* represent the normalized correlations of sporulation genes under 25 antibiotics from different classes, not including the former set

is the key regulator in triggering the late competence genes. When performing the FH analysis with all the competence genes (results not shown), *comK* was located in the cluster of all the late competence genes, showing the high correlation between this gene and the other competence genes. Note that even though *comG* operon is also negatively controlled by *comZ* (Ogura and Tanaka 2000), the correlation between *comK* and *comG* is not affected under stress of antibiotics.

DegU is one of *comK* positive regulators and was expected to have high correlation to *comK*. Nevertheless, in our findings we show that after 10 min of exposure, *degU* is correlated only with *abrB*. Correlation between *comK* and *comG* evolved over time under all antibiotics, but still the two genes are distant. Although DegU has a main role in *comK* regulation and competence, it was shown that the main function of DegU in the development of competence is to stimulate binding of ComK at the onset of competence development (Hamoen et al. 2000), and overproduction of ComK bypasses the need for DegU. ComK appears to be able to function without DegU if present at sufficiently high concentrations (Roggiani and Dubnau 1993). This might support our finding of the low correlation level between the two genes after 10 min. It has also been suggested in the past that unphosphorylated DegU is required for competence, whereas phosphorylated DegU activates the production of degradative enzymes (Dahl et al. 1992). Mutation that causes hyper phosphorylated form of DegU, by DegS, results in decreased competence, diminished motility, and glucose-insensitive sporulation. For this reason, we assume that correlation between *comK* and *degU* will not necessarily appear in the transcription level but they might exist in the phosphorylation level under non stressful conditions.

Nevertheless, the correlations between *degU* and late competence genes evolve over time under all antibiotics. It is possible that the influence of specific antibiotic that affect the *comK* network become dominant over time, as will be shown below.

The response of the competence to DNA topology ("topo") antibiotics antibiotics was examined, at the three exposure time points – 10, 40, and 80 min.
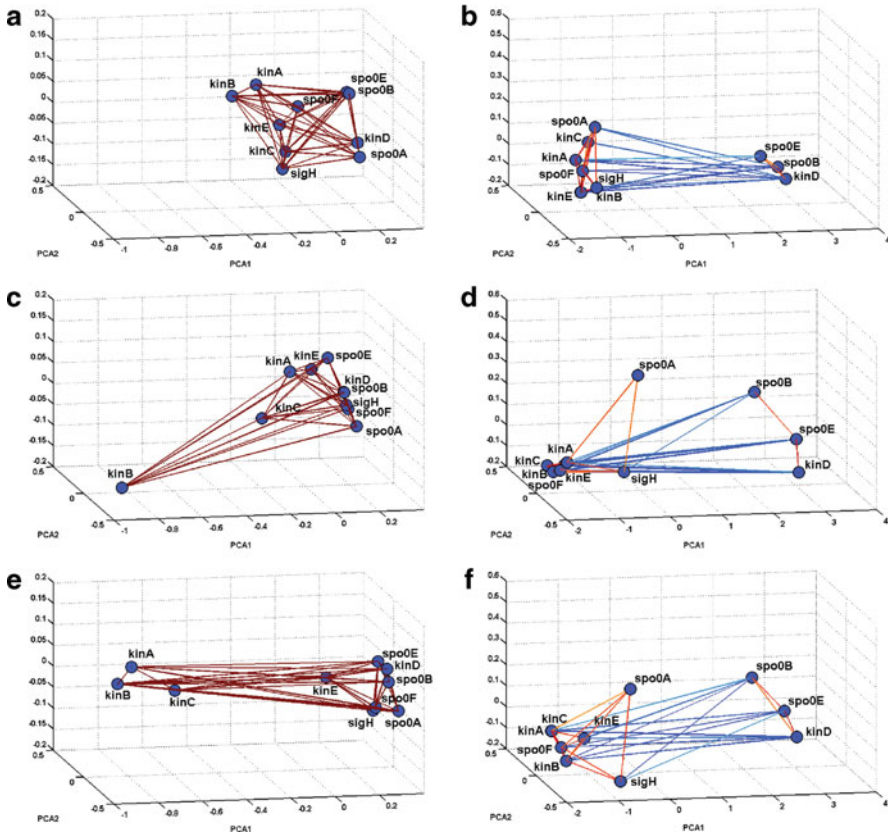
Fig. 10.11 Sporulation initiation genes under the effect of different antibiotics. The *right panels* represent the FH analysis of the genes under specific antibiotics (chloramphenicol, clarithromycin, clindamycin, erythromycin, fusidic acid, neomycin, puromycin, spectinomycin, tetracyclin, novobiocin, nalidixic acid, and penicillin). The *left panels* show the FH analysis of sporulation genes under 25 antibiotics from different classes, not including the former set, at 10 min of exposure (**a**, **b**), after 40 min of exposure (**c**, **d**), and after 80 min of exposure to all antibiotics (**e**, **f**). For visualization purposes, the correlation lines connecting the genes were set by correlation threshold greater than 0.4 and smaller than −0.4

In Fig. 10.15, we show the resulting holographic networks for the three time points after the exposure. Here the main change is detected between the 10 and 40 min time points. At 80 min after exposure, there is an additional effect, as *abrB* and *rok* become functionally more similar (closer in the 3D PCA space) and show higher correlations between them (Fig. 10.16).

The analysis of the response of competence genes to specific case of DNA topology antibiotics shows that *degU* had very high correlation to *comK* and *comG*. This increase in correlation can indicate the advance of synchronization of the competence system, as a result of exposure to antibiotics.
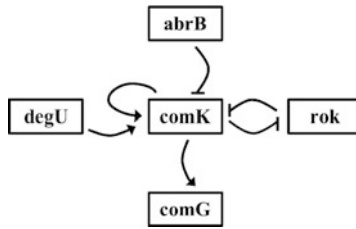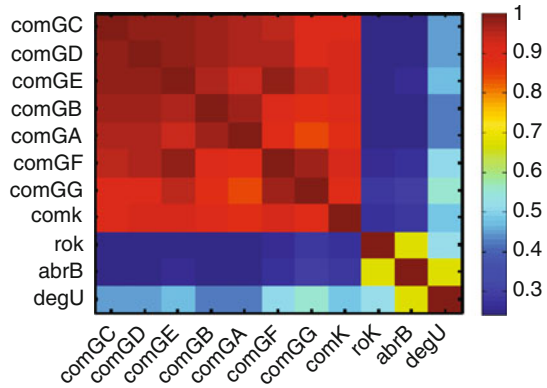
Fig. 10.13 A representation of the known regulation relations in our regulation gene network (Errington 1993; Grossman 1995; Stragier and Losick 1996). *Arrows* and *blunt arrows* represent positive and negative regulation, respectively. At the center of the chart lies the auto-regulatory gene *comK*. AbrB and Rok are repressors of *comK*, and DegU is a positive regulator. ComK is a positive regulator of the *comG* operon (comprised of the genes *comGA* to *comGG* and *yqze*) and a repressor of the *rok* gene

10.5 Time Progress of Cannibalism Gene Network

Cannibalism refers to behavior that was observed in *B. subtilis* cells in which the Spo0A-ON (sporulating) cells in the population trigger the lysis of nonsporulating bacteria (Spo0A-OFF cells) via the elaboration of a killing factor and a toxin (Gonzalez-Pastor et al. 2003). The Cannibalism network consists of two operons – *skf* and *sdp*. *skf* operon is involved in cannibalism and the production of an extracellular killing factor during sporulation (Fig. 10.17). This operon, through its products SkfE and SkfF, also confers resistance to the killing factor. The second operon, *sdp*, is controlled by SpoOA. SdpC is responsible for producing an extra-cellular factor that acts as a signaling protein among bacteria. SdpC strongly controls the transcription of a two-gene operon, *sdpR* and *sdpI*, located immediately downstream of the *sdp* operon. *sdpI-sdpR* use as an immune genes at the producing bacteria and also inhibit sporulation at other bacteria.
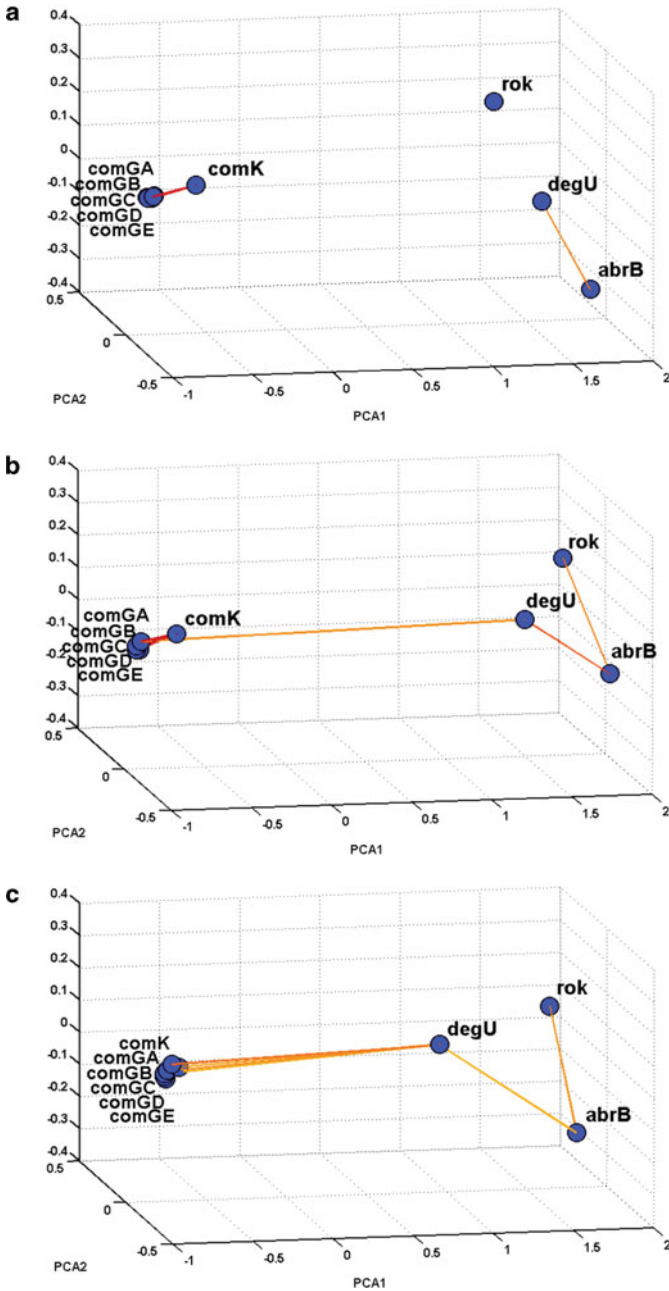
Fig. 10.14 Holographic network of competence genes at three time points after exposure. (**a**) After 10 min of exposure, (**b**) after 40 min of exposure, and (**c**) after 80 min of exposure to all antibiotics. For visualization purposes, the correlation lines connecting the genes were set by correlation threshold greater than 0.4 and smaller than −0.4

Fig. 10.15 Normalized correlation matrix for the competence genes, for all three time points, under exposure to the "topo" antibiotics
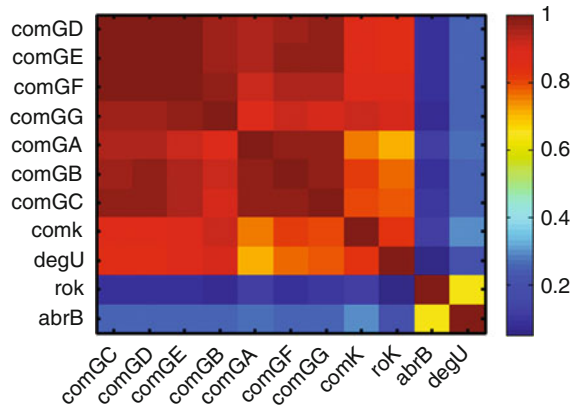
The cannibalism gene network includes the genes *skfA*, *skfB*, *skfC*, *skfD*, *skfE*, *skfF*, *skfG*, *skfH*, *sdpA*, *sdpB*, *sdpC*, *sdpR*, and *sdpI*. In Fig. 10.18, we show the normalized correlation matrix for these genes for all three-time points. In Fig. 10.19, we present the cannibalism gene network at the three different time points. It is possible to observe several changes that are a result of the exposure time to the antibiotics. First, s*po0E* becomes negative correlated to the cannibalism operons (*skf*, *sdp*) and *abrB*; Second, the two cannibalism operons become more correlated; third, *sigH* becomes more correlated to *spo0A*; and last, the gene *sdpI* dissociates from the cannibalism operons and becomes more correlated to *sdpR* and *spo0E*.
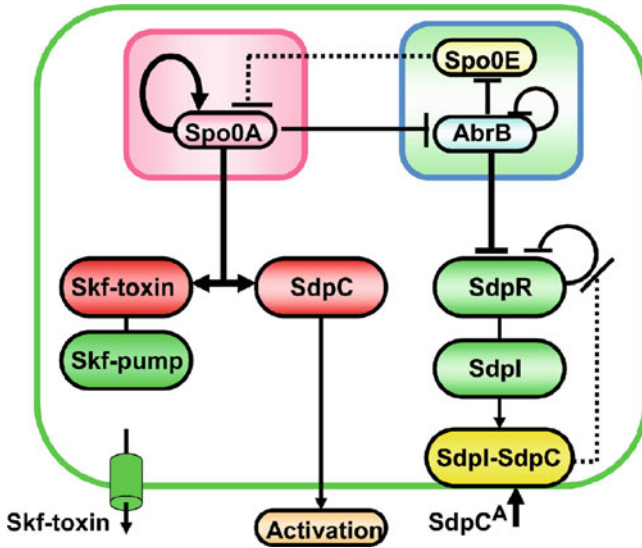
The cannibalism operons, *skf* and *sdp*, transcribe together at the beginning of sporulation. Although they are two separate operons, they cooperatively cause the lysis of Spo0A (off) cells. Our findings show that the two operons separate into two groups that are clustered together with high correlation. Moreover, the correlation becomes stronger as the exposure time to antibiotics grows. The remaining genes (*abrB*, *sigH*, *spo0A*, and *spo0E*) that participate in regulation of these operons are scattered in the PCA space.

Spo0A and abrB are mutual regulators of *skf* and *sdp*. The operons are positively regulated by Spo0A and negatively regulated by abrB (Rok is also a negative regulator of *sdp*, and PhoP is a positive regulator of *skf*). The regulation of *sdp* and *skf* by Spo0A is low-threshold activated, and under high-threshold *sdp* is repressed. This could provide a possible explanation to why *skf* and *sdp* do not have high correlation to *spo0A*. On the other hand, *sigH*, which induce sporulation, is very close to *spo0H*, as was already observed in the analysis of the sporulation gene network.

The gene *spo0E* becomes in time negatively correlated to the *sdp* genes. The function of the two cannibalism operons and Spo0E delays entry into sporulation. However, their role is different. Spo0E dephosphorylates Spo0A and thus maintains a low level of phosphorylated Spo0A. On the other hand, SdpC induces SdpR, which turns on the operons for ATP synthetase (atp), and for lipid catabolism
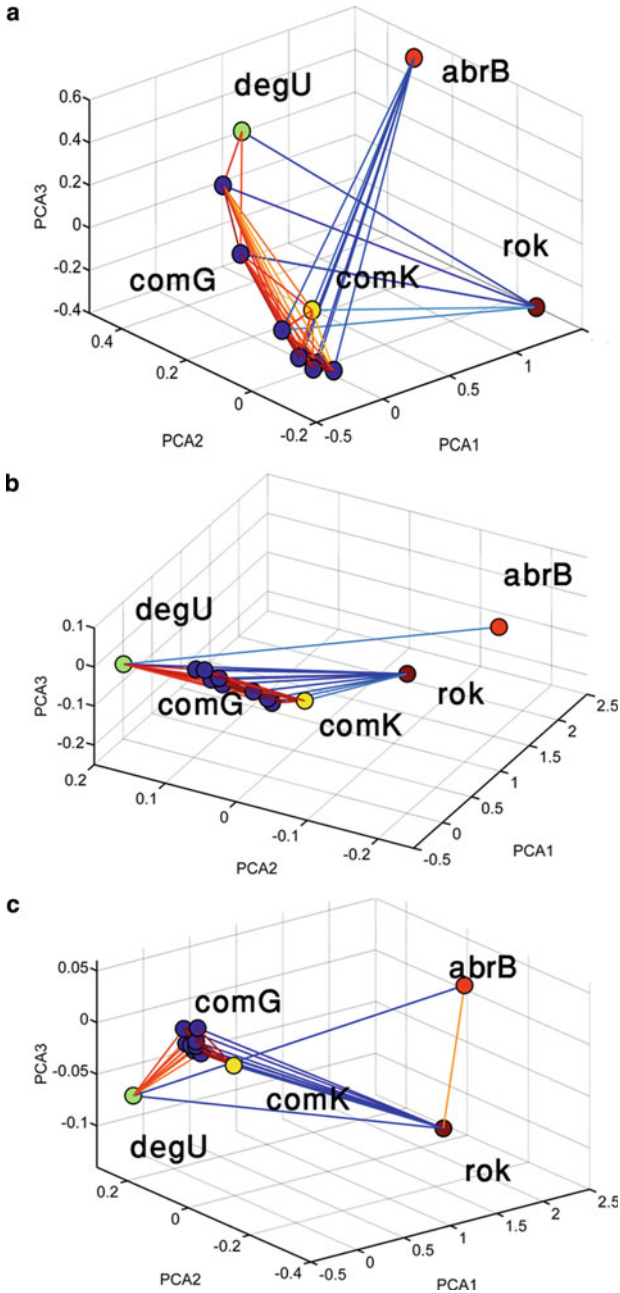
**Fig. 10.16** Time progress of competence genes in response to "topo" antibiotics. The correlation matrices for the competence gene network for three time stages, projected on the joint FH holographic 3D of the PCA space for all time points (Fig. 10.15). (**a**) After 10 min of exposure, (**b**) after 40 min of exposure, and (**c**) after 80 min of exposure to antibiotics interfering with DNA topology

Fig. 10.17 The cannibalism regulatory network. One way of coping with transition state stress conditions in SpoOA-active cells is delaying sporulation by switching on two operons, *skf* and *sdp*. The *skf* operon is involved in the production of an extracellular killing factor (*skf*-toxin). Its other two products, SkfE and SkfF (*skf*-pump), antagonize the lethal action of the killing factor in the SpoOA-active cells. In contrast, cells containing inactive SpoOA are killed by the secreted killing factor, resulting in the release of nutrients that can be consumed by SpoOA-active cells. In both active and inactive SpoOA cells, a gene product of the *sdp* operon (SdpC) switches on expression of the transcription factor sdpR. This factor seems to delay sporulation in SpoOA-active cells, probably by activating lipid catabolism and ATP-producing enzymes (Engelberg-Kulka and Hazan 2003). Moreover, SdpC, which is a toxin, binds to SdpI, thereby triggering sequestration of SdpR at the membrane and activating *sdpRI* transcription. Increasing amounts of SdpI and SdpR molecules trap all SdpC molecules. Free SdpI cannot sequester SdpR, which remains in the cytoplasm, where it shuts off transcription of *sdpRI* (Stragier 2006)
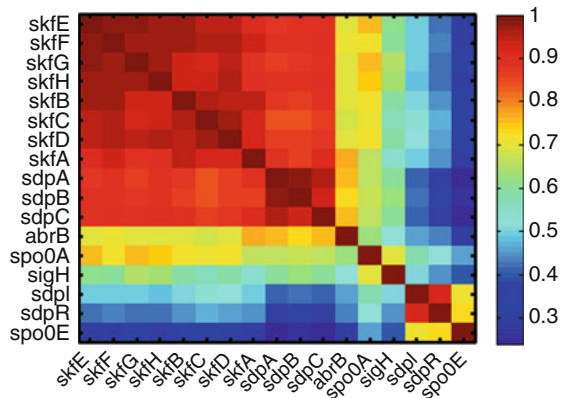
Fig. 10.18 Normalized correlation matrix for the cannibalism genes, for all three time points
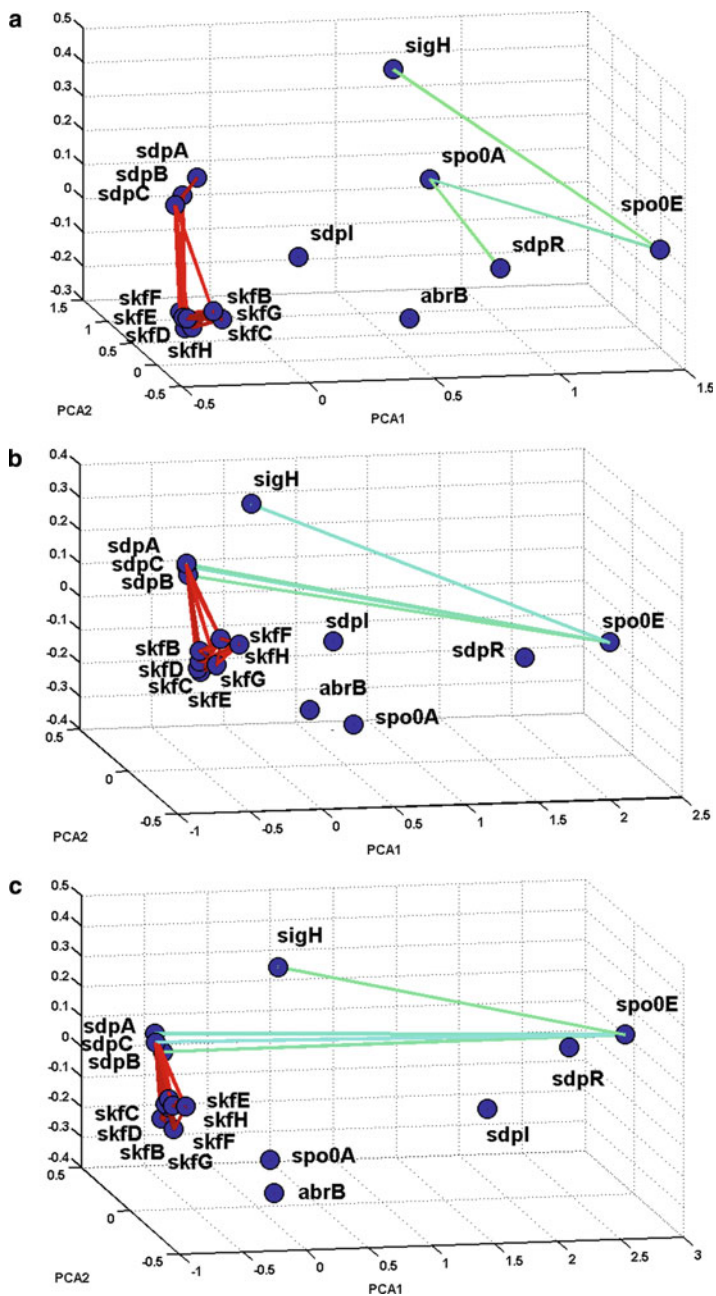
Fig. 10.19 Holographic network of cannibalism genes at three time points after exposure. (**a**) After 10 min of exposure, (**b**) after 40 min of exposure, and (**c**) after 80 min of exposure to all antibiotics. For visualization purposes, the correlation lines connecting the genes were set by correlation threshold greater than 0.8 and smaller than 0

enzymes (YusLKJ). The increase in energy production could be responsible for delaying sporulation, which is triggered by depletion of energy reserves. Since *skf* and *sdp* are activated by Spo0A, repression of the latter by Spo0E probably decreases the expression of those genes that results in negative correlation between *sdp* and *spo0E*.

sdpR, which is known to have a direct effect on sporulation repression, is located close to *spo0E*. Interestingly, *sdpI* that is transcribed on the same operon is not highly correlated with *sdpR*. In fact, *sdpI* changes its position in the PCA space over time, moving from the *skf*, *sdp* cluster at the 10 min interval, getting nearer to *spo0E* at the 80-min interval. The operon *sdpRI* is negatively regulated by AbrB and by auto-repression of SdpR. These two genes should be transcribed simultaneously, as there are no regulators or repressors between them; however, our findings show that they are not correlated (correlation below 0.4). This result could be due to an additional regulation that directly affects *sdpI*.

Once again, it was found that under antibiotic stress, genes that are related to the sporulation system become more synchronized over time and negatively correlated to repressors of sporulation, such as *spo0E*. Still, although the cannibalism network is related to the sporulation initiation phase, it is a separate regulatory network that is regulated also by other genes, so it is not entirely correlated with *spo0A* and *sigH*.

10.6   Discussion

In this chapter, we review and present a new, system-level analysis of the complex gene-network response of *B. subtilis* to environmental stress measured by DNA microarray chips. The method is based on the FH analysis that was originally developed for analyzing multichannels recordings of cultured neural networks activity and of recorded brain activity.

This method was used to analyze gene expression of *B. subtilis* exposed to sublethal levels of 37 different antibiotics. The matrices of gene correlations were computed and analyzed using the FH method. Then, relevant information was extracted from the matrices of normalized correlations by application of the PCA dimension reduction algorithm. The success in retrieving meaningful information proves the assumption that indeed valuable information is embedded in the correlations (similarities) between the expression profiles of different genes.

First, the ability of this method to act as an unsupervised or semi-supervised method was demonstrated. This was achieved by its successful identification and sorting of the genes into operons. This provides for a powerful application of this methodology, which is the identification of functionally related gene groups such as operons in an unsurprised way. In addition, this approach can also classify operons that are more cohesive from others.

Next, it was demonstrated that the approach can also be used to reveal the internal structure of the operons, thus relating the function (expression) to the form. A specific example was given, introducing how this method is able to deduce

information about the existence of unknown structural motif in the case of the *spoVAA* operon. These results demonstrated that the method could be used as a prediction tool to reveal functional similarities of unknown genes or operons.

Next we focus on the ability of the GH method to present the dynamical (functional) correlation motifs in well-known gene regulatory networks. We show that in competence, sporulation, and cannibalism networks, our method is capable of identifying previously known positive and negative gene interactions. Examples of such positive gene interactions include the interactions between the sporulation initiation genes (*kinABCE*, *sigH*, *spo0F*, and *spo0A*), cannibalism operons (*skf*, *sdp*), and late competence genes (*comK*, *comG* operon), and the interactions of *spo0E* are an example of such known negative gene interactions. We also show that under specific antibiotics, this method is able to identify unknown regulations, as was found in the case of *kinD* and *spo0B* that were negatively correlated to other sporulation genes. Furthermore, using this tool to study development across time, we observed specific motifs that were affected as a result of the exposure to the antibiotics. For examples, *sdpIR* that became dissociated from the other cannibalism operons, and *kinD* that became negatively correlated to other sporulation genes. These two examples possibly suggest unknown regulation mechanisms that were uncovered by the FH methodology. Other examples discussed here include the approaching of *spo0A* to sporulation genes in the PCA space, and the increase of correlation over time of *degU* to late competence genes, which possibly indicates a synchronization of the sporulation and competence system under exposure to antibiotic stress.

The affect of different antibiotics on the sporulation and competence networks has been investigated thoroughly in the past (Atsuhiro et al. 2003; Prudhomme et al. 2006). Here we demonstrated the effect of antibiotic stress on the three regulatory networks. Indeed, it was observed that there was a clear effect of specific antibiotics, such as protein synthesis inhibitors, on the interaction between the sporulation genes and DNA topology antibiotics on competence network. These findings are in agreement with previous reports that found the inhibition of sporulation under this class of antibiotics (Madi et al. 2008).

Under all antibiotics, for all three networks, genes that lead to sporulation, cannibalism, and competence become more correlated over time. The regulators Spo0A, DegS, SigH, and AbrB govern several transition state pathways, thus the increase in correlation of these regulators to the network might indicate the synchronization and regulation of the system in prolonging response to antibiotics stress. The increase over time of intra gene network correlation under specific conditions may indicate the functionality of the network in relation to these conditions.

Finally, we integrate the FH methodology with the well-known MST method (Mantegna 1999; Mantegna and Stanley 2000; Tumminello et al. 2007; Xu et al. 2001). Using MST methodology to investigate gene-expression data has been performed in the past (Xu et al. 2001) and can be used as we have shown to identify specific groups of genes. Combining these two methods can provide additional information regarding the makeup of the gene network. We perform each analysis separately, and then use the links between genes calculated from the MST analysis

to connect the genes in the reduced PCA space. This provides us with a simpler visualization of complex gene networks in a 3D in a sense that it sustains the information regarding the relationships between the genes given by the FH method while providing a smart well defined "filter" for the complexity of correlations information that we would like to present. Furthermore, it enabled us to make use of other network theory information methods, such as how central a gene is in the network (as shown qualitatively here).

In conclusion, we present here a new system-level analysis method for the investigation of gene-expression data. Possible applications of this methodology have been discussed in short, and some of the biological results obtained using this methodology were presented.

References

Albert A, Barabási AL (2002) Statistical mechanics of complex networks. Rev Mod Phys 47:47–97

Alon U, Barkai N, Notterman DA, Gish K, Ybarra S et al (1999) Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. Proc Natl Acad Sci USA 96:6745–6750

Atsuhiro A, Eiji K, Michihiro H, Mitsuo O (2003) Inhibition of *Bacillus subtilis* aprE expression by lincomycin at the posttranscriptional level through inhibition of ppGpp synthesis. J Biochem 134:691–697

Azevedo V, Sorokin A, Ehrlich SD, Serror P (1993) The transcriptional organization of the *Bacillus subtilis* 168 chromosome region between the spoVAF and serA genetic loci. Mol Microbiol 10:397–405

Baruchi I, Ben-Jacob E (2004) Functional holography of recorded neuronal networks activity. Neuroinformatics 2:333–352

Baruchi I, Grossman D, Volman V, Shein M, Hunter J et al (2006) Functional holography analysis: simplifying the complexity of dynamical networks. Chaos 16(1):015112

Ben-Dor A, Shamir R, Yakhini Z (1999) Clustering gene expression patterns. J Comput Biol 6:281–297

Burbulys D, Trach K, Hoch J (1991) Initiation of sporulation in *B. subtilis* is controlled by a multicomponent phosphorelay. Cell 8:545–552

Caldwell R, Sapolsky R, Weyler W, Maile RR, Causey SC et al (2001) Correlation between *Bacillus subtilis* scoC phenotype and gene expression determined using microarrays for transcriptome analysis. J Bacteriol 183:7329–7340

Chander P, Halbig KM, Miller JK, Fields CJ, Bonner HK et al (2005) Structure of the nucleotide complex of PyrR, the pyr attenuation protein from *Bacillus caldolyticus*, suggests dual regulation by pyrimidine and purine nucleotides. J Bacteriol 187:1773–1782

Claverys J, Prudhomme M, Martin B (2006) Induction of competence regulons as a general response to stress in gram-positive bacteria. Annu Rev Microbiol 60:451–475

Coronnello C, Tumminello M, Lillo F, Micciche S, Mantegna RN (2005) Sector identification in a set of stock return time series traded at the London Stock Exchange. Acta Phys Pol B 36:2653–2679

Dahl MK, Msadek T, Kunst F, Rapoport G (1992) The phosphorylation state of the DegU response regulator acts as a molecular switch allowing either degradative enzyme synthesis or expression of genetic competence in *Bacillus subtilis*. J Biol Chem 267:14509–14514

de Jong H (2002) Modelling and simulation of genetic regulatory systems: a literature review. J Comput Biol 9:67–103

Donetti L, Hurtado PI, Munoz MA (2005) Entangled networks, synchronization, and optimal network topology. Phys Rev Lett 95:188701

Dowds B, Murphy P, McConnell D (1987) Relationship among oxidative stress, growth cycle, and sporulation in *Bacillus subtilis*. J Bacteriol 169:5771–5775

Eisen M, Spellman P, Brown P (1998) Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci USA 95:14863–14868

Engelberg-Kulka H, Hazan R (2003) Microbiology. Cannibals defy starvation and avoid sporulation. Science 301:467–468

Errington J (1993) *Bacillus subtilis* sporulation: regulation of gene expression and control of morphogenesis. Microbiol Rev 57:1–33

Everitt BS (1993) Cluster analysis. Edward Arnold, London

Fawcett P, Eichenberger P, Losick R, Youngman P (2000) The transcriptional profile of early to middle sporulation in *Bacillus subtilis*. Proc Natl Acad Sci USA 97:8063–8068

Fuchs E, Ayali A, Ben-Jacob E, Boccaletti S (2009) Formation of synchronization cliques during development of modular neural networks. Phys Biol 6(3):036018

Fujita M, Losick R (2005) Evidence that entry into sporulation in *Bacillus subtilis* is governed by a gradual increase in the level and activity of the master regulator Spo0A. Genes Dev 19:2236–2244

Furey TS, Cristianini N, Duffy N, Bednarski DW, Schummer M et al (2000) Support vector machine classification and validation of cancer tissue samples using microarray expression data. Bioinformatics 16:906–914

Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M et al (1999) Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 286:531–537

Gonzalez-Pastor JE, Hobbs EC, Losick R (2003) Cannibalism by sporulating bacteria. Science 301:510–513

Graham RL, Hell P (1985) On the history of the minimum spanning tree problem. IEEE Ann Hist Comput 7:43–57

Grossman A (1995) Genetic networks controlling the initiation of sporulation and the development of genetic competence in *Bacillus subtilis*. Annu Rev Genet 29:477–508

Hamoen LW, Smits WK, de Jong A, Holsappel S, Kuipers OP (2002) Improving the predictive value of the competence transcription factor (ComK) binding site in *Bacillus subtilis* using a genomic approach. Nucleic Acids Res 30:5517–5528

Hamoen LW, Van Werkhoven AF, Venema G, Dubnau D (2000) The pleiotropic response regulator DegU functions as a priming protein in competence development in *Bacillus subtilis*. Proc Natl Acad Sci USA 97:9246–9251

Hansen P, Jaumard B (1997) Cluster analysis and mathematical programming. Math Program 79:191–215

Hartigan JA (1975) Clustering algorithms. Wiley, New York

Hartuv E, Shami R (2000) A clustering algorithm based on graph connectivity. Inf Process Lett 76:175–181

Herwig R, Poustka AJ, Müller C, Bull C, Lehrach H et al (1999) Large-scale clustering of cDNA-fingerprinting data. Genome Res 9:1093–1105

Hoch J (1993) Regulation of the phosphorelay and the initiation of sporulation in *Bacillus subtilis*. Annu Rev Microbiol 47:441–465

Hutter B, Schaab C, Albrecht S, Borgmann M, Brunner NA et al (2004) Prediction of mechanisms of action of antibacterial compounds by gene expression profiling. Antimicrob Agents Chemother 48(8):2838–2844

Jiang M, Shao W, Perego M, Hoch J (2000) Multiple histidine kinases regulate entry into stationary phase and sporulation in *Bacillus subtilis*. Mol Microbiol 38:535–542

Jonas R, Holt S, Haldenwang W (1990) Effects of antibiotics on synthesis and persistence of sigma E in sporulating *Bacillus subtilis*. J Bacteriol 172:4616–4623

Kocabas P, Calik P, Calik G, Ozdamar TH (2009) Microarray studies in *Bacillus subtilis*. Biotechnol J 4:1012–1027

Kruskal JB (1956) On the shortest spanning subtree of a graph and the traveling salesman problem. Proc Am Math Soc 7:48–50

Lorenz M, Wackernagel W (1994) Bacterial gene transfer by natural genetic transformation in the environment. Microbiol Mol Biol Rev 58:563–602

Madi A, Friedman Y, Roth D, Regev T, Bransburg-Zabary S et al (2008) Genome holography: deciphering function-form motifs from gene expression data. PLoS ONE 3:e2708

Madi A, Hect I, Bransburg-Zabary S, Merbl Y, Zucker-Toledano M et al (2009) Organization of the autoantibody repertoire in healthy newborns and adults revealed by system level informatics of antigen microarray data. Proc Natl Acad Sci USA 106:14484–14489

Mandal M, Boese B, Barrick JE, Winkler WC, Breaker RR (2003) Riboswitches control fundamental biochemical pathways in *Bacillus subtilis* and other bacteria. Cell 113:577–586

Mantegna RN (1999) Hierarchical structure in financial markets. Eur Phys J B 11:193–197

Mantegna RN, Stanley HE (2000) An introduction to econophysics: correlation and complexity in finance. Cambridge University Press, Cambridge, UK

Mirkin B (1996) Mathematical classification and clustering. Kluwer Academic Publishing, Dordrecht, The Netherlands

Molle V, Fujita M, Jensen ST, Eichenberger P, Gonzalez-Pastor JE et al (2003) The Spo0A regulon of *Bacillus subtilis*. Mol Microbiol 50:1683–1701

Newman MEJ (2003) The structure and function of complex networks. SIAM Rev 45:167–256

Ogura M, Tanaka T (2000) *Bacillus subtilis* comZ (yjzA) negatively affects expression of comG but not comK. J Bacteriol 182:4992–4994

Ortega GJ, Sola RG, Pastor J (2008) Complex network analysis of human ECoG data. Neurosci Lett 147:129–133

Parego M, Hanstein C, Welsh KM, Djavakhishvili T, Glaser P et al (1994) Multiple protein-aspartate phosphatases provide a mechanism for the integration of diverse signals in the control of development in *B. subtilis*. Cell 79:1047–1055

Prudhomme M, Attaiech L, Sanchez G (2006) Antibiotic stress induces genetic transformability in the human pathogen Streptococcus pneumoniae. Science 313(5783):89–92

Quinlan J (1992) Programs for machine learning. Morgan Kaufmann, San Mateo, CA

Rodionov DA, Vitreschak AG, Mironov AA, Gelfand MS (2003) Regulation of lysine biosynthesis and transport genes in bacteria: yet another RNA riboswitch? Nucleic Acids Res 31:6748–6757

Rogers P, Liu T, Barker K, Hilliard G (2007) Gene expression profiling of the response of Streptococcus pneumoniae to penicillin. J Antimicrob Chemother 59:616–626

Roggiani M, Dubnau D (1993) ComA, a phosphorylated response regulator protein of *Bacillus subtilis*, binds to the promoter region of srfA. J Bacteriol 175:3182–3187

Rumelhart DE, Hinton GE, Williams RJ (1986) Learning internal representations by error propagation. In: Rumelhart DE, McClelland JL (eds) Parallel distributed processing, vol 1. MIT Press, Cambridge, MA, pp 318–362

Ruzal S, Sanchez-Rivas C (1998) In *Bacillus subtilis* DegU-P is a positive regulator of the osmotic response. Curr Microbiol 37:368–372

Schultz D, Wolynes PG, Ben Jacob E, Onuchic JN (2009) Deciding fate in adverse times: sporulation and competence in *Bacillus subtilis*. Proc Natl Acad Sci USA 106:21027–21034

Serizawa M, Yamamoto H, Yamaguchi H, Fujita Y, Kobayashi K et al (2004) Systematic analysis of SigD-regulated genes in *Bacillus subtilis* by DNA microarray and Northern blotting analyses. Gene 329:125–136

Serror P, Sonenshein A (1996) CodY is required for nutritional repression of *Bacillus subtilis* genetic competence. J bacteriol 178:5910–5915

Serror P, Wong K, Sonenshein A (2001) *Bacillus subtilis* CodY represses early-stationary-phase genes by sensing GTP levels. Genes Dev 15:1093–1103

Shapira Y, Kenett DY, Ben-Jacob E (2009) The index cohesive effect on stock market correlations. Eur Phys J B 72:657–669

Sharan R, Maron-Katz A, Shamir R (2003) CLICK and EXPANDER: a system for clustering and visualizing gene expression data. Bioinformatics 19:1787–1799

Sonenshein A (2000) Control of sporulation initiation in *Bacillus subtilis*. Curr Opin Microbiol 3:561–568

Stephenson K, Hoch JA (2002) Evolution of signalling in the sporulation phosporelay. Mol Microbiol 2:297–304

Stragier P (2006) To kill but not be killed: a delicate balance. Cell 124:461–463

Stragier P, Losick R (1996) Molecular genetics of sporulation in *Bacillus subtilis*. Annu Rev Genet 30:297–341

Tamayo P, Slonim D, Mesirov J, Zhu Q, Kitareewan S et al (1999) Interpreting patterns of gene expression with self-organizing maps. Proc Natl Acad Sci USA 96:2907–2912

Tumminello M, Coronnello C, Lillo F, Micciche S, Mantegna RN (2007) Spanning trees and bootstrap reliability estimation in correlation-based networks. Int J Bifurcat Chaos 17:2319–2329

Vazquez-Ramos J, Mandelstam J (1981) Inhibition of sporulation by DNA gyrase inhibitors. Microbiology 127:11–17

Veening JW, Hamoen LW, Kuipers OP (2005) Phosphatases modulate the bistable sporulation gene expression pattern in *Bacillus subtilis*. Mol Microbiol 56:1481–1494

West DB (2001) An introduction to graph theory. Prentice-Hall, Englewood Cliffs, NJ

Xu Y, Olman V, Xu D (2001) Minimum spanning trees for gene expression data clustering. Genome Inform 12:24–33