Biochimica et Biophysica Acta 1810 (2011) 924–932

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbagen

Review

Information theory based approaches to cellular signaling☆

Christian Waltermann, Edda Klipp *

Theoretical Biophysics, Humboldt-Universität zu Berlin, Invalidenstr. 42, 10115 Berlin, Germany

ARTICLE INFO

Article history:
Received 1 November 2010
Received in revised form 8 June 2011
Accepted 13 July 2011
Available online 23 July 2011

Keywords:
Systems biology
Signaling pathway
Signal integration
Shannon entropy
Mutual information
Sensitivity analysis

ABSTRACT

Background: Cells interact with their environment and they have to react adequately to internal and external changes such changes in nutrient composition, physical properties like temperature or osmolarity and other stresses. More specifically, they must be able to evaluate whether the external change is significant or just in the range of noise. Based on multiple external parameters they have to compute an optimal response. Cellular signaling pathways are considered as the major means of information perception and transmission in cells.
Scope of Review: Here, we review different attempts to quantify information processing on the level of individual cells. We refer to Shannon entropy, mutual information, and informal measures of signaling pathway cross-talk and specificity.
Major Conclusions: Information theory in systems biology has been successfully applied to identification of optimal pathway structures, mutual information and entropy as system response in sensitivity analysis, and quantification of input and output information.
General Significance: While the study of information transmission within the framework of information theory in technical systems is an advanced field with high impact in engineering and telecommunication, its application to biological objects and processes is still restricted to specific fields such as neuroscience, structural and molecular biology. However, in systems biology dealing with a holistic understanding of biochemical systems and cellular signaling only recently a number of examples for the application of information theory have emerged. This article is part of a Special Issue entitled Systems Biology of Microorganisms.

1. Introduction

Since the structure and wiring of biochemical signaling systems are better understood [1,2] than their dynamics, the yet unknown role of connecting and shared elements discovered in the interactome is currently subject of great interest in systems biology. In particular the dynamic and static mechanisms of signal integration and signal separation at work in signaling pathways are subject to intense research (examples include [3]). An overview of general signaling mechanisms is provided in [4,5]. In the recent years, however, an increasing amount of research has been devoted to computational frameworks which aim at the quantification of signaling cross-talk and specificity [6,7]. On the experimental frontier the focus shifted to the perturbation of individual signaling pathway components in signaling networks to study their effect on the larger contextual network [2,8]. Experimental techniques and computational frameworks will continue to generate qualitative and quantitative insight and hypotheses about information flows into cells and within cells. It is nevertheless noticeable that the advantages of widely employing the mathematically stringent framework of Shannon's information theory [9] have been largely overlooked. For instance in neuroscience information theory is routinely applied to compare the information content of experimentally measured neural responses with model predictions [10] and the maximally transferable information according to Shannon's coding theorems [11].

However, in recent years also in systems biology new approaches have been based on information theory to contribute to elucidating the characteristics of signal transfer in cellular systems [12–14]. Tkačik and colleagues studied noise levels in gene expression and asked for optimal conditions in the dynamic range of transcription factors and regulatory modules to enhance information flow [15]. Lestas and colleagues used information theory to compute lower limits for noise compression in biological networks that can be achieved by control mechanisms such as negative feedback [16]. Tostevin and ten Wolde estimated the mutual information between input and output trajectories using a Gaussian model and applied it to information transmission in the *E. coli* chemotaxis network [17]. Graeber et al. employ the principle of entropy maximization to inform oncogenicity predictions based on site-specific phosphorylation events in the relevant signaling cascades [18]. Ziv et al. investigated network motifs and topologies and quantified the quality of biochemical computation [19].

This mini-review gives an overview over information theory frameworks, provides an introduction into the field, and comments on

☆ This article is part of a Special Issue entitled Systems Biology of Microorganisms.

* Corresponding author. Tel.: +49 30 2093 9040.
E-mail addresses: christian.waltermann@hu-berlin.de (C. Waltermann), edda.klipp@hu-berlin.de (E. Klipp).

0304-4165/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.bbagen.2011.07.009

future potential applications of information theory based approaches regarding cellular signaling pathways as well as on its limitations. We will also highlight examples, which we consider to be seminal works regarding the diverse impact of information theory on systems biology:

- The information theoretical analysis of possible mechanisms of minimization of signal interference in the marine bacteria *Vibrio harveyi* quorum sensing network [14];
- The improvement of information transmission by a Fus3p mediated negative feedback in the *S. cerevisiae* pheromone pathway [20];
- The analysis of the effects of cooperativity on the assembly of macromolecular complexes in signaling systems [21]; and
- The application of an information-theoretical sensitivity analysis approach to the NFκB signaling pathway [13].

2. The fundamentals of information theory

2.1. Shannon's general communication model, Shannon entropy, information and biochemical systems

The basic paradigm of information theory is that data — an encoded message — is transferred *via* a noisy channel from a sender to a receiver, where it is decoded back to a message, which can for instance automatically determine a response (Shannon's general communication model [9], Fig. 1). Information theory in general aims at quantifying information, signal processing operations, and their limits. For a more general review on the topic, we refer to recent textbooks (e.g. [22]).

When applied to cellular signaling systems, any input into a pathway system can be regarded as a sender. The receiver is the pathway output and the intermediate biochemical reactions form the noisy channel. Typical examples for the sender/channel/receiver architecture are the system of ligand–receptor/signaling pathway/changes of gene expression level, or DNA-damage/protein–protein-interactions/cell cycle checkpoint.

Models of biochemical circuits are often set up under the assumption that pathway component concentrations or activity levels might vary due to extrinsic and intrinsic noises[23,24]. Intrinsic noise arises mainly from transcriptional variations, translational variations, and the noisy properties of chemical reactions, whereas extrinsic noise is caused by environmental influences such as temperature, pH, or ligand concentration variations and its effects on levels and activation states of pathway components. These variations induce a finite amount of uncertainty to pathway response such that output activation depends on pathway inputs in a probabilistic fashion.

The input itself can exhibit stochasticity since both the environmental variables which lead to receptor activation and the amount of receptor activity when given a certain environmental input are subject to fluctuations.

For illustration of these concepts we introduce three very simple example pathway architectures, the logical OR, AND, and XOR gates (Fig. 2). Examples of pathways processing several inputs at the same time include the yeast high osmolarity glycerol (HOG) pathway which is responsible for adaptation to osmotic stress and employs the two sensors Sho1 and Sln1 [25], the *S. cerevisiae* cell wall integrity (PKC) pathway which can be activated by up to five different receptors [26], or the PKA signaling pathway in yeast which may be activated by both the Ras signaling branch and *via* the Gpa1 receptor [27].

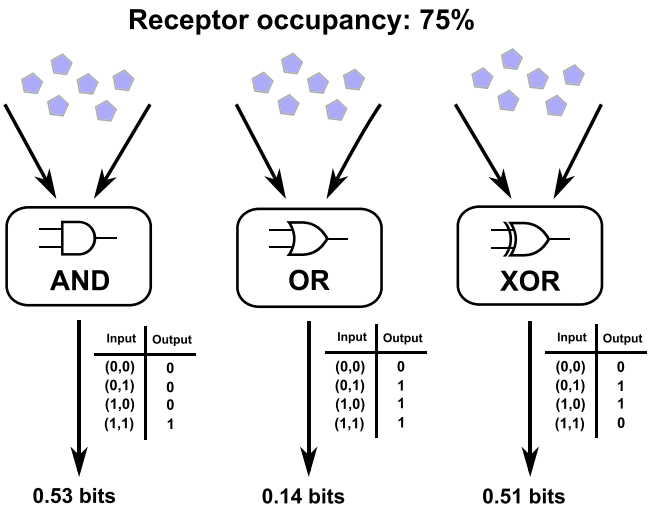

Fig. 2. Example to demonstrate the application of information theory to three small Boolean pathway models, the AND, OR, and XOR gates. The two inputs represent receptors, which bind to a chemical agent in the environment of the circuits with a probability of 75%, thus defining the input distributions to the circuit. The circuits are modeled as noisy channels, which are faithful to their truth tables with a probability of 90%. With a probability of 10%, the response is the opposite of the outcome according to the truth table. With the XOR architecture, the highest amount of information is transmitted at 75% receptor occupancy. The truth tables and the mutual information transmitted can be found at the circuit outputs.

Our logical gate example shall illustrate the simplified case of pathway architecture with two inputs and one output. We assume that our circuits are exposed to an agent that is able to activate each pathway receptor with a probability of 75%. This premise reflects the stochastic nature of environmental variables and receptor activation. Furthermore, we make the assumption that the circuits have a transmission reliability of 90%, i.e. the probability with which they follow the deterministic truth tables provided in Fig. 2. With a probability of 10% our model pathways compute the outcome inaccurately which means that the outcome according to the deterministic truth table is simply inverted. This behavior was introduced to emulate the effect of biochemical noise.

To portrait the usage and the meaning of the newly introduced information theoretical concepts we will now start applying them to the Boolean example circuits.

To achieve quantification of information, several measures can be used. The most common one is defined by the Shannon entropy

$$H = E[I(X)] = -\sum_{x \in X} p(x)\log_2 p(x) \tag{1}$$

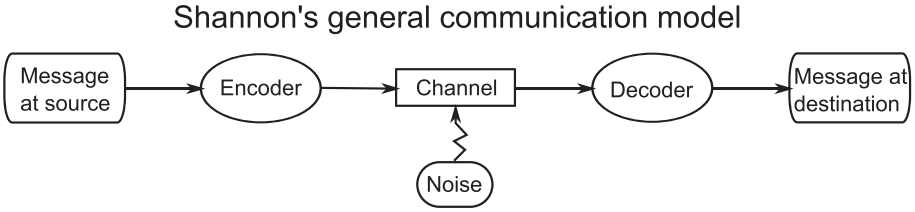

Fig. 1. Schematic representation of Shannon's general communication model. At the source, a message is encoded and sent via a noisy channel to the source where it is decoded. This process can be quantified with the help of the Shannon entropy or other adequate information measures. Analogies to biochemical systems can be drawn: pathway inputs encode information, which is sent *via* the network in the form of biochemical reactions. Extrinsic noise, intrinsic noise, and the balance of biochemical species inside the pathway are sources of noise in a biochemical communication channel. At the pathway outputs information is decoded and can elicit a programmed response. The response can be detected experimentally as a signal that the message has been received.

as the expected value of the information content $I(X)$ of a distribution of random variables $x \in X$. The $p(x_i)$are the probabilities of observing the quantity $x \in X$. The Shannon entropy measures the average amount of information which is coded by the set X. The (self-) information content or "surprisal" $I(x)$of the quantity x is defined as

$$I(x) = -\log_2 p(x). \tag{2}$$

When, like in this case, a logarithm to the base "2" is used, the unit of Shannon entropy and self-information is "bit". The Shannon entropy defines an upper limit on lossless compression of data, information transmission, and communication.

We can now start applying information measures to our three circuits and compute the Shannon entropies (Eq. (1)) of the input sets and the output sets of the circuits as a measure for the average amount of information which is known about a set of variables.

Let X be the set of possible input configurations to our logical pathway models. Then the set X comprises the following elements: (0,1) and (1,0) if only one of the pathway branches is active, (1,1) if both branches get activated, and (0,0) if none of the branches is active. Since we assumed that the activation probability of an individual branch in the presence of the activating agent is 75%, the probability of the state (1,1) that both branches are active is $(75\%)^2 = 56.25\%$, the probabilities for (0,1) and (1,0) that only one branch is active is $75\% \cdot 25\% = 18.75\%$, and the probability of the state (0,0) that none of the pathway branches is stimulated is $(25\%)^2 = 6.25\%$. In possession of the input probabilities we can now directly compute the Shannon entropy of the input set X using Eq. (1):

$H_{Input} = H(X) = -5625 \cdot \log2(0.5625) - 2 \cdot 0.1875 \cdot \log2(0.1875) - 0.0625 \cdot \log2(0.0625) = 1.63$ bit. Since the probabilities of the inputs being active in the absence of the agent are all zero under our assumptions, they do not contribute to the input entropy. Per definition it holds $0 \cdot \log2(0) = 0$.

Analogously, let Y be the set of the possible output states 1 and 0. To compute the probabilities of the output states we first need to introduce the concept of conditional probabilities $p(y|x)$. The conditional probability $p(y|x)$ is the probability that the pathway is in a certain output state y given that its input is x. For instance for our sample XOR gate the probability that the output y is zero given that the input x is (0,0) is 90% since according to the XOR truth table the noise-free output would be 0, however we assumed a transmission reliability of 90%. According to the same scheme the conditional probabilities for all input and output combinations can be determined. For the output probabilities $p(y)$ we obtain

$$p(y) = \sum_{x \in X} p(y|x) \cdot p(x) \tag{3}$$

if we sum up all products of input probabilities with conditional probabilities for all possible pathway inputs $x \in X$. The individual summands are called joint probabilities $p(x,y) = p(y|x) \cdot p(x)$.

The value $p(y = 0)$ for the XOR model is 60% according to Eq. (3). For the AND and OR gates we obtain values of 45% and 15%, respectively. Accordingly, $p(y = 1)$ is 40% for the XOR model, 55% for the AND model, and 85% for the OR model. Applying the Shannon entropy Eq. (1) to the output sets Y of the logical gate models yields values of $H(Y) = 0.97$ bit for the XOR gate, $H(Y) = 0.99$ bit for the AND gate, and $H(Y) = 0.61$ bit for the OR model. These results state that the OR gate model output Y can encode significantly less information than the AND and XOR gates with the input set X to all three gates being identical.

Furthermore, the example calculations demonstrated how sensitively the input entropies $H(X)$ depend on the assumptions made regarding the receptor occupancy: A binary tuple of the form (a,b) which we use to model the inputs into the example circuits can normally encode 2 bit of information if all possible input configurations are equally likely. Under the assumption, however, that the ligand in the environment is present and has a pathway branch activation probability of 75% this value is reduced to a mere 1.63 bit. Our example highlights the importance of correct identification of the input set X and the probability distribution $p(x)$ on X. This issue will be further discussed along with the literature examples.

Similarly, the output entropies $H(Y)$ depend sensitively on the conditional probabilities $p(y|x)$ of the pathway system. The $p(y|x)$ ultimately encodes the properties of the model system. How they can be obtained will also be subject to discussion in the literature example section.

Although we now know the input and output entropies and with them the maximal amount of information encoded in the input and output configurations of the three logical gates, we do not know to what extent the output sets Y depend on the input sets X in each of the three cases. In other words, we have not yet determined, how much can be learned about the presence of the chemical agent in the environment of the signaling systems when the output of the pathway system is known. Information theory provides the concept of mutual information to tackle this question.

Mutual information between two sets of variables X and Y is defined formally as

$$I(X,Y) = H(X) + H(Y) - H(X,Y), \tag{4}$$

where $H(X,Y)$ is the joint entropy between the sets X and Y. The joint entropy is associated to the joint probabilities $p(x,y)$ of the variables x in X and y in Y occurring together and is mathematically defined as

$$H(X,Y) = -\sum_{x \in X} \sum_{y \in Y} p(x,y) \log_2 p(x,y). \tag{5}$$

$H(X,Y)$ measures the average amount of information which is not available when observing variables x in X and y in Y together. It is, therefore, maximal if X and Y are independent, which actually means that $p(x,y) = p(x) \cdot p(y)$. It can also be demonstrated that in this case of independence of X and Y it holds $H(X,Y) = H(X) + H(Y)$ by applying the entropy Eq. (1) and elementary rules for the calculation with logarithms. Else, $H(X,Y)$ is smaller than the sum of the individual entropies.

This insight provides a graphic interpretation of the mutual information Eq. (4): The more X and Y are dependent on each other in statistical sense the lower the joint entropy $H(X,Y)$ becomes and the closer the mutual information $I(X,Y)$ gets to its maximal value $H(X) + H(Y)$, i.e. the sum of information content that input and output of the system are able to encode.

Back in the picture of biochemical pathways where X denotes the set of possible inputs and Y the set of observed pathway output states, mutual information therefore quantifies how well information is transmitted through the biochemical channel. To illustrate the concept of mutual information using our sample Boolean pathway architectures, the missing joint entropies $H(X,Y)$ need to be evaluated from the joint probabilities

$$p(x,y) = p(y|x) \cdot p(x). \tag{6}$$

The joint probabilities quantify to which extent individual pathway inputs X coincide with possible system outputs. For instance, the probability $p((0,0),0)$ for the XOR gate — the probability that both input branches are off and the XOR pathway is off at the same time in the presence of the activating agent — is given by $p((0,0),0) = p(0|(0,0)) \cdot p((0,0)) = 90\% \cdot 6.25\% = 5.625\%$. Analogously, all other computations are carried out, and for the joint entropy $H(X,Y)$ a value of 2.09 bit is acquired after the application of Eq. (5). It can be shown that $H(X,Y) = 2.09$ bit for the AND and OR gates as well. Now, Eq. (4) is used to calculate the mutual information for the XOR gate: It holds $I(X,Y) = H(X) + H(Y) - H(X,Y) = 1.63$ bit $+ 0.97$ bit $- 2.09$ bit $= 0.51$ bit. For the AND circuit the mutual information

between input and output is $I(X,Y) = 0.53$ bit, and for the OR circuit it is a mere 0.15 bit.

How can we interpret the results for the Boolean example circuits? Clearly, the mutual information between input X and output set Y in the OR circuit is significantly lower than the respective values for the AND and XOR gates. The reason is that the OR gate is *on* under almost all input configurations and its output therefore encodes much less information which is also reflected in the lower output entropy $H(Y)$ for the OR gate in comparison with the other gates.

To complete the theoretical introduction of information theoretical concepts we introduce alternative ways to calculate mutual information $I(X,Y)$ and summarize the probability relationships. By explicitly expanding the Shannon entropies in Eq. (4) depending on the probabilities $p(x)$, $p(y)$ and joint probabilities $p(x,y)$ and by subsequently applying calculation rules for the logarithm we can obtain the mutual information directly as a function of these probability distributions:

$$I(X,Y) = \sum_{x\in X}\sum_{y\in Y} p(x,y)\log_2\left(\frac{p(x,y)}{p(x)p(y)}\right). \tag{7}$$

Here again $p(x,y) = p(y|x)\cdot p(x)$ is the joint probability of X and Y; $p(x)$ is provided by the input distribution X, and $p(y)$ can be calculated *via* $p(y) = \sum_{x\in X} p(y|x)\cdot p(x)$.

For biochemical systems with known (stochastic) differential equations, $p(y|x)$ can be directly obtained from the solution of the system by statistical sampling (Monte Carlo) or by analytical approximations such as the Gaussian channel [12].

It can furthermore be demonstrated that mutual information and Shannon entropies are linked by

$$I(X,Y) = H(Y) - H(Y|X), \tag{8}$$

where $H(Y)$ reflects the information content of the output variable Y and $H(Y|X)$ the conditional entropy. $H(Y)$ is given in Eq. (1); $H(Y|X)$ can be obtained as follows:

$$H(Y|X) = -\sum_{x\in X} p(x)\sum_{y\in Y} p(y|x)\log_2 p(y|x). \tag{9}$$

If the pathway system has more than one input, it is often only of interest, how an output Y changes with respect to a combination (X_1, X_2) of inputs, as in our Boolean circuit example. In this case, an adequate conditional probability distribution $p(y|(x_1,x_2))$ should be determined from the solution.

The information theoretical analysis of a conditional probability distribution can be conducted equivalently to the case with just one input as discussed above. If the pathway has multiple inputs, or if the pathway is a result of converging branches upstream, the information transmitted *via* the pathway might depend on what branch is used as the communication input, such that for each individual input the conditional probabilities might take a different value.

Thus, for any given input X_i, a probability $P(Y|X_i)$ of output Y conditional on the input can be determined. Analogously, for any combination of inputs $X_1, X_2,\ldots, X_n$, the probabilities of output Y can be given, for example: $P(Y|X_1,X_2)$.

Signal integration can also be analyzed in terms of how much of the information transmitted from just the individual inputs differs from the information when all inputs are combined (total information), or to what extent information transmission is limited by noise inside the pathway. This way, conclusions about signal interference and other communication limiters can be drawn from the information theoretical analysis of the signaling network.

However, if the input distributions X_1, X_2, …, X_n are not strictly independent and the system is, therefore, capable of transmitting information from one input to another, the multivariate mutual information approach [28] is a more appropriate choice. This approach allows the computation of information transfer (mutual information) from any input or output to any other input or output as follows:

$$I(X_1,\ldots,X_{n+1}) = I(X_1,\ldots,X_n) - I(X_1,\ldots,X_n|X_{n+1}), \tag{10}$$

where $I(X_1,\ldots, X_n|X_{n+1})$ is a general conditional mutual information, while $I(X_1,\ldots, X_n)$ is a n-th order mutual information, employed in an iterative approach.

3. Examples for successful application of information theoretical concepts in systems biology

Below, we describe a number of cases where information theory has been used to further understand signal transmission in quantitative and qualitative terms.

3.1. Alternative, informal measures of information transmission: specificity and cross-talk in signaling cascades

The notion of signaling pathways has been instrumental to conceptualize the view that specific cascades of events contribute to sensing and processing of external signals. However at some point it became clear that there are no isolated routes — pathways — but signaling networks and that more than one input signal can be processed by a signaling network [29]. In this context, one may ask how different routes maintain specificity from the signal down to the cellular response and how to measure cross-talk between pathways. Komarova et al. [6] have suggested the measure of specificity

$$S_1 = \frac{Y_2|X_1}{Y_1|X_1} \tag{11}$$

to quantify how much of input X_1 to signal cascade 1 ends up in the output Y_2 of cascade 2 (Fig. 3). Eq. (11) calculates the ratio of the gene or generally the output response Y_1 of a native input X_1 into branch 1 of a signaling system to the spurious output Y_2 in branch 2 given a defined input X_1 and no native input into branch 2.

Analogously the measure fidelity

$$F_1 = \frac{Y_1|X_1}{Y_1|X_2} \tag{12}$$

tells us how much the output of cascade 1 is determined by the input X_1 of cascade 1 compared to input X_2 of cascade 2.

Schaber and colleagues [7] took into account that different inputs of signaling cascades might be of different nature (e.g. different ligands) or even have different physical units (such as temperature or osmolarity) and may, therefore, be not directly compatible. To measure the combined effect of different types of input signal $X_k$on output quantities Y, the intrinsic and extrinsic specificities

$$S_i(k) = \frac{Y(X_k)}{Y(X_1,\ldots X_n)} \text{ and } S_e(k) = \frac{Y(X_1,\ldots,X_{k-1},X_{k+1},\ldots,X_n)}{Y(X_1,\ldots X_n)} \tag{13a, b}$$

were defined. If both intrinsic and extrinsic specificity are larger than 1, we have a case of mutual signal inhibition, if both are smaller than one, it is mutual signal amplification; $S_i(k)>1$ and $S_e(k)<1$ means that the intrinsic signal dominates (which is in accordance with the intuitive understanding of a signaling pathway); $S_i(k)<1$ and $S_e(k)>1$ indicate that the (so-perceived) extrinsic signal dominates the pathway output, which is counterintuitive. The concept was applied to the yeast pheromone and filamentous growth pathways.

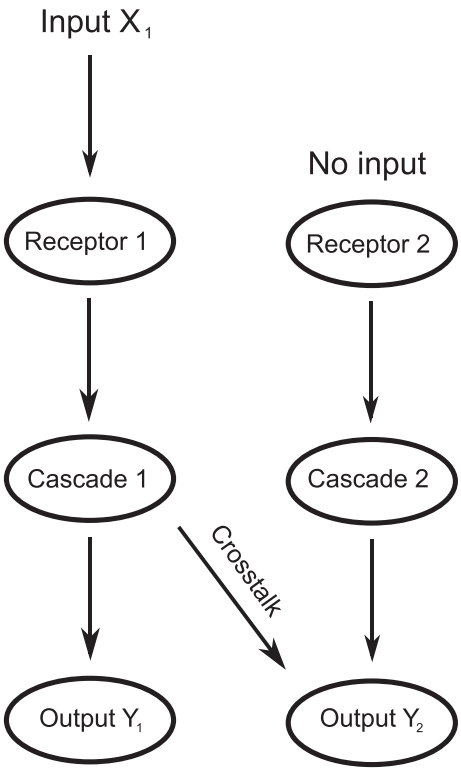

Fig. 3. Crosstalk between two branches "1" and "2" in a biochemical signaling network. The presented frameworks by Kolmorova et al. and Schaber et al. aim at quantifying to what extent signal leakage from branch 1 to branch 2 occur and study what effect the incorporation of feedbacks, compartments, and scaffolds might have on the respective crosstalk measures. For mathematical details please refer to main text. We propose that the information theoretical treatment of these frequent motifs could increase the precision of the estimation of cross-signaling and, therefore, facilitate the generation of experimentally testable hypotheses.

The precise relationship between these informal measures of specificity, fidelity, and cross-talk on the one hand and information theoretical quantities on the other hand still remains to be elucidated and should be subject to future studies. It is of interest to what extent and under which conditions fidelity and specificity correlate with mutual information since their computation is often much easier than the determination of the Shannon entropy.

The presented approaches appear useful to quantify signal pathway crosstalk for specific given values of stimulus and input and provide an intuitive understanding of the effect of crosstalk. They don't require knowledge about input distributions as necessary in information theory. However, in principle, measures (11) to (13) could also be extended for application to such distributions. In their current version, they are easy to calculate for input and output values obtained from experiments.

3.2. Signal separation in overlapping pathways employed in quorum sensing

One recent example for the effective usage of information theory is a study published by Mehta et al. [14]. The authors tackle the question, how the marine bacterium *V. harveyi* integrates and separates signals elicited by three different autoinducer molecules (AI-1, AI-2, and CAI-1) involved in its quorum sensing network and, even more importantly, how it can focus its attention to different autoinducers during different phases of growth or altered environmental conditions. Bacteria use quorum sensing to change their behavior depending on the population density and colony size. In *V. harveyi*, the three autoinducers are sensed by three distinct receptors, however the signals of these receptors converge on one single quorum sensing pathway. (Fig. 4 left). Only AI-1 is produced by *V. harveyi*; CAI-1 and AI-2 are released by other *Vibrio* bacteria and a wide range of non-*Vibrio* species, respectively. This way, *V. harveyi* can obtain information about its own population via AI-1, about other *Vibrio* species via CAI-1, and the total bacterial population through AI-2. However, it remained unclear by what mechanism *V. harveyi* is able to distinguish different input signals in single pathway architecture.

To elucidate this question, the authors developed an analytical steady-state response function $Y = f(X_1, X_2)$ for the *V. harveyi* quorum sensing pathway with a focus on only two of the inputs: The LuxN receptor, which senses the presence of the autoinducer molecule AI-1, and the LuxPQ input, which is able to detect AI-2. Both receptors are active when not exposed to autoinducer molecules and phosphorylate the phosphorelay protein LuxU which in turn indirectly suppresses the expression of quorum sensing related genes in *V. harveyi*. In the

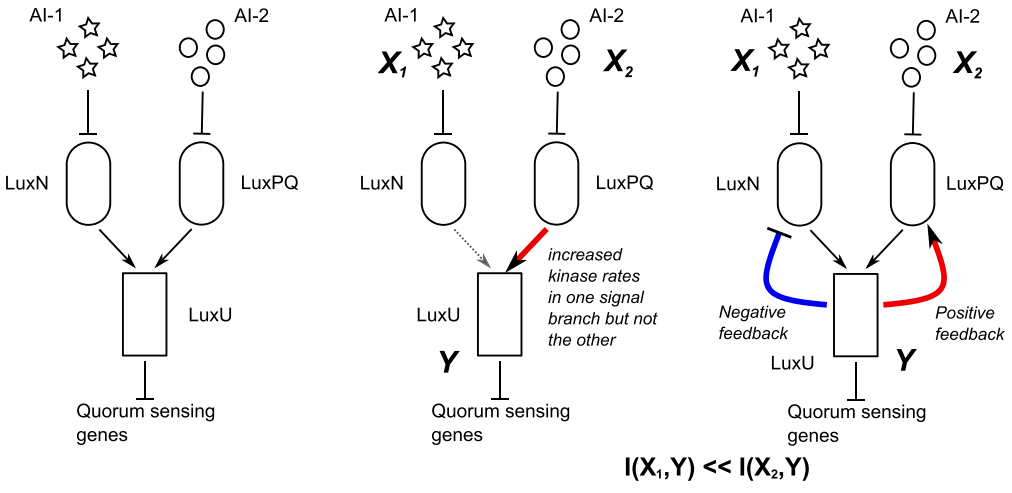

Fig. 4. Information processing in quorum sensing of *V. harveyi*. *Right panel*: Simplified quorum sensing network. The presence of the autoinducer molecule AI-1 inhibits the LuxN receptor input branch and the AI-2 molecule exercises its inhibition on the LuxPQ input branch of the *V. harveyi* signaling network. Both receptors converge on a shared phosphorelay protein LuxU, which they are able to phosphorylate, when the autoinducer molecules are not present. LuxU then subsequently suppresses the expression of quorum sensing genes. *Center panel*: *V. harveyi* could increase the information flow through one of the branches of the quorum sensing network by manipulating the kinase strength in one input branch. In this example, the kinase strength within input branch 2 is increased. This way, the bacteria are able to direct their focus to a specific input. *Right panel:* possible strategies of bacterial input manipulation: A positive feedback on the receptor number in the X_2 input branch increases information transfer through branch 2. The mutual information $I(X_2,Y)$ is also increased by a negative feedback on the receptor number in branch 2 of the quorum sensing pathway.

presence of autoinducer molecules, the corresponding receptors become inactive and the repression on the quorum sensing genes is released. From the response function $Y=f(X_1, X_2)$, which quantifies the system output, the authors compute the noisy (Gaussian channel) transfer function $P(Y|X_1, X_2)$. It corresponds to the system-dependent conditional probability $p(y|x)$ in Section 2, which enables the authors to obtain the mutual information measures $I(X_1,Y)$ and $I(X_2,Y)$ for each pathway branch, and $I((X_1,X_2),Y)$ for the total information transmission using the integral versions of the mutual information Eq. (10). For the total information, the authors find that its transmission is limited by biochemical noise.

Subsequently, they evaluate the information transfer from the individual autoinducer input branches $I(X_1,Y)$ and $I(X_2,Y)$in the quorum sensing pathway depending on the following assumptions:

1. The bacteria are able to tune the kinase strengths in the individual input branches.
2. Presence of positive feedback of the pathway output on the receptor number in only one input branch.
3. Presence of negative feedback of the pathway output on the receptor number in only one input branch.

The information theoretical analysis reveals that by increasing the kinase strength in one pathway branch the information transfer through that branch can be increased (Fig. 4, middle). Analogously, a positive feedback of the pathway output on the receptor number in one branch of the network increases the information transfer *via* that branch compared to the opposite branch; whereas a negative feedback in one branch enhances the information transfer through the opposite branch (Fig. 4 left).

The authors conclude that all three strategies could be employed by bacteria to reduce signal interference in quorum sensing circuits with a multiple input — one output architecture. In *V. harveyi*, hypotheses derived from the theoretical analysis of the network are already supported by experimental observations: Equal kinase activities where observed in the LuxN and LuxPQ branches [30], such that a permanent imbalance of kinase strength can be ruled out. However, the precise mechanisms of kinase strength tuning as proposed by Mehta et al. [14] still remains to be experimentally validated.

The study by Mehta et al. highlights possible future applications of information theory in systems biology: The elucidation of cellular decision making processes using techniques with high precision and the ability to generate predictions which are largely independent from the choice of the model parameters. Often, when dealing with a class of problems in which the prior (input) distribution is known and the question to be answered is what kind of pathway or network architecture is optimal to let the cellular system adapt to the given prior, the approach by Mehta et al. represents a valuable to existing system identification strategies.

3.3. Quantification of information exchange via macromolecular complex formation in signaling pathways

In another recent study [21], information theory is used to quantify the information exchange in macromolecular complexes by linking the mutual information to cooperativity. The authors assume that any protein complex ABC can be considered as a noisy communication channel, where the steady state concentration of free proteins, partially assembled complexes, and the full complex as well as the degradation rates of these species define the rate at which information can be transferred through the complex in a biochemical pathway (Fig. 5 upper left). A new measure of cooperativity is introduced in the following way: The maximal amount of information — that is the mutual information — which is exchanged between an upstream ligand A to a downstream effector C *via* an intermediate protein B through formation of the macromolecular complex ABC can be considered as an alternative measure for the degree of cooperativity in the protein complex assembly process.

To be able to evaluate the mutual information in the complex assembly, the input (prior) probabilities $p(A=0)$, $p(A=1)$, the output probabilities $p(C=0)$, $p(C=1)$, and the joint probabilities $p(A=0,C=1)$, $p(A=0,C=0)$, $p(A=1,C=0)$, $p(A=1,C=1)$ are computed from the equilibrium concentrations [A], [B], [C], [AB], [BC], and [ABC] of the system (for an example please refer to Fig. 5 lower left), which can be determined using an adequate ordinary differential equation (ODE) model of the complex formation process under appropriate numerical values for the relevant binding affinities and the overall concentrations of the species. $A=0$, or $A=1$ describes if the protein

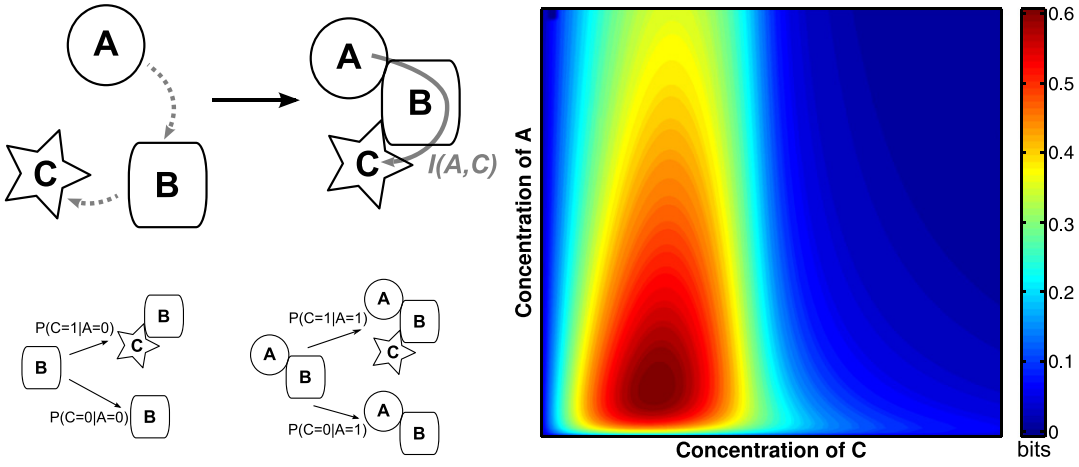

Fig. 5. Information processing in protein complex formation. *Upper left panel:* Complex formation between the upstream activator A, an intermediate protein B, and the downstream target C can be considered information transfer, which can be quantified using the mutual information framework. *Lower left panel:* definitions of conditional probabilities used to compute the mutual information $I(A,C)$. The probability for the case $C=1|A=0$ is derived from the probability of C binding to B under the premise that A is not bound to B. Consequently, the probability for the case $C=0|A=0$ is obtained from the probability that C does not bind to B, when A is not bound. On the other hand, the probability for the case $C=1|A=1$ is derived from the probability that C binds to B with A bound. The probability $C=0|A=1$ is simply the probability that C does not bind to the AB complex. The first two probabilities define the "noisy channel" property of the complex formation process, whereas the latter two characterize the input distribution. *Right panel:* arbitrary example for a mutual information heat map which can be obtained by using the approach. It shows the information in bit transferred through complex formation per input bit for the formation of the ABC complex in dependence on the total concentration of A and the total concentration of C with given B. Maximal information is transmitted, when the concentrations of A and C are adequately balanced.

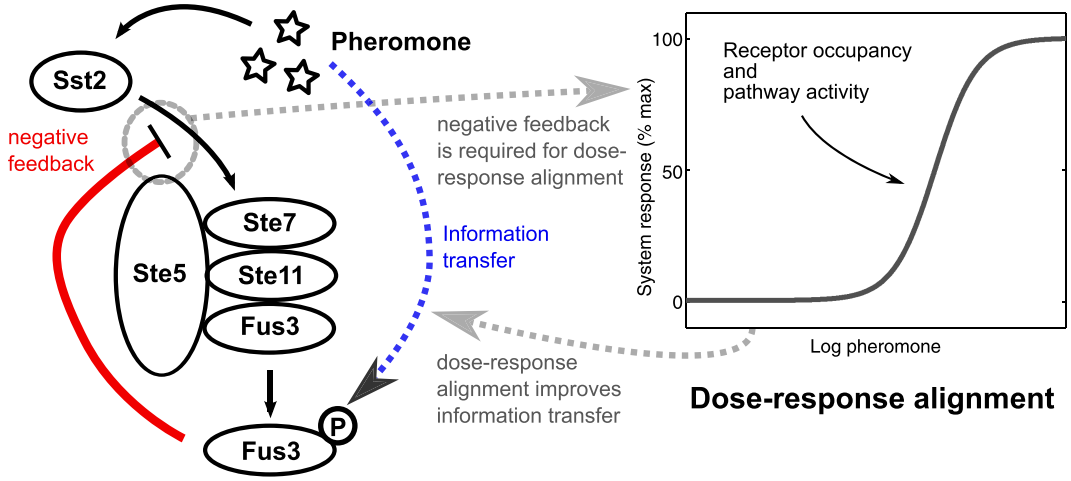

Fig. 6. Hypothesized increase of mutual information due to dose–response alignment. Yu and coworkers were able to identify a negative feedback in the yeast pheromone pathway which is responsible for the feature of dose–response alignment. In dose–response alignment, the output of the system and the concentrations of components which evoke pathway activity demonstrate a highly similar behavior. In budding yeast, this is the case for the pheromone concentration and the activity of the MAP kinase Fus3p in the pheromone pathway.

component is bound to B or not. Analogously, $C=0$, or $C=1$ refers to the condition if C is bound to B or not.

The mutual information exchanged over the complex is then computed by equation

$$I(A,C) = H(A) + H(C) - H(A,C) \tag{14}$$

from the corresponding probability distributions analogously to Eq. (4). Fig. 5 (right panel) contains an illustration of the possible readout for the information exchange *via* the macromolecular complex for ranges of total concentrations of the input protein A and the output protein C at fixed concentration of the central component B.

The authors find that low information exchange takes place, when the concentration of the B component is out of balance with the concentration of the ABC complex as a whole or when the concentration of intermediate complexes AB or AC is out of balance with the concentration of the formed ABC complex at steady state. It is noteworthy that suboptimal mutual information is caused by concentration balances exclusively. No additional noise is added.

The authors exemplify their approach with the p27 regulatory pathway, which controls degradation of the Cdk2 inhibitor p27 during cell cycle progression. It should be noted, that the approach can be easily scaled up to complexes with more than three components using a multivariate mutual information approach.

The study by Lenaerts et al. is not only of interest because it presents mutual information as an alternative way of quantification of cooperativity as a contribution to structural biology, but also because it can be used to quantify the information transfer *via* protein complex formation in a pathway context. In a previous study, it was shown that the information flow through a protein can also be understood on a more fundamental level by analyzing mutual information on the protein domain level [31].

3.4. A Fus3p-mediated negative feedback in the pheromone pathway improves information transmission

Yu and co-workers [20] could demonstrate that Fus3p mediates a fast-acting negative feedback loop in the pheromone pathway of *S. cerevisiae*. It achieves an adjustment of the downstream response of the system to the receptor–pheromone binding upstream, a phenomenon the authors name “dose–response alignment” (Fig. 6). They argue that dose–response alignment improves the information transfer in the pheromone pathway and other signaling systems; however, they do not provide a thorough mathematical analysis based on information theory. Intuitively, such a conclusion could be drawn because one could expect that a negative feedback stabilizes system outputs given substantial variations of system components and might have a dampening influence on noise [32–34]. Furthermore, the authors make it plausible that negative feedback-mediated dose–response alignment does not reduce the dynamic range of the system. They argue that both these features can be contributors to an improvement of information transmission through the pathway.

Increased mutual information due to dose response alignment could have far-reaching implications: In drug discovery and design, a potential drug could disrupt one of the mechanisms leading to dose–response alignment in a target system required for cell proliferation and thereby decrease the information transfer through that system. This would likely decrease the proliferation rate.

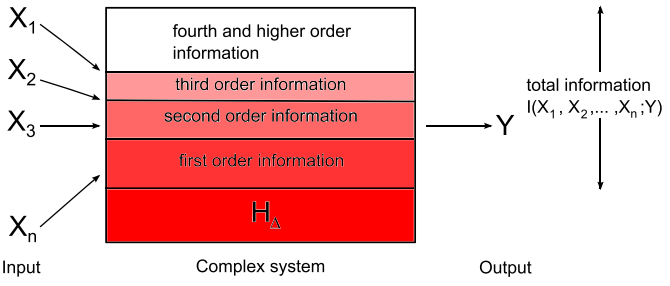

Fig. 7. A complex biochemical system might have multiple inputs $X_1,\ldots X_n$. Global sensitivity analysis aims at quantifying the contributions of each of those inputs as well as the contributions of a combination of inputs on an output Y. With the help of information theory, a framework can be defined which allows the assessment of sensitivities in terms of transmission of information. In the framework presented by Lüdtke and colleagues, first order sensitivities correspond to the mutual information, which is transmitted from one input X_i to the output Y. The second order sensitivities assess the mutual information from two inputs X_i and X_j combined on Y. Higher order sensitivities are defined correspondingly. The authors show that the total information is made up from the sum of all these contributions. However, due to statistical sampling, a term H_Δ remains, which cannot be computationally assessed.

3.5. An information-theory based sensitivity analysis

The last example presents the work of Lüdtke et al. [13] who developed a sensitivity analysis framework based on information theory (see Fig. 7). They start with the premise, that any given deterministic complex system can be analyzed as a stochastic system by random sampling of the input space. A random distribution of the input space consequently translates to a random distribution of outputs by application of the response function of the system or its numerical solution and can then be analyzed within the framework of information theory.

The goal of a global sensitivity analysis is the assessment of contributions of input distributions to the output of the system when taking into account a wide range of inputs within the operating range of the system. Thus, mutual information provides an excellent framework to compute these contributions as the amount of information $I(X_i,Y)$ transferred from one input X_i to an output Y. The authors define this approach as first-order sensitivity analysis.

However, since it is often of interest what the impact of pairwise inputs on an output is — a problem frequently encountered when dealing with biochemical systems — a second order sensitivity analysis is defined by the conditional mutual information between two inputs X_i and X_j and the output Y:

$$I\left(X_i,X_j|Y\right) = \sum_{y\in Y} p(y) \sum_{x\in X} p\left(x_i,x_j|y\right)\log_2 \frac{p\left(x_i,x_j|y\right)}{p(x_i|y)p\left(x_j|y\right)} \tag{15}$$

The second order sensitivities relate to first order influences, joint first and second order influences, and the mutual information between the inputs as follows:

$$I\left(X_i,X_j|Y\right) = I\left(X_i,X_j;Y\right) - I(X_i,Y) - I\left(X_j,Y\right) + I\left(X_i,X_j\right) \tag{16}$$

Here, $I(X_i, X_j; Y)$ is the joint mutual information between (X_i, X_j) and Y. In the sensitivity interpretation this is the joint first and second order sensitivity; $I(X_i, Y)$ and $I(X_j, Y)$are first order sensitivities, and $I(X_i, X_j)$is the mutual information between the two inputs i and j quantifying the correlation between them.

The authors subsequently define higher order sensitivities from the multivariate mutual information approach in an analogous manner. They demonstrate that the total mutual information in the system amounts to the sum of sensitivities of all orders minus a residual H_Δ which results from the errors made by discretization (Fig. 7).

Of particular interest might be the fact, that this information-theory based sensitivity analysis approach is applicable to a wide range of system outputs and responses. It is not just limited to the activities of systems components, but can be employed to global systems properties such as periods of oscillations as long as an adequate deterministic mathematical description of the system output features can be formulated.

4. Discussion

The presented applications of information theory in systems biology can be summarized under the following goals:

1. Identification of optimal pathway structures, properties, and conditions with respect to their ability to maximize mutual information through a larger biochemical network (e. g. [14,15,20]).
2. Formulation of balance between individual pathway components such that mutual information is maximal (e. g. [21,31]). This approach indirectly allows predictions regarding the contextual network.
3. Mutual information/multivariate information as a global system response, and input entropies as global system parameters (e. g. [13]).
4. Quantification of how much information contained in inputs, e.g. a ligand concentration, is coded by the system output, e.g. gene transcription level, i.e. what does the cell really learn about the presence of a stimulus (e.g. [14]).

There are some general problems regarding information theory which might become relevant when applying it to biochemical systems:

1. Entropy is non-unique. This means that for instance the quantification of information $H(X)$ of an input X into a pathway depends on the probability distribution $p(x)$ on X. Often the receptor activation as depending on ligand concentration is not precisely known, but needs to be estimated in order to obtain $p(x)$. In this case a bias might be introduced into the input entropy which will ultimately affect $H(X)$. Also when estimating $p(x)$ from experimental data, the experimentalist's choice of the stimulus/input set X will determine $H(X)$.
2. Defining the channel. This includes the questions: What is defined as input set X, what will be the output Y, how will the transfer function $p(y|x)$ be set up.
3. Time resolution of the transfer function $p(y|x)$. Adaptation processes in response to stimuli can change the parametrization of the mathematical model which is used as the basis for estimating $p(y|x)$. This adds additional complication to the computation of mutual information, for instance if a time delay should be added when calculating $I(X,Y)$.

Currently, information theory has not unfolded its full potential within systems biology and, particularly, as far as signal integration and differentiation is concerned. Nevertheless, the studies featured in this mini-review show that information theory approaches are capable of generating far-reaching testable hypotheses and contribute to a more holistic understanding of cellular signaling systems.

Other biological areas outside systems biology, particularly the fields of computational neuroscience, structural and molecular biology, employ information theory routinely when dealing with problems of high complexity. Frequent applications in neuroscience include the model of neural systems as optimized communication channels [10]. In structural biology, it is used for the evaluation of protein structure predictions [35] and in molecular biology for DNA and protein sequence analysis [36]. These examples are just few out of many.

On first sight information theory might have the appeal of extracting more abstract hypotheses about design principles of biochemical systems within a mathematically sound and complete framework. When applying information theory to systems biology, potential limitations can include the fact that setting up a reasonable transfer function $p(y|x)$ might be as challenging as setting up a conventional mathematical model for the intermediate steps between inputs and outputs of a studied biochemical system. The choice of parameters in the transfer function will influence the results of its information theoretical analysis equally strongly as it is the case for the simulation results of an ODE pathway model. However, new molecular biology techniques may allow direct measurements of the transfer function without knowing or modeling the details of the signaling pathways as shown, e.g., by experiments employing the transcription factor Bicoid as input and the expression levels of the gene Hunchback as output in early *Drosophila* development [15].

With the advent in qualitative and quantitative investigation of cellular signaling pathways and cellular information processing, it becomes both more interesting and more feasible to understand not

only whether information has been perceived, but how much information was transmitted, whether it was relevant, and which route it has taken through the numerous alternative pathways in a signaling network. Here, we have reviewed recent attempts to analyze and quantify the amount of information transmitted through cellular systems. Such studies pave the way for deeper quantitative understanding of the multifarious levels of regulation relevant for cellular signaling and information processing.

Acknowledgements

Work in E.K.'s lab is supported by the German Research Council (SFB 618, SFB 740) and the European Union 7th Framework Programme (project Unicellsys (HEALTH-2007-201142), BaSysBio (LSHG-CT-2006-037469) and Cellcomput (NEST-2005-Path-2-043310)).

References

[1] S. Bandyopadhyay, C.Y. Chiang, et al., A human MAP kinase interactome, Nat. Methods 7 (10) (2010) 801–805.
[2] K. Kaizu, S. Ghosh, et al., A comprehensive molecular interaction map of the budding yeast cell cycle, Mol. Syst. Biol. 6 (2010) 415.
[3] M.N. McClean, A. Mody, et al., Cross-talk and decision making in MAP kinase pathways, Nat. Genet. 39 (3) (2007) 409–414.
[4] C. Waltermann, E. Klipp, Signal integration in budding yeast, Biochem. Soc. Trans. 38 (5) (2010) 1257–1264.
[5] L. Bardwell, X. Zou, et al., Mathematical models of specificity in cell signaling, Biophys. J. 92 (10) (2007) 3425–3441.
[6] N.L. Komarova, X. Zou, et al., A theoretical framework for specificity in cell signaling, Mol. Syst. Biol. 1 (2005) 0023.
[7] J. Schaber, B. Kofahl, et al., A modelling approach to quantify dynamic crosstalk between the pheromone and the starvation pathway in baker's yeast, FEBS J. 273 (15) (2006) 3520–3533.
[8] M. Krantz, D. Ahmadpour, et al., Robustness and fragility in the yeast high osmolarity glycerol (HOG) signal-transduction pathway, Mol. Syst. Biol. 5 (2009) 281.
[9] C.E. Shannon, A mathematical theory of communication, Bell Syst. Tech. J. 27 (1948) 379–423.
[10] A. Borst, F.E. Theunissen, Information theory and neural coding, Nat. Neurosci. 2 (11) (1999) 947–957.
[11] M.J. Wainwright, Visual adaptation as optimal information transmission, Vision Res. 39 (23) (1999) 3960–3974.
[12] A.M. Walczak, G. Tkacik, et al., Optimizing information flow in small genetic networks. II. Feed-forward interactions, Phys. Rev. E Stat. Nonlin. Soft Matter Phys. 81 (4 Pt 1) (2010) 041905.
[13] N. Ludtke, S. Panzeri, et al., Information-theoretic sensitivity analysis: a general method for credit assignment in complex networks, J. R. Soc. Interface 5 (19) (2008) 223–235.
[14] P. Mehta, S. Goyal, et al., Information processing and signal integration in bacterial quorum sensing, Mol. Syst. Biol. 5 (2009) 325.
[15] G. Tkacik, C.G. Callan Jr., et al., Information flow and optimization in transcriptional regulation, Proc Natl Acad Sci USA 105 (34) (2008) 12265–12270.
[16] I. Lestas, G. Vinnicombe, et al., Fundamental limits on the suppression of molecular fluctuations, Nature 467 (7312) (2010) 174–178.
[17] F. Tostevin, P.R. ten Wolde, Mutual information between input and output trajectories of biochemical networks, Phys. Rev. Lett. 102 (21) (2009) 218101.
[18] T.G. Graeber, J.R. Heath, et al., Maximal entropy inference of oncogenicity from phosphorylation signaling, Proc. Natl. Acad. Sci. U. S. A. 107 (13) (2010) 6112–6117.
[19] E. Ziv, I. Nemenman, et al., Optimal signal processing in small stochastic biochemical networks, PLoS One 2 (10) (2007) e1077.
[20] R.C. Yu, C.G. Pesce, et al., Negative feedback that improves information transmission in yeast signalling, Nature 456 (7223) (2008) 755–761.
[21] T. Lenaerts, J. Ferkinghoff-Borg, et al., Information theoretical quantification of cooperativity in signalling complexes, BMC Syst. Biol. 3 (2009) 9.
[22] D. McKay, Information Theory, Inference, and Learning Algorithms, Cambridge University Press, Cambridge, 2003.
[23] M.B. Elowitz, A.J. Levine, et al., Stochastic gene expression in a single cell, Science 297 (5584) (2002) 1183–1186.
[24] P.S. Swain, M.B. Elowitz, et al., Intrinsic and extrinsic contributions to stochasticity in gene expression, Proc. Natl. Acad. Sci. U. S. A. 99 (20) (2002) 12795–12800.
[25] E. Klipp, B. Nordlander, et al., Integrative model of the response of yeast to osmotic shock, Nat. Biotechnol. 23 (8) (2005) 975–982.
[26] R. Rodicio, J.J. Heinisch, Together we are strong — cell wall integrity sensors in yeasts, Yeast 27 (8) (2010) 531–540.
[27] H. Tamaki, Glucose-stimulated cAMP-protein kinase A pathway in yeast *Saccharomyces cerevisiae*, J. Biosci. Bioeng. 104 (4) (2007) 245–250.
[28] W.J. McGill, Multivariate mutual information, Psychometrika 19 (1954) 97–116.
[29] K. Kaizu, H. Moriya, et al., Fragilities caused by dosage imbalance in regulation of the budding yeast cell cycle, PLoS Genet. 6 (4) (2010) e1000919.
[30] T. Long, K.C. Tu, et al., Quantifying the integration of quorum-sensing signals with single-cell resolution, PLoS Biol. 7 (3) (2009) e68.
[31] T. Lenaerts, J. Ferkinghoff-Borg, et al., Quantifying information transfer by protein domains: analysis of the Fyn SH2 domain structure, BMC Struct. Biol. 8 (2008) 43.
[32] A. Lobner-Olesen, Distribution of minichromosomes in individual *Escherichia coli* cells: implications for replication control, EMBO J. 18 (6) (1999) 1712–1721.
[33] A. Becskei, L. Serrano, Engineering stability in gene networks by autoregulation, Nature 405 (6786) (2000) 590–593.
[34] J. Paulsson, Summing up the noise in gene networks, Nature 427 (6973) (2004) 415–418.
[35] R. Swanson, M. Vannucci, et al., Information theory provides a comprehensive framework for the evaluation of protein structure predictions, Proteins 74 (3) (2009) 701–711.
[36] O. Weiss, M.A. Jimenez-Montano, et al., Information content of protein sequences, J. Theor. Biol. 206 (3) (2000) 379–386.