

Adaptive gene regulatory networks

F. STAUFFER¹ and J. BERG^{1,2(a)}

¹ *Institut für Theoretische Physik, Universität zu Köln - Zùlpicher Straße 77, 50937 Köln, Germany, EU*

² *Physikalisches Institut, Albert-Ludwigs-Universität Freiburg - Herrmann Herder-Str. 3, 79104 Freiburg, Germany, EU*

received 8 April 2009; accepted in final form 5 November 2009

published online 3 December 2009

PACS 87.16.Yc – Regulatory genetic and chemical networks

PACS 87.23.Kg – Dynamics of evolution

PACS 87.18.Sn – Neural networks and synaptic communication

Abstract – Regulatory interactions between genes show a large amount of cross-species variability, even when the underlying functions are conserved: there are many ways to achieve the same function. Here we investigate the ability of regulatory networks to reproduce given expression levels within a simple model of gene regulation. We find an exponentially large space of regulatory networks compatible with a given set of expression levels, giving rise to an extensive *entropy of networks*. Typical realisations of regulatory networks are found to share a bias towards symmetric interactions, in line with empirical evidence.

Copyright © EPLA, 2009

Introduction. – The expression of genes is regulated such that the right combinations of gene products are generated at the right time and place of an organism. Key regulators of gene expression are *transcription factors*, proteins which bind to specific sites on DNA and influence the expression of nearby genes. Typically, the expression of a gene is effected by a combination of several transcription factors, and conversely, a transcription factor regulates several genes. Expression levels can thus depend on the entire set of regulatory interaction between transcription factors and their target genes, referred to as a *regulatory network*. These intracellular reaction networks process extracellular information to induce specific gene expression patterns, allowing, for instance, the development of a complex body plan, or responses to external conditions.

Even though regulatory networks are tuned carefully to produce specific expression patterns, there are in general many networks fulfilling a regulatory task. One example is the control of mating type in different yeast species: The same set of genes controlled in *S. cerevisiae* by an activator which is upregulated in a certain state is controlled by a repressor which is downregulated in that state in *C. albicans* [1]. A second prominent example is the development of the anterior patterning in insect embryos, leading to the formation of the insect's head. The gene crucial to this process in the fruit fly *Drosophila*, called bicoid,

is absent in many other insects, where a combination of different genes take on the same task [2]. Even whole sets of genes which are co-expressed across the entire yeast family can have different regulatory interactions in different species [3]. Across bacteria, widespread 'tinkering' at the level of individual interactions is found, even though similar subnetwork topologies (network motifs) appear in organisms with similar lifestyle [4]. Source of these changing interactions is a rapid evolutionary turnover of transcription factor binding sites at the level of DNA sequences [5,6]. This can generate new regulatory interactions. A recent essay on the degeneracy of regulatory networks can be found in [7].

The large number of networks resulting in a viable organism (viable regulatory networks) does not imply that there are many dispensable interactions: viable networks may occupy only a small fraction of the set of all possible regulatory networks, and elaborate counter-changes may be needed to restore viability once an interaction has been altered. An analogy is the set of all RNA sequences which fold into a given secondary structure, which stretches across the entire sequence space [8]. Numerical studies in regulatory networks, based on simple models of gene regulation [9–12] found that the space of viable networks can be crossed in small steps, and that a wide range of new expression patterns can be generated by small changes to different viable networks. The large number of viable regulatory networks is particularly relevant from

^(a)E-mail: berg@thp.uni-koeln.de

an evolutionary perspective, as neutral evolution gradually explores different viable networks.

These observations call for a statistical approach based on the ensemble of all viable networks, which is the topic of this paper. Statistical mechanics forms a natural framework for exploring the ensemble of viable networks: The set of all possible networks forms a phase space (analogous to, *e.g.*, the set of all possible spin configurations in an Ising model). An indicator function, equal to one for a viable network and equal to zero otherwise, replaces the Boltzmann factor. The corresponding partition function then generates the ensemble of viable networks.

In order to study the ensemble of viable networks on a concrete example, we first develop a model regulatory network. Then the partition function is constructed and solved analytically. An order parameter emerges from this calculation, which quantifies the symmetry of interactions in viable networks. Finally, a number of evolutionary consequences of the properties of the set of viable networks are discussed.

A model regulatory network. – Fundamentally, a network is viable if it produces the required gene products in every phase of the life cycle of an organism, and in every environmental condition the organism is likely to encounter. A model regulatory network has to produce a prescribed set of expression levels on the basis of a model regulatory mechanism. We discuss these components in turn.

Gene expression levels. For every external or internal condition, the organism has to produce a certain set of gene products. For instance, when nutrients are available, specific enzymes have to be produced to digest these nutrients. We model these distinct external or internal states of an organism as a discrete set of numbers $\mu = 1, 2, \dots, P$. For each state μ of the organism, ξ_i^μ denotes the (log-) expression level of gene i . ξ_i^μ positive for high steady-state concentrations of the gene product and negative for low steady-state concentrations. Figure 1 gives a graphical representation.

Regulatory interactions. The task of producing the correct expression levels for every gene and every condition falls to the regulatory network. We restrict the discussion to networks with pairwise regulatory interactions between genes. A regulatory interaction between gene j , coding for a transcription factor, and its target gene i is denoted by J_{ij} . The regulatory interaction J_{ij} is positive when gene j codes for a transcription factor which enhances the expression of gene i , it is negative when j represses the expression of i . The matrix $\{J_{ij}\}$ thus encodes the regulatory network. As only transcription factors affect the expression levels of genes, $J_{ij} = 0$ unless j actually codes for a transcription factor. (Conversely, of course, genes coding for transcription factors can be target genes of other transcription factors.) We thus distinguish between genes coding for *transcription factors*, and genes coding for enzymes, structural components of the cell, motor

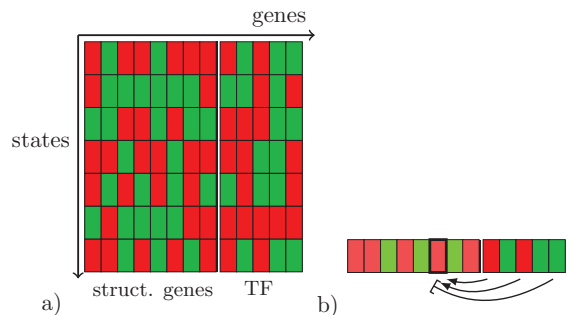


Fig. 1: (Colour on-line) Expression levels and regulatory networks. a) We list genes along the x -axis, and states of the organism along the y -axis. Following established convention in expression analysis, expression levels are colour-coded with high expression levels shown in red (dark), low levels in green (light). b) Regulatory interactions must be compatible with gene expression levels in all states of the organism. The schematic example shows interactions between transcription factors and a single target gene; two enhancing interactions (\rightarrow) with upregulated transcription factors, and a repressive interaction (\rightarrow) with a downregulated transcription factor lead to the activation of the target gene.

proteins, etc. The latter class, consisting of genes which are not involved in cellular information processing, is conventionally referred to as *structural genes*. For convenience, we sort the genes into structural genes and transcription factors; genes are labelled $i = 1, \dots, N$ for structural genes and $i = N + 1, \dots, \beta N$ for transcription factors. Typical values for the parameter β , giving the fraction of transcription factors to structural genes, are about 0.1 and larger. In bacteria, an intriguing interplay of the number of transcription factors and the genome size has been found [13].

Gene regulatory mechanism. In a given state μ , the combination of transcription factors present determines whether a given gene will be transcribed or not. We assume the effects of transcription factors to combine linearly: a gene i is transcribed (positive value of ξ_i^μ) if the sum of present activators of that gene, minus the sum of repressors reaches a threshold. Otherwise gene i is not transcribed (negative value of ξ_i^μ). Combining these two cases gives the condition

$$\xi_i^\mu / \sqrt{N} \sum_j J_{ij} \xi_j^\mu > \kappa, \quad \forall i, \mu. \quad (1)$$

Threshold condition (1) has been used extensively to model neural [14] and gene regulatory networks [10,11, 15–17]. For each state μ , it provides a condition for the self-consistency of an expression patterns —genes with high expression levels are transcribed, genes with low expression levels are not transcribed— but does not describe how the organism proceeds from one state to another [18].

An *indicator function* for viable networks equals one, if condition (1) is fulfilled, and zero otherwise. This indicator function can be written in terms of the Heaviside

step-function $\Theta(x)$ as

$$\begin{aligned} I(\{J_{ij}, \xi_i^\mu\}) &= \prod_{i,\mu} \Theta \left(\frac{1}{\sqrt{N}} \xi_i^\mu \sum_j J_{ij} \xi_j^\mu - \kappa \right) \\ &= \prod_{i,\mu} \int_{\kappa}^{\infty} d\lambda \int \frac{dx}{2\pi} e^{i \sum_{i\mu} \lambda_i^\mu x_i^\mu - \frac{i}{\sqrt{N}} \sum_{ij\mu} x_i^\mu \xi_i^\mu J_{ij} \xi_j^\mu}, \end{aligned} \quad (2)$$

the second step using the integral representation of $\Theta(x)$.

The ensemble of viable networks: partition function. – The two classes of genes introduced above, structural genes and transcription factors, have very different constraints acting on their expression levels. In a given state, structural genes have their expression levels *fixed by external constraints*; e.g. when a specific enzyme is required, the corresponding gene must be expressed. Expression levels of transcription factors, on the other hand, can be *adapted* to take on any value, provided the viability condition (1) is fulfilled. For transcription factors, thus, viable expression levels depend on the regulatory interactions: for example expressing a repressor at a high level has the same effect as downregulating an enhancer.

For a given set of expression levels of structural genes, the ensemble of viable networks is characterized by the microcanonical partition function

$$Z \equiv e^S = \frac{\text{Tr}_{\mathbf{J}, \boldsymbol{\xi}_t} I(\mathbf{J}, \boldsymbol{\xi})}{\text{Tr}_{\mathbf{J}, \boldsymbol{\xi}_t}}. \quad (3)$$

This partition function gives the *fraction of viable networks* in terms of the trace $\text{Tr}_{\mathbf{J}, \boldsymbol{\xi}_t}$ over the phase space (regulatory interactions \mathbf{J} and the expression levels of transcription factors $\boldsymbol{\xi}_t$) and the indicator function (2) of couplings and all expression levels. The corresponding entropy $S = \ln Z$ quantifies how sparse viable networks are in the space of all regulatory interactions and transcription factor expression levels. The average of the entropy over expression levels of structural genes is the central quantity, it is calculated using the replica trick [19]. This so-called *quenched average* describes typical instances from a distribution of expression levels.

For simplicity, we constrain vectors of regulatory interactions \mathbf{J}_i and of expression levels $\boldsymbol{\xi}_i$ to lie on hyperspheres. This defines the trace over phase space (4)

$$\text{Tr} = \prod_i \int d\mu(\mathbf{J}_i) \prod_{i>N} \int d\mu(\boldsymbol{\xi}_i) \quad (4)$$

with $d\mu(\mathbf{J}_i) = \prod_{j>N} dJ_{ij} \delta((\beta-1)N - \sum_{j>N} J_{ij}^2)$ and analogously for the average over transcription factor expression levels; $\langle \langle S \rangle \rangle = \prod_{i \leq N} \int d\mu(\boldsymbol{\xi}_i) \ln Z$.

Analysis of the model: the role of transcription factors. – Transcription factors play a special role; their expression levels provide the regulatory input for *every gene* in the regulatory network. This produces an effective coupling between regulatory interactions of different

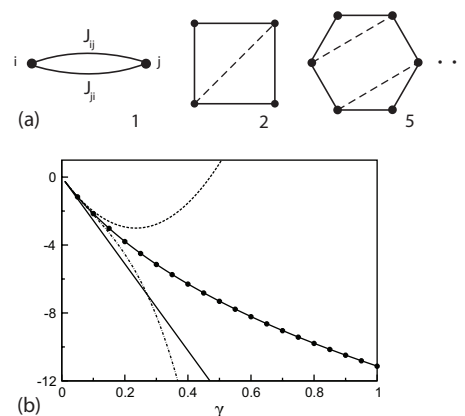


Fig. 2: Averaging over expression levels. a) The diagrams corresponding to the first three terms in (5) are shown along with their combinatorial factors. Nodes represent variables i, j, \dots , solid lines indicate the corresponding matrix entries G_{ij} , dashed lines are contractions $i = j$. b) Plotting the logarithm of (5) against γ shows the contribution of different diagrams. The first diagram gives a linear term (thin solid line), the series up to second and third order are shown by the dashed and dash-dotted curves, respectively. These are valid approximations up to some finite values of γ only. The thick solid line gives the full series to infinite order along with a numerical computation of (5), where G_{ij} was taken a random matrix of size $N = 50$ with i.i.d. normally distributed elements.

genes. One consequence emerges already at the level of the average of (2) over the expression levels ξ . As an illustration, we consider a toy problem, where the average of $\exp\{-i\sqrt{\gamma/(2N)} \sum_{ij} \xi_i G_{ij} \xi_j\}$ is computed over a distribution of independent normally distributed variables ξ_i . G_{ij} is a symmetric matrix with uncorrelated random entries:

$$\begin{aligned} \langle e^{-\frac{i}{2} \sqrt{\gamma/N} \sum_{ij} \xi_i G_{ij} \xi_j} \rangle &= 1 / \sqrt{\det(\mathbf{1} + i\sqrt{\gamma/N} \mathbf{G})} \\ &= \exp \left\{ -\frac{1}{2} \sum_{n=1}^{\infty} \frac{1}{n} \text{Tr}(i\sqrt{\gamma/N} \mathbf{G})^n \right\} \\ &= \exp \left\{ \gamma/4 \sum_i z_i - \gamma^2/4 \sum_i z_i^2 + 5\gamma^3/12 \sum_i z_i^3 + \dots \right\} \\ &= \exp \left\{ \sum_i w(\gamma z_i) \right\} \end{aligned} \quad (5)$$

with shorthand $z_i = 1/N \sum_j G_{ij} G_{ji}$ and Tr denoting the matrix trace. The successive terms in the power series (5) can be represented diagrammatically; fig. 2(a) shows the first three diagrams. Figure 2(b) shows how the different powers in (5) contribute to the average and how for finite values of γ the series has to be taken to infinite order, giving $w(z) = \frac{2z+1-\sqrt{1+4z}}{8z} - \frac{1}{2} \log \left(\frac{1}{2} + \frac{1}{2} \sqrt{1+4z} \right)$. This is in contrast to the standard situation in fully-connected disordered models, where in the thermodynamic limit the series in (5) terminates after the first term.

The approach (5) applied to the full model (2),(3) gives $\sum_{i \leq N, \mu} w(z_i^\mu) + \sum_{i > N, \mu, a} w(z_i^{\mu a})$ with $z_i^\mu = \sum_a (x_i^{\mu a})^2$ and $z_i^{\mu a} = (x_i^{\mu a})^2 + \frac{1}{N} \sum_j (x_i^{\mu a})^2 (J_{ji}^a)^2 + \frac{2}{N} \sum_j x_i^{\mu a} x_j^{\mu a} J_{ij}^a J_{ji}^a$. Neglecting fluctuations of $z_i^{\mu a}$ across genes, the entropy of viable networks $\langle \langle S \rangle \rangle \equiv N^2 s$ can be computed in the thermodynamic limit $N \rightarrow \infty$ by standard methods. s is the intensive entropy; the entropy S scales as N^2 , since there are of order N^2 degrees of freedom. $(\beta(\beta-1)N^2$ regulatory interactions and $PN(\beta-1) \equiv \alpha N^2(\beta-1)$ transcription factor expression levels. We consider the number of patterns to scale linearly with the number of genes, $P \equiv \alpha N$. This is a natural scaling, in the sense that changes of order one of α turn out to lead to changes of order one in the order parameters of the system, see below.)

Within a replica-symmetric ansatz we obtain

$$s = \text{extr}_{X_1, \hat{X}_1, X_2^\pm, \hat{X}_2^\pm, F, h} \left[\frac{1}{2} \alpha (\beta-1) (F - \ln F) + \frac{1}{2} \alpha \hat{X}_2^- X_2^- \right. \\ \left. + \alpha (\beta-1) (\hat{X}_2^+ X_2^+ / 2 + \hat{X}_1 X_1) + \frac{(\beta-1)^2}{4} \ln(1-h^2) \right. \\ \left. + \alpha (\beta-1) \ln \left[\int_\kappa^\infty d\lambda \int \frac{dx}{2\pi} \exp \left\{ w \left(\frac{1}{F^2} [x^2 + X_2^+ \right. \right. \right. \right. \right. \\ \left. \left. \left. - 2ixhX_1] + \frac{X_2^-}{(\beta-1)F} \right) + i(\lambda - \hat{X}_1)x - \frac{1}{2} \hat{X}_2^+ x^2 \right\} \right] \right. \\ \left. + \alpha \int d\hat{y} dy e^{[iy\hat{y} + w(\frac{y}{F})]} \ln \left[H \left(\frac{\kappa}{\sqrt{\hat{X}_2^- + 2iy}} \right) \right] \right] \quad (6)$$

as the effective single-particle contribution to the entropy of viable networks in (3). $H(x)$ denotes the cumulative Gaussian measure $\int_x^\infty \frac{dy}{\sqrt{2\pi}} \exp\{-y^2/2\}$. Typical for mean-field models, eq. (6) contains an extremum over order parameters (akin to the magnetisation emerging as the order parameter in the Curie-Weiss model of the ferromagnet). Most of the order parameters here, $X_1, \hat{X}_1, X_2^\pm, \hat{X}_2^\pm$, describe the statistics of the variable x in the integral representation of the indicator function in (2), another, F , stems from the spherical constraint on expression levels. Solely the order parameter h has an intuitive interpretation in terms of the symmetry of regulatory interactions, which will be discussed in detail below.

Network entropy: encoding an unlimited number of expression patterns. – The entropy of viable networks (6) decreases with increasing number of patterns $P = \alpha N$, see fig. 3. This is to be expected, as each set of expression patterns induces a new set of constraints on the network. However, the entropy remains finite even as the number of patterns becomes large with $\alpha \rightarrow \infty$: there (typically) *always exists a viable network*, so there is *no transition* to a phase where solutions of (1) no longer exist¹. In contrast, phase transitions between

¹The statement holds provided α and β remain finite in the thermodynamic limit.

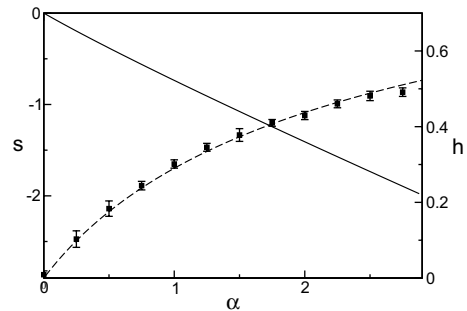


Fig. 3: Entropy and symmetry of viable networks. With increasing number of patterns $P = \alpha N$, the space of viable networks shrinks, and the networks become increasingly symmetric, see text. Here the entropy s per structural gene (solid line) and symmetry parameter h (dashed line) are plotted against α for $\beta = 2, \kappa = 0$. The \blacksquare symbols stem from numerical simulations with $N = 80$, $\kappa = 0$. The \blacksquare symbols stem from numerical simulations with $N = 80$, averaged over 20 realizations of the quenched disorder (mean and standard error).

learnable and unlearnable regimes are well known in neural networks and many combinatorial problems [14,20]. At these phase transitions the solution space of a system shrinks to zero, and the entropy diverges.

At first sight, the ability of regulatory networks to store a virtually unlimited number of expression patterns of structural genes is striking. This ability stems directly from the freedom to choose expression levels of transcription factors: transcription factor expression levels adapt in such a way that regulatory interactions compatible with expression levels of all genes can be found. Each set of expression levels of structural genes (resulting in new constraints to the viability condition (1)) also brings a new set of adaptive variables (the expression levels of transcription factors). The balance between these two contributions results in a finite entropy of viable networks.

One consequence of the large set of solutions to the viability condition (1) concerns the evolution of regulatory interactions under constraints on expression levels of structural genes. At first sight, one would expect such constraints to lead to the evolutionary conservation of regulatory interactions, in analogy with the evolution of biological sequences. However, our model shows why regulatory interactions alone show little such conservation: regulatory interactions specify expression levels of structural genes only when taken together with the expression levels of transcription factors. Both sets of degrees of freedom *co-evolve* and, considered separately, show a fast decay of their correlation (see below). To detect conservation of regulatory systems, one would thus need to consider both interactions and expression levels.

Symmetry of regulatory networks. – The order parameter $h = \frac{1}{(\beta-1)^2 N^2} \sum_{i,j} J_{ij} J_{ji}$ is the *symmetry parameter* of the resulting regulatory network. A positive value of h indicates that if gene i regulates gene j , and also j regulates i , the signs of these interactions are correlated,

with like signs occurring more frequently than opposite signs. The origin of this symmetry lies in condition (1) for a viable network, where a positive value of $\xi_i^\mu \xi_j^\mu$ for some i, j, μ gives rise to positive values for both J_{ij} and J_{ji} , and analogously for negative values. Thus, the symmetry parameter h increases with the number of expression patterns; fig. 3 shows the analytical outcome for h along with the result of numerical simulations.

This statistical bias towards symmetric interactions is compatible with empirical data on regulatory networks. A literature search for well-documented cases of mutually interacting genes with known interaction sign² finds 9 cases of mutually interacting gene pairs with like interaction sign [21] compared to only 3 cases with different sign [22]. A nontrivial statistics of reciprocal interactions has also been found in neural and metabolic networks [23], where, however, the signs of the interactions are generally unknown.

It may be that these results stem from very specific mechanisms, and are realized only in a few instances of regulatory networks. On the other hand, they may also derive from a generic mechanism, whereby any stability constraint of a form similar to (1) induces a symmetry in the regulatory network. Future genomewide data of high quality will show if the reported bias is indeed generic.

The numerical simulations in fig. 3 are based on Monte Carlo dynamics of the regulatory interactions J_{ij} and the expression levels of transcription factors $\xi_{i>N}^\mu$. As energy function we use the sum of squared deviations from the viability condition (1)

$$\mathcal{H}(\{J_{ij}, \xi_i^\mu\}) = \sum_{i,\mu} \Xi \left(\frac{1}{\sqrt{N}} \xi_i^\mu \sum_j J_{ij} \xi_j^\mu - \kappa \right), \quad (7)$$

with $\Xi(x)$ equal to x^2 for $x < 0$, and zero otherwise. The minima $\mathcal{H} = 0$ satisfy the viability condition (1) by construction [14]. Spherical constraints on regulatory interactions and expression levels are implemented using Lagrange multipliers. Simulated annealing was used to find viable networks: random changes in the regulatory interactions and expression levels of transcription factors were drawn from a univariate Gaussian distribution and were accepted or rejected according to the Metropolis rule at an inverse temperature which was linearly increased from zero to 7 over 1000 Monte Carlo sweeps, causing gradual convergence to the ensemble of viable networks.

Adaptation to evolving constraints. – Over long evolutionary time scales, the required expression levels of structural genes can change. In the case of enzymes, for instance, changing nutrient availability or changing metabolic rates alter the required expression levels. Such changes of the expression levels of structural genes induce

²We exclude data from high-throughput experiments, such as ChIP-on-chip assays, which are noisy and give only the presence/absence of an interaction.

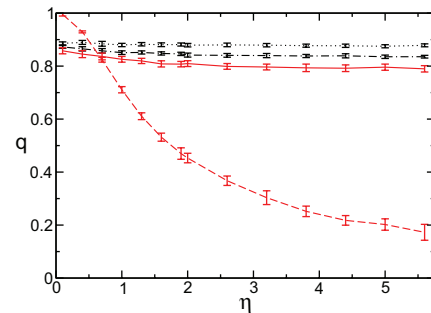


Fig. 4: (Colour on-line) Response to changing expression levels. The overlaps of perturbed and unperturbed systems (see text) are plotted against the perturbation strength η : structural genes expression level overlaps (red dashed line) tend to zero with increasing η by construction, whereas transcription factor expression level overlaps (red solid line) quickly reach a plateau. The same holds for regulatory interactions to transcription factors (black dotted line), and interactions to structural genes (black dash-dotted line). The plateau value decreases with the fraction $1 - \beta$ of transcription factors in the genome. The data stem from Monte Carlo simulations with $N = 40$, $\alpha = 1$, $\beta = 2$, and $\kappa = 0$, averaged over 20 samples.

adaptive changes both of the regulatory network, and of the expression levels of transcription factors. To investigate the adaptation to changing expression levels of structural genes, we systematically perturb the expression levels of structural genes of a viable network, rendering it, in general, at first unviable. (Expression levels $\xi_{i \leq N}^\mu$ are perturbed by adding i.i.d. Gaussian random variables with mean zero and standard deviation η and normalizing their variances to one again.) Subsequently, regulatory interactions and transcription factor expression levels are adapted until the viability condition (1) is satisfied again. The adaptation was again performed by simulated annealing under the energy function (7). The overlap $q_\xi^< = \frac{1}{NP} \sum_{i \leq N, \mu} \xi_i^\mu \xi_i^{\prime \mu}$ of structural gene expression levels of the unperturbed (unprimed) and the perturbed (primed) system quantifies the strength of the perturbation, the analogously defined $q_\xi^>$, $q_J^<$, $q_J^>$ quantify the response of the system to this perturbation. Figure 4 shows the overlaps as a function of perturbation strength. One finds that already small perturbations with $q_\xi^< \approx 1$ result in a drop of the overlaps to a plateau value. Larger perturbations, and even the limit $q_\xi^< \rightarrow 0$ induce only a slow decay of $q_\xi^>$, $q_J^<$, $q_J^>$ from their plateau values. Accordingly, close to any viable network for one set of expression levels of structural genes, there exists a viable network for any other, even unrelated set of expression levels. This effect allows fast adaptation to changes in the required expression levels.

Another consequence of the observed drop of the transcription factor expression level overlap to a plateau is that expression levels of transcription factor change *more* than those of structural genes for *small perturbations*. For *large perturbations*, the expression levels of transcription factors change *less* than those of structural genes. This

effect may explain an apparent contradiction in the cross-species comparison of experimentally measured expression levels across different *Drosophila* species and across different primates. The *Drosophila* species not only have a larger sequence divergence between them (compared to primates), but also live in more diverse habitats. As a result, the genes of the *Drosophila* species also show, on average, larger expression level changes between them than the primates do [24,25]. A comparison of humans with other primates shows large changes of transcription factor expression levels [24] compared to structural genes, different *Drosophila* species show only small changes of transcription factor expression levels compared to structural genes [25].

In summary, we have investigated the degeneracy of regulatory networks within a simple model of genetic regulation. We have found that typical instances of regulatory networks show a symmetry of reciprocal interactions. Future data on regulatory networks will enable a detailed comparison between the statistics of empirical networks and the statistical ensemble of viable networks defined by constraints such as the viability condition (1). This comparison could include, *e.g.*, the connectivity distribution in models with a finite connectivity, or the statistics of loop lengths. Such a comparison would require high-quality data beyond the current high-throughput experiments such as ChIP-on-chip, which are plagued by signal-to-noise problems.

Given the simplicity of the model, one may ask how robust the present results are with respect to changes of the model. In general any constraint, of the form (1) or of a different form, will induce a networks statistics with a *characteristic bias* compared to random networks. The characteristic bias induced by the particular constraint (1) is the symmetry bias of reciprocal interactions. Here, a statistical bias in empirical data on both regulatory networks *and* expression patterns may help us guess the form of the viability condition. The connection between transcription factor expression levels acting as degrees of freedom, and the large space of viable networks may well persist also in more complex models, as each new pattern introduces new degrees of freedom. However, the geometry of the solution space may change. In particular, models taking into account physical interactions between transcription factors to implement logical functions [16] lead to p-spin interactions $J_{ijk\dots}$ and may result in a disconnected solution space and combinatorial complexity.

Many thanks to A. ALTLAND, M. COSENTINO LAGOMARSINO, and M. ŁUKSZA for discussions. Funding from the DFG is acknowledged under grants BE 2478/2-1 and SFB 680.

REFERENCES

- [1] TSONG A., MILLER M., RAISNER R. and JOHNSON A., *Cell*, **115** (2003) 389.
- [2] SCHRÖDER R., *Nature*, **422** (2003) 621.
- [3] TANAY A., REGEV A. and SHAMIR R., *Proc. Natl. Acad. Sci. U.S.A.*, **102** (2005) 7203.
- [4] MADAN BABU M., TEICHMANN S. and ARAVIND L., *J. Mol. Biol.*, **358** (2006) 614.
- [5] TAUTZ D., *Curr. Opin. Genet. Dev.*, **10** (2000) 575.
- [6] MUSTONEN V. and LÄSSIG M., *Proc. Natl. Acad. Sci. U.S.A.*, **102** (2005) 15936.
- [7] CHOUARD T., *Nature*, **456** (2008) 300.
- [8] SCHUSTER P., FONTANA W., STADLER P. F. and HOFACKER I. L., *Proc. Biol. Sci.*, **255** (1994) 279.
- [9] BORNHOLDT S. and SNEPPEN K., *Phys. Rev. Lett.*, **81** (1998) 236.
- [10] CILIBERTI S., MARTIN O. and WAGNER A., *PLoS Comput. Biol.*, **3** (2007) e15.
- [11] CILIBERTI S., MARTIN O. and WAGNER A., *Proc. Natl. Acad. Sci. U.S.A.*, **104** (2007) 13591.
- [12] LAU K.-Y., GANGULI S. and TANG C., *Phys. Rev. E*, **75** (2007) 051907.
- [13] MOLINA N. and VAN NIMWEGEN E., *Trends Genet.*, **25** (2009) 243.
- [14] HERTZ J., KROGH A. and PALMER R., *Introduction to the Theory of Neural Computation* (AddisonWesley, Reading Mass.) 1991.
- [15] WAHDE M. and HERTZ J., *Biosystems*, **55** (2000) 126.
- [16] BUCHLER N., GERLAND U. and HWA T., *Proc. Natl. Acad. Sci. U.S.A.*, **100** (2003) 5136.
- [17] BOLDHAUS G. and KLEMM K., <http://arxiv.org/abs/0904.4843> (2009).
- [18] NOVAK B., TYSON J., GYORFFY B. and CSIKASZ-NAGY A., *Nat. Cell Biol.*, **9** (2007) 724.
- [19] MÉZARD M., PARISI G. and VIRASORO M., *Spin Glass Theory and Beyond* (World Scientific, Singapore) 1987.
- [20] MÉZARD M. and MONTANARI A., *Information Physics and Computation* (Oxford University Press, Oxford) 2008.
- [21] GOUDREAU G. *et al.*, *Proc. Natl. Acad. Sci. U.S.A.*, **99** (2002) 8719; MIKELADZE-DVALI T. *et al.*, *Cell*, **122** (2005) 775; OLGUIN H. C., YANG Z., TAPSCOTT S. J. and OLWIN B. B., *J. Cell Biol.*, **177** (2007) 769; CHEW J.-L. *et al.*, *Mol. Cell Biol.*, **25** (2005) 6031; FELDSER D. *et al.*, *Cancer Res.*, **59** (1999) 3915; LEMEILLE S., GEISELMANN J. and LATIFI A., *BMC Microbiol.*, **5** (2005) 18; TO A. *et al.*, *Plant Cell*, **18** (2006) 1642; PANCHANATHAN R., XIN H. and CHOUBEY D., *J. Immunol.*, **180** (2008) 5927; CANAPLE L. *et al.*, *Mol. Endocrinol.*, **20** (2006) 1715.
- [22] ALABADI D. *et al.*, *Science*, **293** (2001) 880; SCHMIDT J. E., VON DASSOW G. and KIMELMAN D., *Development*, **122** (1996) 1711; BORRELLI S. *et al.*, *BMC Mol. Biol.*, **8** (2007) 85.
- [23] GARLASCHELLI D. and LOFFREDO M. I., *Phys. Rev. Lett.*, **93** (2004) 268701.
- [24] GILAD Y., OSHLACK A., SMYTH G. K., SPEED T. P. and WHITE K. P., *Nature*, **440** (2006) 242.
- [25] RIFKIN S. A., KIM J. and WHITE K. P., *Nat. Genet.*, **33** (2003) 138.