FEBS Letters 498 (2001) 140–144 FEBS 24903

Minireview

Transcriptomes, transcription activators and microarrays

Frédéric Devaux[1], Philippe Marc[1], Claude Jacq*

Laboratoire de genetique moleculaire, Ecole Normale Superieure, 46 rue d'Ulm, 75005 Paris, France

Received 10 April 2001; accepted 27 April 2001

First published online 16 May 2001

Edited by Gunnar von Heijne

Abstract Gene-specific transcription activators are among the main factors which specifically shape the transcriptome profiles. It is tempting to take advantage of their properties to decipher the genome expression circuitry. The advent of microarray technology has offered fantastic opportunities to quickly analyze the expression profiles dictated by specific transcription factors. This review will first focus on the strategies which have been devised to control the activity of transcription factors and in the second part on the microarray experiments which addressed the role of these transcription factors in the genome-wide expression profile. This last part will mainly consider the case of the yeast *Saccharomyces cerevisiae* genome. All the collected data are available through the on-line database yTAFNET (http://transcriptome.ens.fr/ytafnet/). yTAFNET is designed to help the characterization of connections between the different yeast regulatory networks.

Key words: yTAFNET; Transcription factor; Regulation network; Database; DNA microarray

1. Introduction

Since future progress in our understanding of gene expression processes will undoubtedly be derived from genetic and biochemical approaches, it seems likely that genome-wide expression analyses will provide the most significant new dimension to the field in the near future. One of the most fundamental questions is to understand how transcriptional activators and the many components of the transcription initiation machinery collaborate to regulate gene expression. The wealth of information coming from genome sequencing projects coupled with the powerful DNA microarray technology could be used to devise new rationale approaches. Two quantitatively different levels of regulation can be distinguished when one considers either the modifications of the chromatin structure or the involvement of gene-specific transcription factors. Several genome-wide studies have already addressed the role of chromatin modifying enzymes and of various components of the initiation apparatus (Fig. 1). Among at least half a dozen chromatin modifying enzymes, the role of the histone deacetylase (HDAC) Rpd3 has been examined by treating wild type yeast cells with the trichostatin A, an HDAC inhibitor, or by analyzing the transcription profile of yeast deletion strains [1]. It was thus found that the yeast HDACs, Rpd3, Sir2 and Hda1, play distinct roles in regulating genes involved in cell cycle progression, amino acid biosynthesis and carbohydrate transport and utilization. It was also shown that the *Saccharomyces cerevisiae* Swi–Snf complex, involved in nucleosome remodelling, exerts its effect at the level of individual genes rather than over large chromosomal domains [2,3]. Similarly, Chd1 has been shown to exert both positive and negative effects on transcription of 2–4% of all genes [4]. The global effects of depleting nucleosomal histones and silencing factors (Sir2, Sir3, Sir4 and Rap1) were also investigated and it was found that sub-telomeric genes were derepressed over regions extending 20 kb from the telomeres indicating that histones make Sir-dependent contributions to telomeric silencing, and that the role of histones located elsewhere in chromosomes is gene-specific rather than generally repressive [5]. The gene-specific involvement was also demonstrated by microarray analyses in the cases of different elements of the transcriptional machinery like the yeast mediator [6,7] or different TFIID or SAGA complexes [8,9]. However, most of these microarray experiments involving yeast strains with mutations in specific transcriptional proteins do not allow the determination of whether the observed positive or negative effects are due to the direct action of the transcriptional regulatory protein at the affected promoter. Thus, relative promoter association studies using chromatin immunoprecipitation could demonstrate that the NC2 heterodimer directly affects 17% of *S. cerevisiae* genes in a pattern that resembles the response to general environmental stress [10].

More generally these experiments reveal an unanticipated level of regulation which is superimposed on that due to gene-specific transcription factors. The connections between these two classes of regulations are a fundamental question in the field of eukaryotic gene expression. This review will be focussed on the genome-wide properties of the gene-specific transcriptional factors. In the simple eukaryote *S. cerevisiae* roughly 4% of the genes encode proteins which are likely to be specific transcriptional factors because they have a typical DNA-binding domain. About 30% of them have been demonstrated to act as transcription activators and in most of the cases very little information is available concerning the complete set of target genes under the control of a given transcriptional factor. A more complete knowledge of the different regulation networks is however a prerequisite if one wishes to describe and control the expression of the genome. We will review the genome-wide analyses which have addressed

*Corresponding author. Fax: (33)-1-44 32 37 30.
E-mail: jacq@biologie.ens.fr

[1] Both authors contributed equally to this work.

PII: S0014-5793(01)02478-4

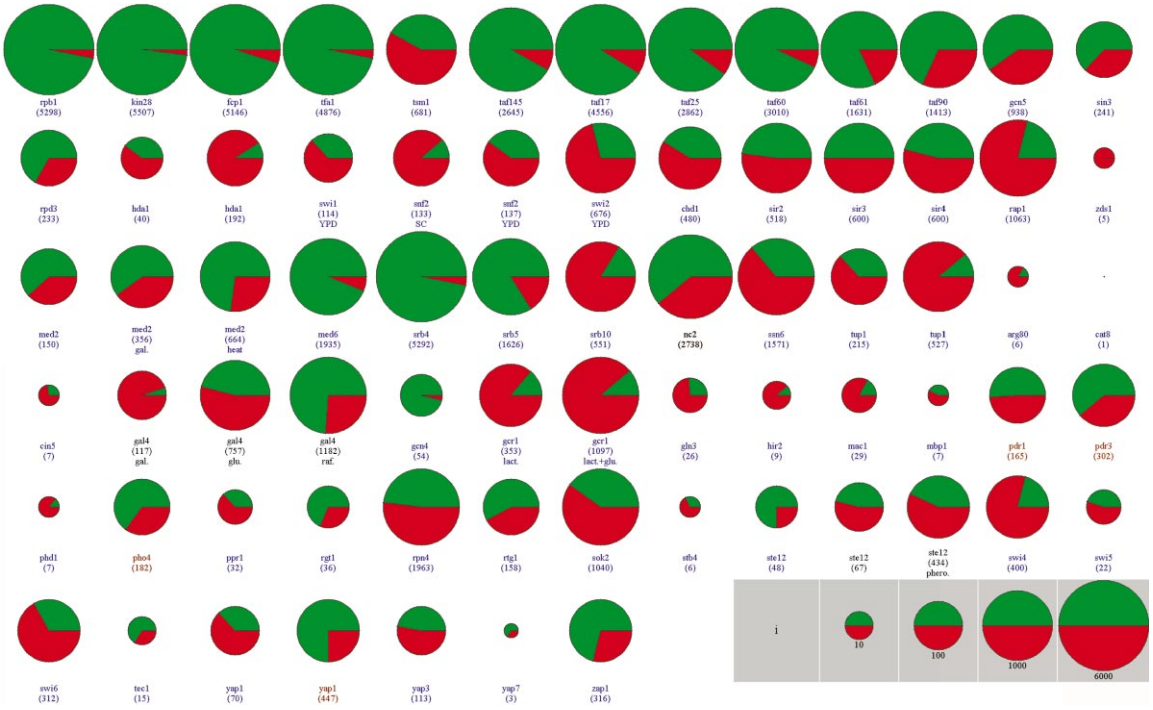

Fig. 1. How many genes are regulated by a yeast transcription factor? Seventy two published experiments (complete list available at http://transcriptome.ens.fr/ytafnet/) have been analyzed to extract the number of up- or down-regulated genes corresponding to different states of a transcription factor. Only genes which were two-fold induced or repressed are considered. Each pie represents the proportion of genes that are up-regulated (in red) or down-regulated (in green) in the mutant compared to the wild type condition. The color of transcription factor names symbolizes the type of experiment: blue = deletion, black = chromatin immunoprecipitation, brown = gain of function or overexpression. The size of the pie is dependent on the total number of genes whose expression changed in the experiment (indicated in brackets). The logarithmic scale of pie size is represented at the bottom of the figure.

the regulatory properties of gene-specific transcription factors after a brief review of the methods which have been used to control the activity of these factors.

2. Control of the transcription factor activity

Addressing the specific role of a transcription activator is always a difficult question since it is known that transcription factors are tightly regulated and act through several types of interactions which, in a combinatorial association, determine the specificity of the activation process. The rather naive approach in which a disrupted strain devoid of a transcription factor is compared to the wild type strain in standard growth conditions may generally not be very informative because of compensatory effects [11]. However, microarray analysis of gene expression in developing retinal tissue from Crx(+/+) and Crx(−/−) mice demonstrated that cDNA microarrays could be used in these conditions to define the transcriptional networks controlled by transcription factors in vertebrate tissue in vivo [12]. Several experimental approaches have been devised to reveal the set of target genes which are under the control of a specific transcription activator. The less artificial endeavours to reveal the properties of gene-specific transcription activators are to systematically study the transcriptome variations within a natural process like the human or yeast cell cycle [13–15] or the yeast sporulation [16]. Cluster analyses and characterization of common elements in the promoter of the co-regulated genes constituted good evidence in favor of the role of specific transcription activators like SBF, MBF, Mcm1, SFF, Ace2, and Swi5 [15]. The direct interaction of some of these factors with their cognate promoter could be further documented by immunoprecipitation experiments [17], such an information being essential to a correct analysis of systematic target gene analysis [18]. Another general approach has been used in the case of the yeast transcription factor Zap1 which senses cellular zinc status [19]. It was reasoned that Zap1-regulated genes would be expressed at higher levels in zinc-limited wild type cells than in zinc-replete cells. Moreover, these genes would be expressed at higher levels in zinc-limited wild type cells vs. zinc-limited *zap1* mutants. Forty six target genes of Zap1 could thus be identified. This was an efficient way to distinguish direct and indirect target genes of a transcription factor, a problem which is not properly addressed when overexpression of the transcription factor, in a wild type or gain of function form, is compared to its total absence in a disruptant strain. Direct overexpression of a transcription factor compared to its total absence was used in the early microarray experiments [20,21] but, more recently, time-course expression of the activated transcription factor associated with a kinetic analysis of the evolution of the transcriptome was designed to distinguish the direct target genes, early activated, from the indirect effects which lead to late activations or repressions. The different time-dependent expression systems relied on conditional promoter control [22] like in the cases of the tumor suppressors WT1 [23] or p53 [24]. In the case of the yeast transcription factor Pdr1, this kinetic approach revealed that all the early induced target genes contain at least one pleiotropic drug responsive element in their promoter in agreement with a direct regulatory process [25]. More generally, the activity of the transcription fac-

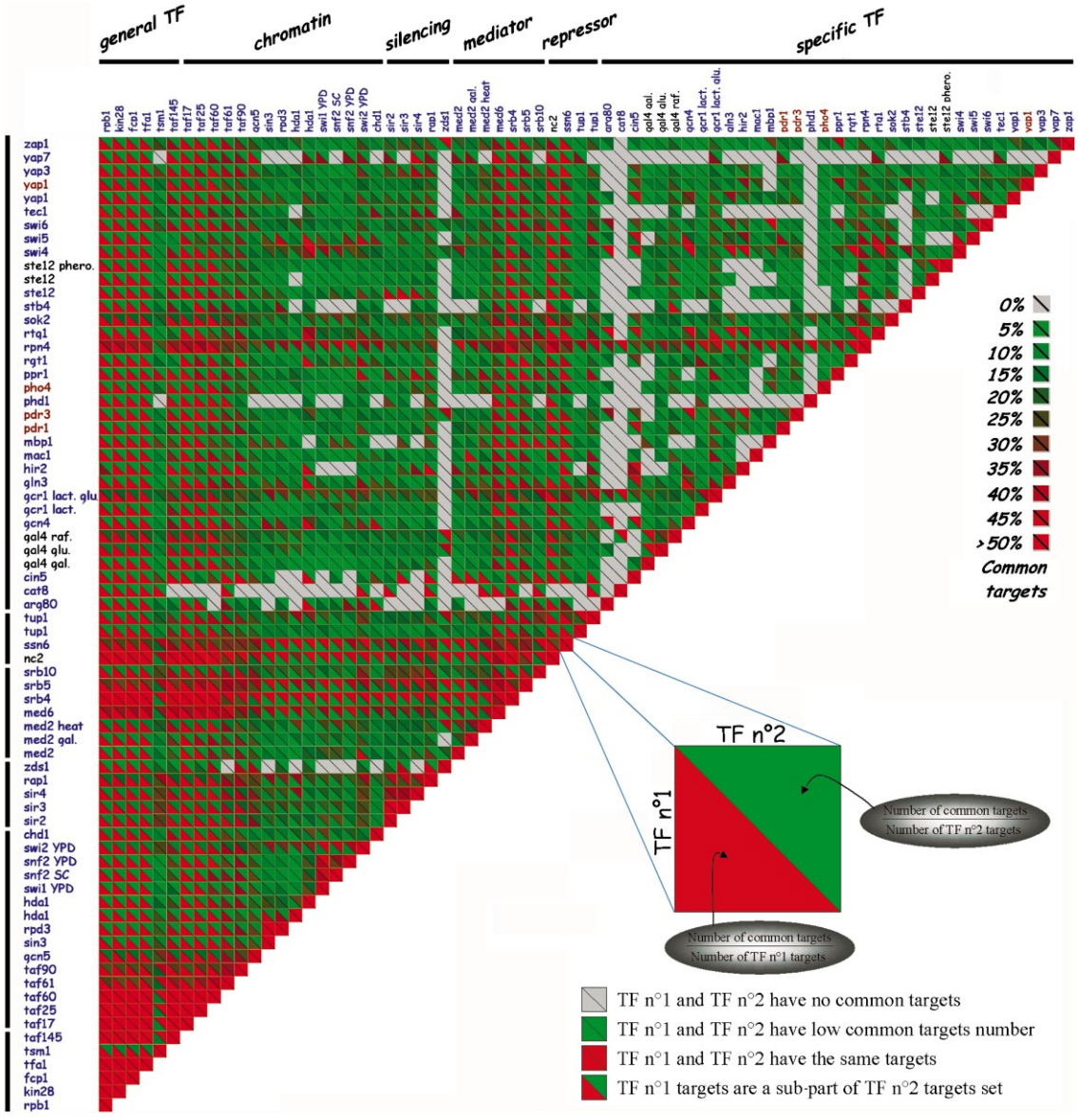

Fig. 2. Connectivity matrix of yeast regulatory networks. Each line or column represents a published microarray experiment in which the target genes of a transcription factor (TF) were identified. Only genes which were two-fold induced or repressed are considered. The intersection of two experiments is figured by a square divided in two parts: the upper and lower parts of the square represent, respectively, the percentage of the transcriptome associated to TF no. 2 that is common with TF no. 1 and vice versa (see model square at the bottom right). The color of each half square region is related to the value of this percentage. The scale of colors is indicated at the right of the figure. For example, a square which is half green and half red means that the transcriptome associated to the TF on the red side is a small part of the transcriptome associated to the TF on the green side. The more both half squares are red, the more the two corresponding TFs share perfectly overlapping transcriptomes. The colors for the experiment names are identical to Fig. 1.

tor rather than its expression can be controlled. Especially interesting is the possibility to control the regulation of transcription factors through expression of chimeric forms [26,27]. The hormone-binding domain of steroid receptors can be used to master the function of heterologous transcription factors by hormonal control in *cis* [28]. One of the most interesting applications of this strategy was followed to identify the MYC target genes [29]. A MYC–ER fusion protein was induced with the estrogen analog 4-hydroxy-tamoxifen and 27 genes were consistently induced and nine genes were repressed. More generally, systematic domain swap experiments have revealed that activation domains are functionally autonomous units able to stimulate RNA polymerase II activity when fused to a heterologous DNA-binding domain [30]. Genome-wide expression analyses have recently shown that the DNA-binding domain of a transcription factor like Pdr1 can specifically recognize all the cognate promoters and activate transcription, provided that it is linked to a constitutive activation domain [25]. More sophisticated approaches aimed at constructing artificial transcription factors [31] by designing artificial DNA-binding domains [32], artificial activating regions [33] or both [34]. In the design of this last synthetic activator, the protein DNA-binding module was replaced with a hairpin-polyamide composed of *N*-methylpyrrole and *N*-methylimidazole amino acids that bind in the minor groove of DNA with a great specificity, a dimerization element that is known to form a coiled-coil, residues 251–281 of the yeast protein Gcn4 and a synthetic activation domain composed

of a 20 amino acid amphipathic peptide [34,35]. It will be interesting to analyze the genome-wide regulatory of such artificial transcription factors which certainly represent a great promise to control the genome properties.

3. Microarrays and transcription activators

The yeast *S. cerevisiae* certainly offers today the widest collection of genome-wide expression analyses. As of April 2001, 50 independent studies encompassing a large spectra of experimental conditions (more than 1000) have been published. Several databases presenting systematic management and analysis of yeast expression data were proposed [36]. Recently, a database called yMGV was constructed to allow a direct visualization and data mining of the microarray data [37]. Among these 50 studies, nine specifically address the role of gene-specific transcription factors. The strategies used to study these different genetic networks have been presented above and they take advantage of the many mutants which are available in yeast. Since all the corresponding results could not be presented in this review, a special database was made available at the following web address: http://transcriptome.ens.fr/ytafnet/. The yeast transcription activator factor network (yTAFNET) is an on-line database providing a synthetic view of the transcription expression profiles resulting from the published genome-wide yeast studies directly related to the role of gene-specific transcription factors. yTAFNET supports queries across the data from all the updated experiments. It allows the user to ask several questions relevant to the role of gene-specific transcription factors in the genetic regulatory circuitry of the genome. Thus it is possible to: (i) establish the list of target genes activated or repressed by a given transcription factor and corresponding to a chosen induction/repression level (1.5, 2, 3), (ii) compare different expression profiles, and (iii) determine the overlap between different transcription regulatory networks. A schematic analysis of these overlaps is presented in Fig. 2. For each overlap and for a two-fold induction or repression level the intensity of the color reflects the percentage of target genes which are shared between the two considered regulatory networks. More precise information containing the exact number of genes involved and the name of these genes can be obtained from the website database. It should be emphasized that assembling a single database from microarray data of different origins, generally not standardized and of irregular quality, could lead to misinterpretations if the user does not go back to the original publication. Nevertheless it can be useful to have a synthetic view of these scattered data. One important aspect of such an analysis is to directly figure the connections between the two main levels of regulation presented in this review: the regulatory networks controlled by gene-specific factors and the general control of the chromatin modifying enzymes or various components of the initiation apparatus.

4. Perspective

It has been proposed [38] that gene regulatory circuits with virtually any desired properties can be constructed from networks of simple regulatory elements. DNA chip technologies associated with the specific properties of transcriptional activators offer extremely powerful approaches to characterize these regulatory elements and decipher the interactions between the regulatory networks which compose the different facets of genome expression. This approach, however, should be able to discriminate between direct and indirect effects of the activation of a gene-specific transcription factor by designing, for instance, time-course experiments. Moreover, direct assessment of relative promoter association in vivo using chromatin immunoprecipitation analysis will be a necessary complement of the expression profiling data. Quite interestingly, these genome-wide approaches can be extended to completely unknown new putative transcription factors which have a typical DNA-binding domain [25] and which have been discovered in the frame of the genome sequence studies. These studies should eventually lead to the complete graph of regulatory networks associated with model genomes and to the description of the functional modules which most probably compose the living cells [39].

References

[1] Bernstein, B.E., Tong, J.K. and Schreiber, S.L. (2000) Proc. Natl. Acad. Sci. USA 97, 13708–13713.
[2] Sudarsanam, P. and Winston, F. (2000) Trends Genet. 16, 345–351.
[3] Sudarsanam, P., Iyer, V.R., Brown, P.O. and Winston, F. (2000) Proc. Natl. Acad. Sci. USA 97, 3364–3369.
[4] Tran, H.G., Steger, D.J., Iyer, V.R. and Johnson, A.D. (2000) EMBO J. 19, 2323–2331.
[5] Wyrick, J.J., Holstege, F.C., Jennings, E.G., Causton, H.C., Shore, D., Grunstein, M., Lander, E.S. and Young, R.A. (1999) Nature 402, 418–421.
[6] Myers, L.C., Gustafsson, C.M., Hayashibara, K.C., Brown, P.O. and Kornberg, R.D. (1999) Proc. Natl. Acad. Sci. USA 96, 67–72.
[7] Myers, L.C. and Kornberg, R.D. (2000) Annu. Rev. Biochem. 69, 729–749.
[8] Holstege, F.C. et al. (1998) Cell 95, 717–728.
[9] Lee, T.I. et al. (2000) Nature 405, 701–704.
[10] Geisberg, J.V., Holstege, F.C., Young, R.A. and Struhl, K. (2001) Mol. Cell Biol. 21, 2736–2742.
[11] Hughes, T.R. et al. (2000) Cell 102, 109–126.
[12] Livesey, F.J., Furukawa, T., Steffen, M.A., Church, G.M. and Cepko, C.L. (2000) Curr. Biol. 10, 301–310.
[13] Cho, R.J. et al. (2001) Nat. Genet. 27, 48–54.
[14] Cho, R.J. et al. (1998) Mol. Cell 2, 65–73.
[15] Spellman, P.T. et al. (1998) Mol. Biol. Cell 9, 3273–3297.
[16] Chu, S., DeRisi, J., Eisen, M., Mulholland, J., Botstein, D., Brown, P.O. and Herskowitz, I. (1998) Science 282, 699–705.
[17] Iyer, V.R., Horak, C.E., Scafe, C.S., Botstein, D., Snyder, M. and Brown, P.O. (2001) Nature 409, 533–538.
[18] Ren, B. et al. (2000) Science 290, 2306–2309.
[19] Lyons, T.J., Gasch, A.P., Gaither, L.A., Botstein, D., Brown, P.O. and Eide, D.J. (2000) Proc. Natl. Acad. Sci. USA 97, 7957–7962.
[20] DeRisi, J.L., Iyer, V.R. and Brown, P.O. (1997) Science 278, 680–686.
[21] DeRisi, J., van den Hazel, B., Marc, P., Balzi, E., Brown, P., Jacq, C. and Goffeau, A. (2000) FEBS Lett. 470, 156–160.
[22] Vallier, L. et al. (2001) Proc. Natl. Acad. Sci. USA 98, 2467–2472.
[23] Lee, S.B. et al. (1999) Cell 98, 663–673.
[24] Zhao, R. et al. (2000) Genes Dev. 14, 981–993.
[25] Devaux, F., Marc, P., Bouchoux, C., Delaveau, T., Hikkel, I., Potier, M.C. and Jacq, C. (2001), EMBO Reports, in press.
[26] Picard, D. (1994) Curr. Opin. Biotechnol. 5, 511–515.
[27] Natesan, S., Molinari, E., Rivera, V.M., Rickles, R.J. and Gilman, M. (1999) Proc. Natl. Acad. Sci. USA 96, 13898–13903.
[28] Picard, D. (2000) Methods Enzymol. 327, 385–401.
[29] Coller, H.A., Grandori, C., Tamayo, P., Colbert, T., Lander, E.S., Eisenman, R.N. and Golub, T.R. (2000) Proc. Natl. Acad. Sci. USA 97, 3260–3265.
[30] Gaudreau, L., Keaveney, M., Nevado, J., Zaman, Z., Bryant,

G.O., Struhl, K. and Ptashne, M. (1999) Proc. Natl. Acad. Sci. USA 96, 2668–2673.
[31] Nevado, J., Gaudreau, L., Adam, M. and Ptashne, M. (1999) Proc. Natl. Acad. Sci. USA 96, 2674–2677.
[32] Kamiuchi, T., Abe, E., Imanishi, M., Kaji, T., Nagaoka, M. and Sugiura, Y. (1998) Biochemistry 37, 13827–13834.
[33] Lu, X., Ansari, A.Z. and Ptashne, M. (2000) Proc. Natl. Acad. Sci. USA 97, 1988–1992.
[34] Mapp, A.K., Ansari, A.Z., Ptashne, M. and Dervan, P.B. (2000) Proc. Natl. Acad. Sci. USA 97, 3930–3935.
[35] Giniger, E. and Ptashne, M. (1987) Nature 330, 670–672.
[36] Aach, J., Rindone, W. and Church, G.M. (2000) Genome Res. 10, 431–445.
[37] Marc, P., Devaux, F. and Jacq, C. (2001), Nucleic Acids Res., in press.
[38] Monod, J. and Jacob, F. (1961) Cold Spring Harbor Symp. Quant. Biol. 26, 389–401.
[39] Hartwell, L.H., Hopfield, J.J., Leibler, S. and Murray, A.W. (1999) Nature 402, C47–C52.