Update TRENDS in Biotechnology Vol.22 No.6 June 2004

Research Focus

The regulatory software of cellular metabolism

Daniel Segrè

Department of Genetics, Harvard Medical School, 77 Louis Pasteur Avenue, Boston, MA 02115, USA

Understanding the regulation of metabolic pathways in the cell is like unraveling the 'software' that is running on the 'hardware' of the metabolic network. Transcriptional regulation of enzymes is an important component of this software. A recent systematic analysis of metabolic gene-expression data in *Saccharomyces cerevisiae* reveals a complex modular organization of co-expressed genes, which could increase our ability to understand and engineer cellular metabolic functions.

The wiring of a metabolic network is represented by a static set of chemical reactions, most of which are well-characterized and could be seen as the 'hardware' component of metabolism [1,2]. Much less is understood about its 'software' component – the complex set of regulatory strategies used by the cell to perform a variety of tasks in different environmental conditions [3,4]. Mastering the language of this software would not only constitute a major intellectual achievement, but could also lead to many practical applications, such as making the application of costly metabolic processes more efficient. Although the transcriptional and non-transcriptional regulation of metabolic flow in single pathways has been studied extensively [5], the discovery of recurring regulatory patterns and modular organization in the transcriptional regulatory program of metabolism remains an ongoing challenge in functional genomics [6]. The availability of pathway databases and large expression datasets offers new approaches for reaching this goal, as explored by Ihmels *et al.* for the metabolic network of *Saccharromyces cerevisiae* [7]. Their systematic approach could help bridge the gap between top-down data analysis and bottom-up model building.

Towards a 'modulocentric' approach

Ihmels *et al.* began their study by asking whether all genes within a given pathway (defined using the Kyoto Encyclopedia of Genes and Genomes database [2]) are, as might be expected, co-regulated. It appears that, for most pathways, only a subset of genes displays significant co-expression (e.g. 24 out of 46 genes in glycolysis). The Signature Algorithm [8] (Figure 1) is used to identify systematically any classes of genes that are associated with a given metabolic pathway and whose expression is correlated under a certain set of conditions. These sets of genes, or transcriptional modules, consist mainly of linear arrangements of enzymes in the metabolic network. These modules could also include genes from feeder pathways, transporters and transcription factors. An independent forthcoming study (P. Kharchenko *et al.*, unpublished) recently found even smaller fractions of co-expressed genes within classically defined pathways. Importantly, genes from the same pathway that are not co-regulated transcriptionally could instead be constitutively transcribed [3] and under non-transcriptional control [4].

The large-scale analysis of enzyme co-regulation patterns has revealed an arrangement of functional associations that crosses the boundaries of classically defined pathways [1,2]. Therefore, a 'dromocentric' view of metabolism (from the Greek word *dromos*, meaning pathway) should be complemented by a 'modulocentric' approach, in which modules rather than conventional pathways are the elementary functional units. Grouping of genes by modules rather than by pathways might also provide cleaner and more significant signals when analyzing new metabolic gene expression data.

Choosing the right path

The cellular response to changing conditions often involves altering metabolic fluxes at crucial branching points by using a combination of allosteric, covalent and transcriptional regulation [3,9,10]. How important is the role of gene co-regulation at these junctions? Ihmels' group found that, at diverging branches, co-regulation of the incoming reaction and one of the two outgoing reactions is prevalent (Table 1) [7]. This pattern is suggested to confirm the preference for linear flows along pathways. Although coordinated regulation of enzymes along a linear pathway is a documented and fundamental *modus operandi* [10], it should not be expected to be ubiquitous and to necessarily imply linearity of the metabolic flow. In some cases, simultaneous flow through different branches is necessary, for example, for the allocation of precursors in the synthesis of essential biomass components [11,12]. In other cases, branching pathways might truly represent differentially regulated alternatives. The recurring expression patterns found by Ihmels *et al.* [7] also suggest the existence of general mechanisms for the dynamic control of fluxes. To illustrate these points, I have developed a simplified kinetic model of branch regulation (Figure 2), inspired by LaPorte *et al.* [9]. This addresses the question of how fluxes change under linear co-regulation compared with the case of regulation of just one of the outgoing reactions. The comparison in Figure 2 shows that linear co-regulation of two genes gives a much wider range of flux variability than single gene regulation, and prevents saturation. In addition, in this example, linear co-regulation produces a slower initial flux response, possibly increasing robustness at noisy low enzyme concentrations. Additional insight into the dynamics associated with different co-expression patterns would require a more systematic analysis of all regulatory strategies at branch

Corresponding author: Daniel Segrè (dsegre@genetics.med.harvard.edu).

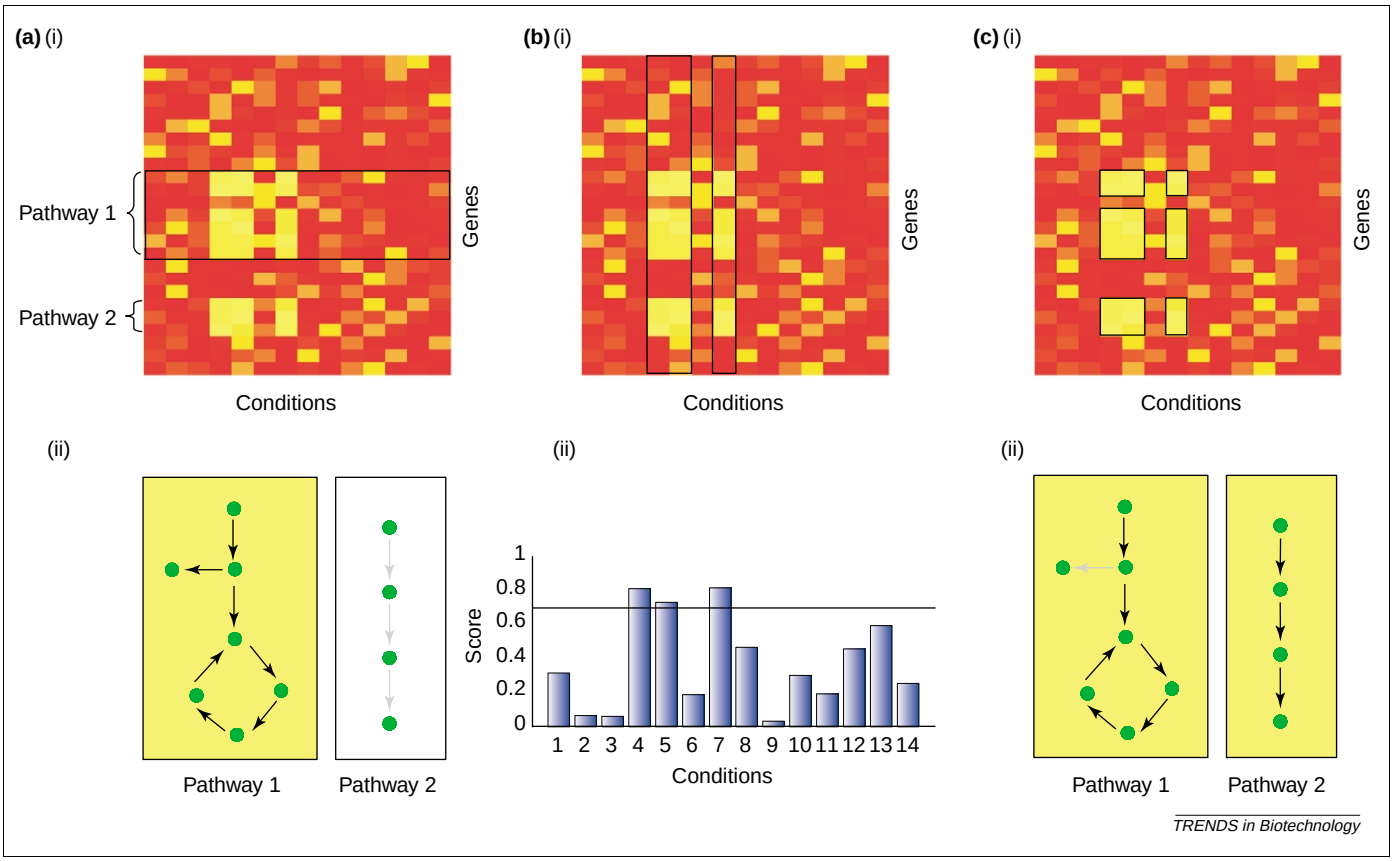


Figure 1. Use of the Signature Algorithm (SA) to identify co-regulation patterns and modular structure. The SA is a method that has been developed for finding transcriptional modules using gene-expression data and can be applied to the study of metabolic networks. **(a)** The first step in using the SA is the selection of a set of genes (and corresponding expression profiles; black rectangle) (i), which constitute the seed for determining the transcriptional module. For example, the choice could be a set of genes that corresponds to a topologically well-characterized metabolic pathway, such as Pathway 1 in (ii). **(b)** In the second step of the algorithm (i) a set of conditions is selected (black rectangles) for which the expression of the initially seeded genes significantly deviates from the average. This involves choosing a threshold parameter (ii) (horizontal line), which defines the cut-off for deciding which conditions are relevant (in this case, numbers 4, 5 and 7).**(c)** In the final step, genes that are highly expressed under the previously chosen conditions (those that are above a second threshold value) are selected (i). The end result is a set of genes and conditions or a module (black rectangles). The final module might contain additional genes that were not introduced in the initial seeding set (ii) (Pathway 2). Conversely, genes belonging topologically to the initial pathway might not belong the final co-regulated module (gray arrow).

points, possibly using metabolic control analysis [10], biochemical systems theory [13] or flux-oriented theory [14].

Isoenzyme switches

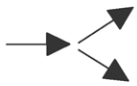
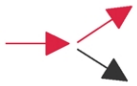
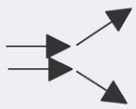
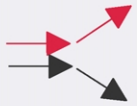
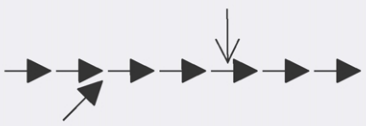
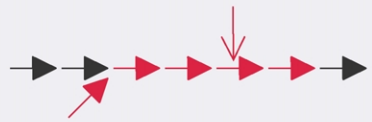
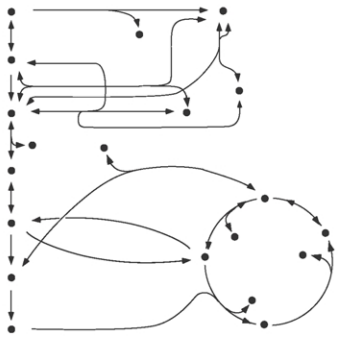
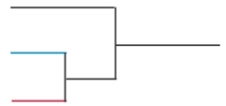
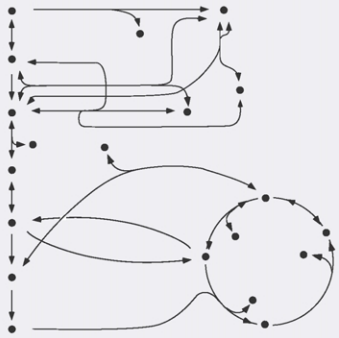
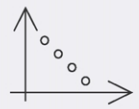
For several metabolic reactions, the catalyst can be one of two (or more) distinct enzymes (isoenzymes). One possible role of isoenzymes could be either the redundancy or enhancement of catalytic function. In such cases, one might expect isoenzymes to be co-regulated. Ihmels *et al.* [7], intrigued by the observation that isoenzymes often correlate with distinct modules, systematically analyzed isoenzyme co-expression patterns. In the majority of the examined pairs of isoenzymes (63%), they found that the two enzymes are not co-regulated. Rather, each of them (or at least one of them) correlates with a distinct enzyme that is an immediate neighbor in the metabolic network. This result indicates that different isoenzymes are frequently involved in distinct cellular responses that are controlled like switch mechanisms. For example, the two hexokinase isoenzymes HXK1 and HXK2 correlate with glycogen storage and glycolysis, respectively. It had also been proposed that this kind of specialized isoenzyme could be mediated by enzyme complex formation and channeling [15]. Interestingly, some pairs of isoenzymes might undergo both transcriptional and non-transcriptional regulation. For example, the glutamate dehydrogenases GDH1 and GDH3 – which, according to Ihmels *et al.*, correlate with different neighbor reactions – were also found to fulfill distinct tasks under allosteric control [16].

A switch in isoenzymes usage could be induced by unanticipated cellular requirements. Yeast grown in the presence of Cd^{2+} produces large quantities of glutathione for detoxification, which increases the need for sulfur [17]. This requirement triggers the replacement of sulfur-rich isoenzymes with sulfur-depleted equivalents. This transcriptional switch, especially evident in the ENO1–ENO2 and PDC1–PDC6 isoenzyme pairs [17], was not detected in the analysis by Ihmels *et al.* This serves as a reminder that the variety of conditions encountered during evolution might cover a much larger range of possibilities than those that are currently sampled experimentally. Conversely, this example suggests that some of the observed isoenzyme switches could be associated with other responses to amino acid shortage.

A global view of the regulatory software

At a whole-cell level, the hierarchical structure of metabolic functional modules can be inferred using an

Table 1. Summary of the metabolic regulation patterns and organization principles found by Ihmels *et al.* [7][a,b]

Level	Property	Observation	Examples
Reaction 	Linear co-regulation at branches 	Prevalence of branches at which one outgoing reaction is correlated with the incoming reaction	PFK1 and FBA1
Reaction 	Differential isoenzyme regulation 	Prevalence of isoenzyme pairs where one (or each) enzyme shows correlation with a connected reaction	(HXK1, YDR001C) and (HXK2 and PGI1)
Pathway 	Co-regulated core 	A subset of the enzymes in a given pathway displays correlation	24 out of 46 glycolysis genes
Pathway 	Transcriptional modules 	Transporters, transcription factors and feeder pathways are co-expressed with a given pathway	Transcription factor YGR067C co-expressed with glyoxylate cycle
Whole metabolism 	Hierarchical modularity 	By varying the stringency of co-regulation (or resolution), small specific modules gradually change and collapse into broader functional classes	Amino acid transport and glycolysis differentially regulated at high resolution, but co-regulated at low resolution (fast growth under rich conditions)
Whole metabolism 	Exponential distribution of module connectivity 	Exponential (rather than power law) distribution of connectivity when restricting to co-expressed subsets	See Ihmels *et al.* paper [7] and website

[a]The properties are sorted by the hierarchical level at which they are found, from small reaction branches to whole-cell metabolic pathways. Detailed information is contained in the databases available at: http://barkai-serv.weizmann.ac.il/MetabolicNetworks/
[b]Abbreviations: FBA, fructose-bisphosphate aldolase; HXK, hexokinase; PFK, 6-phosphofructokinase.

'unsupervised' variant of the Signature Algorithm – the Iterative Signature Algorithm [18]. By continuously decreasing a co-regulation threshold parameter, small modules merge into larger modules with less specific functions and decreasing levels of co-expression. Only three modules are found at low resolution: stress, protein synthesis and amino acid biosynthesis. Particularly striking is the division of these two last modules, possibly reflecting a distinction between fast growth under amino acid-rich media and slow growth under amino acid-poor media, respectively.

The classification proposed by Ihmels and colleagues focuses on the functional roles of the different genes, and is shown to differ from a previous analysis that was based on

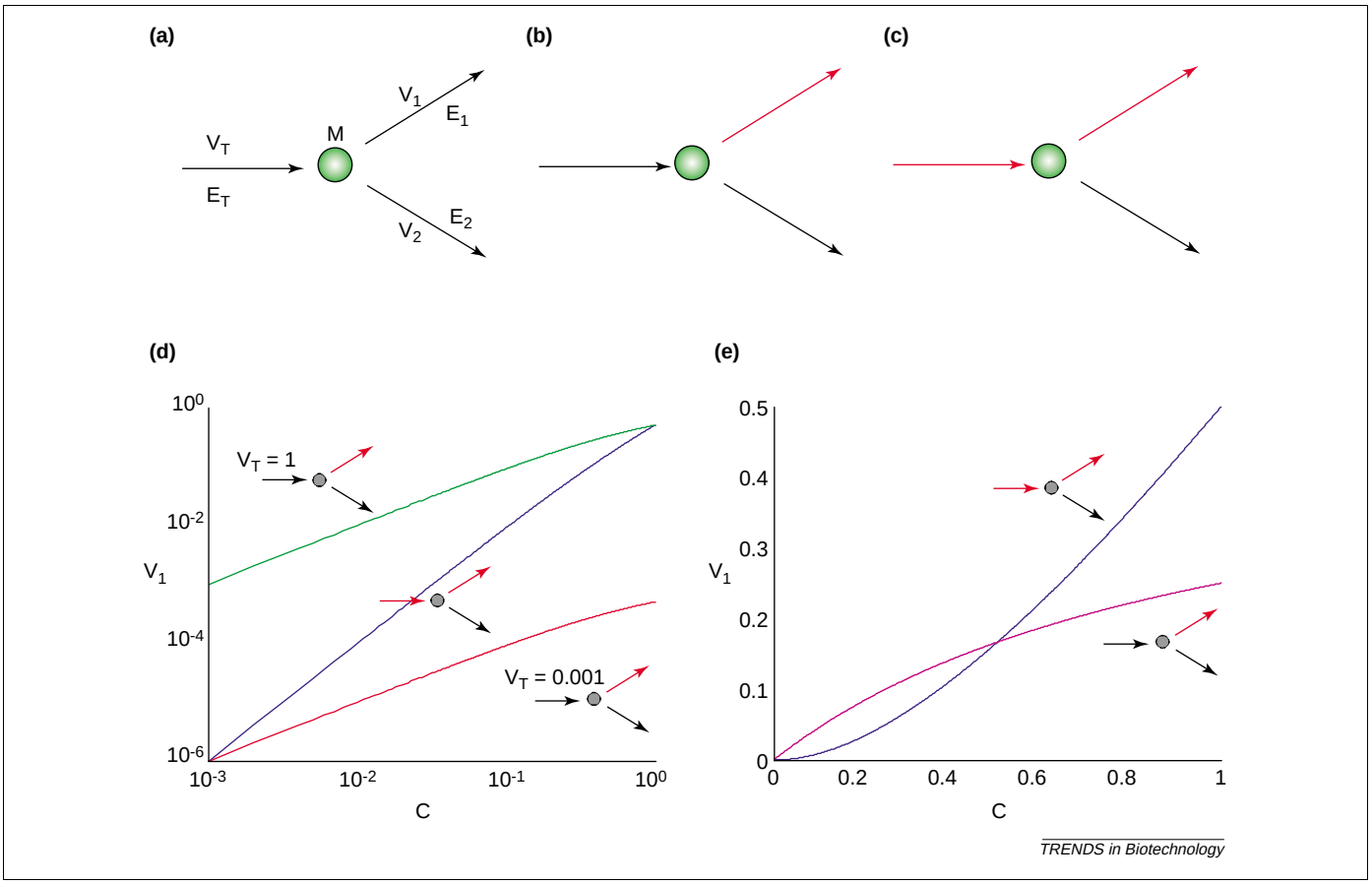


Figure 2. Observed co-expression patterns that could translate into dynamic properties of flux regulation. A simple kinetic model for a branch point in a metabolic network is presented here to illustrate the connection between enzyme concentrations and fluxes. **(a)** The topology of the branch, with one incoming reaction, producing metabolite M (green), and two outgoing reactions. The fluxes (V) and enzymes (E) associated with the reactions are indicated at the side of the corresponding arrows. Under the simplifying assumption that the Michaelis–Menten constants for reactions 1 and 2 are the same, the steady-state equations can be solved analytically, and obtain: $V_1 = V_T V_1^{MAX}/(V_1^{MAX} + V_2^{MAX})$, where V_i^{MAX} is the maximal rate for reaction i, which is proportional to the corresponding enzyme concentration. **(b)** One possible way of regulating the metabolic flow is to increase the concentration of E_1, the enzyme corresponding to flux V_1 (red arrow). The effect of this action is depicted in the red and green curves of (d), and in the magenta curve of (e). **(c)** In a different regulatory scheme, the enzyme for the incoming reaction, E_T, could be co-regulated with E_1 (red arrows). In this case, the response curve is shown in blue in (d) and (e). **(d)** The flux V_1 is plotted as a function of a control parameter C in log–log scale. For the green and red curves, the control parameter corresponds to $c = V_1^{MAX}$, and V_T has a fixed value of 1 and 0.001, respectively. These curves represent flux V_1 for the regulatory mechanism described in (b). For the blue curve [co-regulation, as described in (c)] the abscissa corresponds to $C = V_T = V_1^{MAX}$. We implicitly assume here that the substrate of the incoming reaction is buffered in such a way that V_T can be varied arbitrarily. **(e)** A graph similar to (e) in linear scale. The blue line corresponds to the blue line in (d). The magenta line represents the regulatory mechanism (b), for a fixed value of V_T (0.5). As described by the equation, single enzyme regulation approaches saturation, whereas co-regulation might lead to indefinite increase. Although, on one hand, the relatively slow onset for the co-regulation case might imply a weaker response, on the other, it could represent a mechanism for noise filtering.

network topology [19]. For example, fermentable and non-fermentable carbohydrate degradation pathways are far apart in the function-based hierarchy, despite their proximity in a topology-based classification. A different quantitative analysis of the relationship between gene regulation and network topology has recently been proposed (P. Kharchenko *et al.*, unpublished). They found that, on average, the closer two genes are in the metabolic network, the higher their correlation in expression. Above a characteristic pathway length, genes tend not to be significantly co-regulated. Altogether, these analyses of the connection between network topology and gene expression constitute new examples of data integration, which might be extended to include protein levels or other data.

Conclusions

The work of Ihmels *et al.* [7] helps our understanding of the regulatory program that connects single enzyme functions to cellular metabolic tasks, and provides clues as to which genes should be tinkered-with to produce desired metabolic flows. A clearer picture of the interplay between transcriptional and non-transcriptional regulation might come from comparisons of expression patterns with flux measurements [20]. Technologies for the measurement of *in vivo* fluxes are rapidly developing, mainly based on detection of ^{13}C-labeled biomolecules by mass spectrometry [12,21,22].

Cellular strategies for controlling metabolic flows could also be viewed in the light of optimality criteria. For example, flux balance analysis [11] uses a steady-state approximation and linear programming to find sets of fluxes that maximize or minimize a given metabolic objective. Recent efforts in flux balance modeling have succeeded in incorporating simplified regulatory interactions [23]. Alternative approaches have been proposed for the design of optimal enzyme regulation in a metabolic process [24], and for testing the hypothesis that optimal regulation protocols might be a consequence of natural selection [25]. In addition, the organization of transcriptional modules could have a mathematical counterpart in extreme pathways [26] and elementary mode [27] decompositions.

Our understanding of biological complexity clearly benefits from the representation of cellular organization and dynamics in terms of functional modules [6,8,28]. A major challenge for computational biologists will be to use modular organization to build tractable and predictive cell models. Transcription modules inferred from gene-expression data might be helpful in achieving this goal. In turn, the models would aid both our understanding of the interaction and evolution of modules, and the design of novel metabolic networks with desired functions.

Acknowledgements

I thank George Church, Matthew Wright, Patrik D'haeseleer, Xiaoxia Lin, Dennis Vitkup, Peter Kharchenko, Ayellet Falcovitz Segrè, Nikos Reppas, Dan Fraenkel, Kyriacos Leptos and Bruna Schreiber for their helpful comments and constructive criticism.

References

1 Krieger, C.J. *et al.* (2004) MetaCyc: a multiorganism database of metabolic pathways and enzymes. *Nucleic Acids Res.* 32, 438–442
2 Kanehisa, M. *et al.* (2004) The KEGG resource for deciphering the genome. *Nucleic Acids Res.* 32, 277–280
3 Fraenkel, D.G. (2003) The top genes: on the distance from transcript to function in yeast glycolysis. *Curr. Opin. Microbiol.* 6, 198–201
4 ter Kuile, B.H. and Westerhoff, H.V. (2001) Transcriptome meets metabolome: hierarchical and metabolic regulation of the glycolytic pathway. *FEBS Lett.* 500, 169–171
5 Strathern, J.N., *et al.* eds (1982) *The Molecular Biology of the Yeast Saccharomyces, Metabolism and Gene Expression*, Cold Spring Harbor Laboratory
6 Segal, E. *et al.* (2003) Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat. Genet.* 34, 166–176
7 Ihmels, J. *et al.* (2004) Principles of transcriptional control in the metabolic network of *Saccharomyces cerevisiae*. *Nat. Biotechnol.* 22, 86–92
8 Ihmels, J. *et al.* (2002) Revealing modular organization in the yeast transcriptional network. *Nat. Genet.* 31, 370–377
9 LaPorte, D.C. *et al.* (1984) The branch point effect. Ultrasensitivity and subsensitivity to metabolic control. *J. Biol. Chem.* 259, 14068–14075
10 Fell, D.A. (1998) Increasing the flux in metabolic pathways: a metabolic control analysis perspective. *Biotechnol. Bioeng.* 58, 121–124
11 Schilling, C.H. *et al.* (1999) Toward metabolic phenomics: analysis of genomic data using flux balances. *Biotechnol. Prog.* 15, 288–295
12 Fischer, E. and Sauer, U. (2003) Metabolic flux profiling of *Escherichia coli* mutants in central carbon metabolism using GC-MS. *Eur. J. Biochem.* 270, 880–891
13 Ni, T.C. and Savageau, M.A. (1996) Application of biochemical systems theory to metabolism in human red blood cells. Signal propagation and accuracy of representation. *J. Biol. Chem.* 271, 7927–7941
14 Crabtree, B. *et al.* (1997) Principles of regulation and control in biochemistry; a pragmatic, flux-orientated approach. In *Handbook of Physiology: Cell Physiology* (Hoffman, J. and Jamieson, J.D., eds), pp. 117–180, Oxford University Press
15 Ureta, T. (1991) The role of isozymes in metabolite channelling. *J. Theor. Biol.* 152, 81–84
16 DeLuna, A. *et al.* (2001) NADP-glutamate dehydrogenase isoenzymes of *Saccharomyces cerevisiae*. Purification, kinetic properties, and physiological roles. *J. Biol. Chem.* 276, 43775–43783
17 Fauchon, M. *et al.* (2002) Sulfur sparing in the yeast proteome in response to sulfur demand. *Mol. Cell* 9, 713–723
18 Bergmann, S. *et al.* (2003) Iterative signature algorithm for the analysis of large-scale gene expression data. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 67, 031902
19 Ravasz, E. *et al.* (2002) Hierarchical organization of modularity in metabolic networks. *Science* 297, 1551–1555
20 Daran-Lapujade, P. *et al.* (2003) Role of transcriptional regulation in controlling fluxes in central carbon metabolism of *Saccharomyces cerevisiae*, a chemostat culture study. *J. Biol. Chem.* 279, 9125–9138
21 Fiaux, J. *et al.* (2003) Metabolic-flux profiling of the yeasts *Saccharomyces cerevisiae* and *Pichia stipitis*. *Eukaryot. Cell* 2, 170–180
22 Klapa, M.I. *et al.* (2003) Systematic quantification of complex metabolic flux networks using stable isotopes and mass spectrometry. *Eur. J. Biochem.* 270, 3525–3542
23 Herrgard, M.J. *et al.* (2003) Reconciling gene expression data with known genome-scale regulatory network structures. *Genome Res.* 13, 2423–2434
24 Hatzimanikatis, V. *et al.* (1996) Optimization of regulatory architectures in metabolic reaction networks. *Biotechnol. Bioeng.* 52, 485–500
25 Klipp, E. *et al.* (2002) Prediction of temporal gene expression. Metabolic opimization by re-distribution of enzyme activities. *Eur. J. Biochem.* 269, 5406–5413
26 Price, N.D. *et al.* (2003) Analysis of metabolic capabilities using singular value decomposition of extreme pathway matrices. *Biophys. J.* 84, 794–804
27 Stelling, J. *et al.* (2002) Metabolic network structure determines key aspects of functionality and regulation. *Nature* 420, 190–193
28 Hartwell, L.H. *et al.* (1999) From molecular to modular cell biology. *Nature* 402, C47–C52

doi:10.1016/j.tibtech.2004.04.013

Erratum

Erratum: Organ printing, computer-aided jet-based 3D tissue engineering

Trends in Biotechnology 21 (2003) 4, 157–161

In the April 2003 issue of *Trends in Biotechnology* the source of the gels in the figures was not adequately specified. The gel in Figures 3e and 3f is a thermostable poly (N-[2-hydroxypropyl] methacrylamide)-containing RGD peptide hydrogel, developed by Dr Stephane Woerly, and is the exclusive property and technology of Organogel Canada Ltd (http://www.organogel.com/) [1]. The authors of the paper, and the staff of *Trends in Biotechnology* apologise to the readers and to Organogel Canada Ltd for any confusion this might have caused. DOI of original article: 10.1016/S0167-7799(03)00033-7

References

1 Woerly, S. *et al.* (2001). Spinal cord restoration with NeuroGel containing RGD peptides. *Biomaterials* 22, 1095–1111

doi:10.1016/j.tibtech.2004.04.004