1 A new millenium: functional genomics, cellular bioinformatics, regulation analysis

H.V. Westerhoff[1,2], K.A. Reijenga[1], J.L. Snoep[1,3], B.N. Kholodenko[4], and B.H. ter Kuile[5]

[1]Departments of Molecular Cell Physiology, and
[2]Mathematical Biochemistry, BioCentrum Amsterdam, Faculty of Biology, Free University, De Boelelaan 1087, NL-1081 HV Amsterdam, EU
[3]Department of Biochemistry, University of Stellenbosch, Private bag XI, 7602 Matieland, South Africa
[4]Department of Pathology, Anatomy and Cell Biology Thomas Jefferson University, 1020 Locust St Philadelphia, PA 19107, USA
[5]Leiden Institute of Chemistry, Gorlaeus Laboratories, P.O. Box 9502 NL-2300 RA, Leiden, EU

Introduction

Biology almost made it! Just a year or two after the end of the past millennium mankind shall avail of the complete DNA sequence of its own genome, hence of the complete amino acid sequence of the catalysts of virtually all its own chemical and physical processes. The information explosion is excessive: there are more than 50 000 such processes in a single human being, and these can be compared to similar and other processes in more than 100 other systems that are capable of living and of which the genome sequence is known.

However, the new information itself is 'dead'. Proteins neither assume a linear conformation, nor a single alpha helical structure. Their catalytic properties depend on their three dimensional structure. The latter is partly defined by the primary structure, but the relation between the two types of structure is so complex that the catalytic properties cannot be calculated *ab initio* from the amino-acid sequence. Presently 3-D structure prediction of proteins focuses on homologies with known structures. Moreover, the functioning of the enzymes in the living cell depends less on their precise 3-D structure than on their kinetic properties, such as K_M and K_d, and we are far from calculating these *ab initio*. In addition, the route to the living cell runs through metabolic pathways, which are subject to hierarchi-

cal control systems; consequently they depend on the magnitudes of all kinetic parameters simultaneously, on the gene-expression patterns and on topological organization. Consequently, although we are far from the beginning in terms of the dead information concerning life, we are remote from the end of having the *live* information, and from understanding the latter.

To obtain the 'live' information, sequencing cloned DNA does not suffice. Yet, having the sequences available has led to revolutionary approaches to analyzing the living cell. Through the hybridization array procedure one is now able to determine the expression of all genes at the mRNA level. Proteomics and metabolomics will soon provide expression information at the protein and the metabolic level. Confocal and single molecule fluorescence methods allow inspection of the processes in the living cell [1]. Accordingly, in the early years of the new millennium, we shall begin to obtain more, and more complete, information about the living cell [2]. If all information will be available soon, is there anything left to be done for BioThermoKinetics?

Science is much more than data collecting; it aims at understanding phenomena. The discipline of BioThermoKinetics is strong in effecting such understanding, especially at levels transcending that of the single molecule; it mostly addresses what arises from the interaction of molecules. As the number of molecules about which information is available increases, the number of relevant potential interactions increases more strongly. In addition, the fact that the sequence and expression information is that of *complete* genomes invalidates the sledgehammer argument against BioThermoKinetics that there are too many unknowns even to begin to understand the living cell. Consequently, BioThermoKinetics may exert a high control on 3rd millenium science, provided that it gears up to deal with the new type of experimental data.

Here we address the BioThermoKinetic analysis of two types of the new information. One is the measurement of pathway fluxes together with the activities of all the enzymes in the pathway. The other is the measurement of all mRNA concentrations that correspond to those enzymes. We shall develop a method to disentangle metabolic from hierarchical regulation of the enzyme activities in the living cell, and shall distinguish between translation and transcription components of the latter.

Results

Metabolome and beyond

We begin from the concept [3,4] that the rate (v, which is taken as positive) of an enzyme catalyzed reaction can be written as a function of the concentration of the enzyme that catalyzes the reaction (e), and the concentrations of substrates, products and other metabolic modifiers (X).

$$v = v(e, X) = f(e) \cdot v(X) \tag{1.1}$$

A feature we shall *not* make use of here, is that the dependence on enzyme concentration can usually be written as a multiplier without cross dependence (eq. 1.1). We consider a steady-state flux J running through any enzyme (at stoichiometry 1) that changes in response to a change in intracellular and extracellular conditions. This steady-state flux equals the rate catalyzed by that enzyme at steady state [3]. Dividing the relative change in the (steady-state) rate of the reaction catalyzed by the enzyme, by the relative change in flux, one finds (for a full derivation, cf. Westerhoff *et al.*, in preparation) a new summation theorem for global regulation, *i.e.*, that:

$$1 = \frac{d \ln v}{d \ln J} = \rho_h + \rho_m \tag{1.2}$$

Here ρ_h and ρ_m quantify the hierarchical and the metabolic regulation coefficients, respectively, of the activity of the enzyme in the living cell. Changes in the flux can be caused by (i) changes in substrate and/or end product concentration (metabolic regulation), and (ii) changes in enzyme levels (regulation through gene expression). The theorem states that the sum of hierarchical and metabolic regulation of the rate catalyzed by an enzyme, must equal 1. The definitions of the regulation coefficients are:

$$\rho_h \equiv \left(\frac{\partial \ln v}{\partial \ln e}\right)_X \cdot \frac{d \ln e}{d \ln J} \tag{1.3}$$

$$\rho_m \equiv \left(\frac{\partial \ln v}{\partial \ln X}\right)_e \cdot \frac{d \ln X}{d \ln J} = \sum_{j=1}^{p} \left(\frac{\partial \ln v}{\partial \ln X_j}\right)_{e,X_k} \cdot \frac{d \ln X_j}{d \ln J} \tag{1.4}$$

The latter expression for the metabolic regulation coefficient makes explicit that the middle expression is shorthand for a conglomerate of metabolic regulations of the enzyme activity.

The regulation coefficients defined here are related more closely to co-response coefficients [5,6] than to control coefficients:

$$\rho_m = \varepsilon_X \cdot \Omega^{X:J} = \sum_{j=1}^{p} \varepsilon_{X_j} \cdot \Omega^{X_j:J} \tag{1.5}$$

$$\rho_h = \varepsilon_e \cdot \Omega^{e:J} \tag{1.6}$$

The elasticity in eq. 1.6 has also been denoted by the letter π [7] and is often equal to 1, i.e., when adding copies of the enzyme at unchanged substrate and product concentrations leads to a proportional increase in overall activity.

In the case of metabolic regulation only, ρ_h equals zero and:

$$1 = \sum_{j=1}^{p} \varepsilon_{X_j} \cdot \Omega^{X_j:J} \tag{1.7}$$

For the special case that the regulatory events are due to the modulation of a single enzyme (different from the one the rate of which is being considered), the co-response coefficients, can be replaced by co-control coefficients:

$$1 = \sum_{j=1}^{p} \varepsilon_{X_j} \cdot O^{X_j:J} \tag{1.8}$$

This equation has already been derived by Sauro [8] who called its terms 'partitioned regulatory coefficients' and noted their relationship with the 'conditional elasticities' defined by Holzhütter and colleagues [9]. Sauro showed how this relationship could be used to establish the relative importance of metabolic control through a feedback loop relative to the metabolic control by product inhibition and substrate stimulation [8]. Although Sauro called *this* control 'regulation', we suggest that it be called 'control' (or 'internal regulation') as it refers to the control by an enzyme activity. It does not refer to a regulatory event that the *system* may be experiencing such as through an activation from the outside. For the latter analysis, the expression in terms of co-response coefficients, i.e., eq. 1.7, may be more useful.

Although we present eq. 1.2 as a new *law* for Control Analysis, it should be noted that its analogues for exclusively metabolically regulated systems (e.g., eq. 1.7) were already known. In this sense eq. 1.2 could be seen as a mere extension of eq. 1.8 to systems with regulated gene expression. From the biological point of view though, the added consideration of regulated gene expression as a determinant of fluxes, is likely to prove important.

There is an interesting asset to eq. 1.2, as compared to eq. 1.8. Whereas Sauro noted that the subdivision of eq. 1.7 in partitioned regulatory coefficients could only work for small changes [8], this limitation may not exist for our theorem (eq. 1.2). Eq. 1.1 indicates why not.

We wish to stress that although eq. 1.2 should allow one to decide the relative extents to which metabolic and gene-expression regulation are responsible for the regulation of a steady-state reaction rate (i.e., flux), it does not address all aspects of regulation. As was discussed by Hofmeyr and Cornish-Bowden [6], one may wish to understand homeostasis in the sense of how much in a system a flux can be changed whilst keeping the changes in metabolite concentrations small. Then neither eqs. 1.7, 1.8, nor 1.2, lead to an answer and the partitioning of coresponse coefficients needs to be considered. Or, one may wish to understand how homeostasis of a concentration is effected, for which there is yet another stategy [10].

Transcriptome versus proteome

Most genome-wide expression analyses measure mRNA concentrations. Recent evidence suggests that mRNA levels are not always good predictors of the corresponding protein concentrations, not even at steady state [11]. Here we develop a method to quantify the correlation between message and enzyme levels. This method focuses on the process of translation of mRNA that encodes a certain enzyme in the cell at a rate v_{trans}. When external conditions change, cells progress to a new steady state. mRNA levels, concentrations of amino acids, enzyme concentrations, and rates of translation may all change as well. These changes are related as follows:

$$\frac{1}{\Omega^{e}_{\text{trans}}} = \frac{d\ln v_{\text{trans}}}{d\ln e} = \rho_{rh} + \rho_{rm} + \rho_{rr} \tag{1.9}$$

In the likely event that the protein e is degraded with first order kinetics, translation is not subject to product inhibition by the enzyme it synthesizes, and regulation is through translation, this expression equals 1:

$$1 = 1 - 0 = \varepsilon_e^{\text{degradation}} - \varepsilon_e^{\text{translation}} = \rho_{rh} + \rho_{rm} + \rho_{rr} \tag{1.10}$$

The hierarchical and metabolic regulation coefficients of translation are defined, respectively as:

$$\rho_{rh} \equiv \left(\frac{\partial \ln v_{\text{trans}}}{\partial \ln \text{mRNA}}\right)_{Y,\text{ribo}} \cdot \frac{d\ln \text{mRNA}}{d\ln e} \tag{1.11}$$

$$\rho_{rm} \equiv \left(\frac{\partial \ln v_{\text{trans}}}{\partial \ln Y}\right)_{\text{mRNA,ribo}} \cdot \frac{d\ln Y}{d\ln e} \tag{1.12}$$

$$\rho_{rr} \equiv \left(\frac{\partial \ln v_{\text{trans}}}{\partial \ln \text{ribosome}}\right)_{\text{mRNA},Y} \cdot \frac{d\ln \text{ribosome}}{d\ln e} \tag{1.13}$$

It should be noted that the third term, i.e., ρ_{rr}, is also a hierarchical one, referring to the possibility of the cell to regulate the concentration of its translation machinery. The subscript 'r' refers to t**r**anslation, the subscript 'ribo' to **ribo**some concentration, the subscript 'h' to **h**ierarchical, and 'm' to **m**etabolic, i.e., metabolic (or other) signals affecting translation activity. 'Y' refers to the concentrations of amino acids and ATP.

Discussion

Two new theorems have been added to the arsenal of Hierarchical Metabolic Control Analysis [12,13]. One (eq. 1.2) states that the regulation of *in vivo* enzyme rate consists of two fractions that should add up to 100 %. The fractions are metabolic

and hierarchical. The latter, i.e., eq. 1.9, stems from levels beyond metabolism itself. This hierarchical regulation can be further disentangled into a part that runs through gene expression and a part that runs through signal transduction and covalent modification, but this has not been elaborated here.

In recent experiments [8], one of us has determined how enzyme concentrations varied with metabolic fluxes, i.e., $\Omega^{e;J}$. With the above equations and assumption, this should enable one to quantify for each enzyme how much of the 100 % regulation of its *in vivo* rate is metabolic and how much is hierarchical (Ter Kuile & Westerhoff, in preparation).

The other theorem (Eqs. 1.9 and 1.10) addresses regulation of translation. Also here, the total regulation tends to add up to 100 % (although in the general case the sum may differ from 100 %, cf. above) and consists of the 'hierarchical' regulation through the concentration of the mRNA and the concentration of the translation machinery, and regulation through other factors affecting translation. The silent assumption behind the massive funding of hybridization array machines is that $\rho_{rm} = 0$, which is not what one of us has been finding [11].

In a recent study the intracellular regulation of DNA supercoiling in *E.coli* has been analyzed by a different method that also led to the quantitative discrimination between hierarchical and direct regulation [14]. This is elaborated further in another chapter in this book (cf. eq. 28.2 in [15]).

The reader will note that the hierarchical regulation theory devised here embraces many subtleties that need to be addressed. These include the implications of the control hierarchy being democratic, and the parameter dependence of the regulation coefficients. In addition, a similar treatment of transcription regulation is possible. By virtue of their focus on one enzyme at a time, the approach developed here also relate to the recent Metabolic Design method [16].

Acknowledgement

We thank Jannie Hofmeyr for critical reading of the manuscript and numerous suggestions for its improvement.

References

1. http://www.bio.vu.nl/GRC2000

2. Westerhoff, H.V. (2000) Live Perspectives of Biochemistry. *The Biochemist*, in the press.

3. Heinrich, R. and Rapoport, T.A. (1973) Linear theory of enzymatic chains; its application for the analysis of the crossover theorem and of the glycolysis of human erythrocytes. *Acta Biol. Medica Germ.* **31**, 479–494.

4. Kacser, H. and Burns, J. (1973) In *Rate Control of Biological Processes* (Davies D.D., ed.), pp. 65–104, Cambridge University Press, London.

5. Hofmeyr, J.-H.S., Cornish-Bowden, A. and Rohwer, J.M. (1993) Taking enzyme kinetics out of control; putting control into regulation. *Eur. J. Biochem.* **212**, 833–837.

6. Hofmeyr J.-H.S. and Cornish-Bowden A. (1996) Co-response analysis: a new experimental strategy for metabolic control analysis. *J. Theor. Biol.* **182**, 371–380.

7. Kacser, H., Sauro, H.M. and Acerenza, L. (1990) Enzyme-enzyme interactions and control analysis The case of non-additivity: monomer-oligomer associations. *Eur. J. Biochem.* **187**, 481–491.

8. Sauro H. (1990) Regulatory Responses and Control Analysis: Assessment of the relative importance of internal effectors. In *Control of Metabolic Processes* (Cornish-Bowden, A., ed.), pp. 225–230, Plenum, New York.

9. Holzhütter H.G., Jacobasch G. and Bisdorff A. (1985) Mathematical modelling of metabolic pathways affected by an enzyme deficiency. A mathematical model of glycolysis in normal and pyruvate-kinase-deficient red blood cells. *Eur. J. Biochem.* **149**, 101–111.

10. Kahn, D. and Westerhoff, H.V. (1993) The Regulatory Strength: How to be Precise about Regulation and Homeostasis. *Acta Bioth.* **41**, 85–96.

11. Ter Kuile, B.H. (1999) Regulation and adaptation of glucose metabolism of the parasitic protist *Leishmania donovani* at the enzyme and mRNA levels. *J. Bacteriol.* **181**, 4863–4872.

12. Kahn, D. and Westerhoff, H.V. (1991) Control Theory of Regulatory Cascades. *J. Theor. Biol.* **153**, 255–285.

13. Westerhoff, H.V., Koster, J.G., Van Workum, M. and Rudd, K.E. (1990) On the control of gene expression. In *Control of Metabolic Processes* (Cornish-Bowden, A., ed.), pp. 399–412, Plenum, New York.

14. Ruhdal Jensen, P., Van der Weijden, C.C., Jensen, L.B., Westerhoff, H.V. and Snoep, J.L. (1999) Extensive regulation compromises the extent to which DNA gyrase controls DNA supercoiling and growth rate of *Escherichia coli. Eur. J. Biochem.* **266**, 865–877.

15. Snoep, J.S., Van der Weijden, C.C., Andersen, H.W., Westerhoff, H.V. and Jensen, P.R. (2000) Hierarchical control of DNA supercoiling in *Escherichia coli*; how to study homeostatically controlled systems using control analysis. This volume, p. 181–187.

16. Kholodenko, B.N., Cascante, M., Hoek, J.B., Westerhoff, H.V. and Schwaber, J. (1998) Metabolic design: How to engineer a living cell to desired metabolite concentrations and fluxes. *Biotechnol. Bioeng.* **59**, 239–247.