Interpreting correlations in metabolomic networks

R. Steuer*[1], J. Kurths*, O. Fiehn† and W. Weckwerth†

*Arbeitsgruppe Nichtlineare Dynamik, Institut für Physik der Universität Potsdam, Am Neuen Palais 10, Haus 19, 14469 Potsdam, Germany, and †Max-Planck Institut für Molekulare Pflanzenphysiologie, Am Mühlenberg 1, 14476 Golm, Germany

Abstract

Correlations, as observed between the concentrations of metabolites in a biological sample, may be used to gain additional information about the physiological state of a given tissue. In this mini-review, we discuss the integration of these observed correlations into metabolomic networks and their relationships with the underlying biochemical pathways.

Introduction

Despite the extensive knowledge about fundamental biochemical processes gathered in the last century, the regulation and large-scale organization of metabolism in multicellular organisms is still poorly understood [1]. While much effort has been put into transcript and protein profiling, the identification and quantification of metabolites at the systems level has been largely neglected [2,3]. Only recently, novel experimental techniques based on gas and liquid chromatography coupled to MS have allowed more than 1000 distinct compounds in single plant extracts to be detected and quantified [4,5].

This novel metabolomic approach aims to cover as many individual metabolites as possible, with the ultimate goal of an unbiased and comprehensive quantification of all metabolites present in a biological sample [6]. To investigate the structure of metabolism from measurements of cellular metabolite levels, we recently proposed to use metabolite–metabolite correlation analysis and its graphical visualization via metabolic correlation networks [7,8]. In this mini-review, we will discuss the interpretation and analysis of such data-generated networks in terms of the underlying biochemical pathways.

Metabolomic network analysis

A metabolomic experiment usually results in an $M \times N$ matrix of (relative) metabolite concentrations, where M and N denote the number of metabolites and samples respectively. Remarkably, even when all samples are obtained from an ensemble of identical genotypes under highly controlled conditions, the metabolite levels show a large biological variability. This observed biological variability considerably exceeds the relative technical standard deviation [4,5]. More importantly, for a given metabolite, the variation is not independent, but may correlate with the variation of other metabolites, as shown in Figure 1.

Key words: data mining, mass spectrometry, metabolic networks, metabolomics.

[1]To whom correspondence should be addressed (e-mail steuer@agnld.uni-potsdam.de).

©2003 Biochemical Society

The observed correlation between any two metabolites can be quantified using the Pearson correlation coefficient or other related similarity measures [9]. To visualize the resulting correlation matrix, the metabolites are then integrated into a metabolic correlation network, as described in Figure 1. The resulting networks are highly complex and their relationship to the underlying biochemical system is as yet unclear. While some metabolite–metabolite correlations are easy to understand in terms of classic biochemical knowledge, such as the correlation between glucose 6-phosphate and fructose 6-phosphate found in all data analysed so far, most correlations defy an intuitive biochemical interpretation (see e.g. β-alanine and serine in Figure 1). Thus, how can one deduce novel insights into the structure of metabolism from the resulting metabolomic networks? Is there a systematic connection between the underlying enzymic system and the observed correlations?

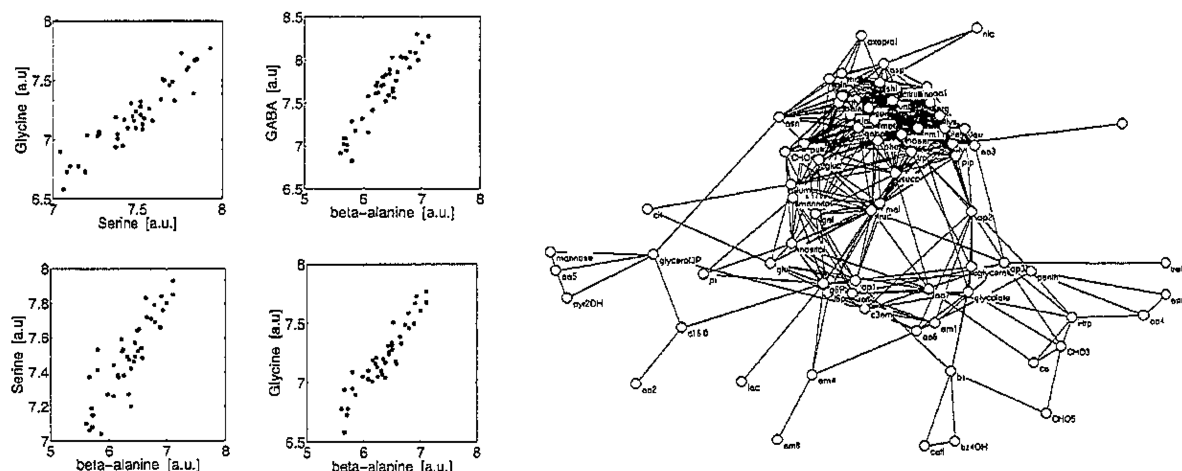
Interpreting the correlations

As a prerequisite for further analysis, we have to connect the experimentally observed correlations to the underlying biochemical system. To this end, we have recently argued that the observed variability in a homozygous plant population must have biological causes, reflecting the intrinsic flexibility of metabolic networks [1]. We thus adopt the following working hypothesis: cell metabolism constitutes a complex dynamical system, which is continuously subject to random fluctuations. These fluctuations induce variability in certain metabolites, propagate through the network, and ultimately create an emergent pattern of correlations. More specifically, we assume that, even under uniform and highly controlled experimental conditions, metabolite levels are never constant, but continuously changing due to subtle variations in the environment, as well as due to complex patterns of regulation in (other parts of) the cellular system.

For example, potato tubers rely on the supply of certain metabolites. This supply is likely to fluctuate. To investigate how the metabolic network responds to such fluctuations and gives rise to observable correlations, we recently studied a prototypical example numerically using stochastic

Figure 1 | Metabolic correlation networks

Left-hand panel: examples of metabolite–metabolite scatter plots using data from 44 potato-tuber samples obtained simultaneously from identical genotypes at the same developmental stage. All data are log-transformed and reported in arbitrary units. Right-hand panel: a metabolic correlation network for potato tuber. Each metabolite is assigned to co-ordinates in a two-dimensional plane, such that the pairwise distances approximately reflect the similarity given by the correlation matrix. If their correlation exceeds a given threshold, two metabolites S_i and S_j are connected with a link.



differential equations [10]. In this case, the time evolution of certain (external) metabolites is assumed to be a time-dependent stochastic variable. Hypothetical experiments are performed by numerically integrating the system and recording the concentration of all metabolites at a given point in time. The resulting pattern of correlations can then be compared with the (in this case known) model of metabolism.

In conformity with experimental findings, the example above clearly demonstrated that there is no straightforward connection between the observed correlations and the underlying reaction network. We observed strong correlations between seemingly distant metabolites, whereas metabolites sharing a common reaction are not necessarily correlated. Thus, a solely intuitive interpretation of the observed correlations must almost unavoidably fail.

However, some information about the underlying reaction system is still embedded in the observed correlation matrix. To elucidate this connection, we can make use of the theory of stochastic dynamical systems [11]. Using a linear approximation of the system, it is possible to give an analytical description that provides a link between the observed correlation matrix and the Jacobian matrix of the system [10]. Thus, our result leads to a systematic relationship between the experimentally observed correlations and the underlying biochemical pathways. While the resulting equations are not directly applicable to reverse engineer metabolic pathways from data, they provide a fundamental conceptual basis for the analysis of metabolomic datasets. Our approach enables researchers to treat the observed correlations as a snapshot of the physiological state of a given tissue at a given point in time [10].

Outlook

One of the primary applications of all -omics technologies is to compare two types of sample, to identify differential behaviour of individual components. This approach usually neglects biological variability by using average concentration levels, with the underlying assumption that identical genotypes should ideally produce the same steady state under highly controlled conditions [6].

In extension to this, we propose using the large observed biological variability for network analysis. According to our hypothesis, the observed correlations between metabolite concentrations are a consequence of the underlying enzymic reaction network, thus providing a fingerprint of the current state of the system [10]. In this way, the organization of metabolites in complex correlation networks exploits the intrinsic flexibility of metabolism to gain additional information about a molecular system.

Indeed several recent experimental findings support our theoretical consideration outlined above. For example, different environmental conditions, or plants suppressed in the expression of specific enzymes, manifest themselves by having (slightly) altered Jacobians. Thus, we must expect those systems to show not only different steady state levels, but also (slightly) altered correlation networks: and this is indeed observed experimentally [12]. On the other hand, we find that the overall structure of metabolic correlation networks is highly preserved under different conditions, indicating a certain robustness of metabolite ratios, despite large variations in average metabolite levels [7].

Thus, despite the considerable technical difficulties that still need to be solved, we can be rather optimistic that

©2003 Biochemical Society

metabolomic methods will lead to a better understanding of metabolic networks. The first steps have been made, but metabolomic analyses have only just begun.

We would like to thank Megan McKenzie for editing the manuscript. R.S. is supported by the HSP-N grant of the State of Brandenburg.

References

1 Weckwerth, W. and Fiehn, O. (2002) Curr. Opin. Biotechnol. **13**, 156–160
2 Fiehn, O. (2001) Comp. Funct. Genom. **2**, 155–168
3 Fiehn, O. (2002) Plant Mol. Biol. **48**, 155–171
4 Fiehn, O., Kopka, J., Dörmann, P., Altmann, T., Trethewey, R.T. and Willmitzer, L. (2000) Nat. Biotechnol. **18**, 1157–1161
5 Roessner, U., Wagner, C., Kopka, J., Trethewey, R.N. and Willmitzer, L. (2000) Plant J. **23**, 131–142
6 Weckwerth, W. (2003) Annu. Rev. Plant Biol. **54**, 669–689
7 Fiehn, O. (2003) Phytochemistry **62**, 875–886
8 Kose, F., Weckwerth, W., Linke, T. and Fiehn, O. (2001) Bioinformatics **17**, 1198–1208
9 Steuer, R., Kurths, J., Daub, C.O., Weise, J. and Selbig, J. (2002) Bioinformatics **18**, S231–S240
10 Steuer, R., Kurths, J., Fiehn, O. and Weckwerth, W. (2003) Bioinformatics **19**, 1019–1026
11 van Kampen, N.G. (1992) Stochastic processes in physics and chemistry, Elsevier Science, Amsterdam
12 Weckwerth, W., Tolstikov, V.V. and Fiehn, O. (2001) Proceedings of the 49th ASMS Conference on Mass Spectrometry and Allied Topics, pp. 1–2, May 27–31, Chicago

Received 22 July 2003

©2003 Biochemical Society