From genome to cellular phenotype—a role for metabolic flux analysis?

Athel Cornish-Bowden and María Luz Cárdenas

More than 20 genomes have been completely sequenced, and the first draft of the complete human genome is expected this year[1]. This barrage of information should eventually lead to a better understanding of cellular physiology, which in turn may then be exploited to advance human health, agricultural production, and industrial fermentation. If the physiology of the organism under study is relatively well understood, such as for *Escherichia coli* or yeast, the compilation of this sequence data to reconstruct metabolic pathways may seem a reasonable challenge. However, if there is only limited biochemical information available, such as for *Treponema pallidum* (the causative agent of syphilis), a partial list of putative enzymes is unlikely to allow metabolic networks and cellular phenotypes to be established. Realistically, a list of genes alone is not enough to understand the pathophysiology or manipulate the metabolism of *T. pallidum*. Thus, a robust method of converting a list of putative enzymes into a set of metabolic pathways could be a powerful bioinformatic tool. In this issue, Schuster et al.[2] discuss a means of defining pathways rigorously and show how to express their constitutive activities as the sum of "elementary flux modes", each of which represents a minimal sequence of steps that can in principle operate independantly of any others.

The path from genome to phenotype involves multiple steps (Fig. 1). Although considerable efforts have been devoted to improving the early steps, we are only beginning to get a handle on the later ones. Sequencing the genome is now just a matter of time and money, and the identification of gene function is becoming easier as a result of advances in comparative sequence analysis[3]. However, the increasing number of orphan sequences will require considerable effort to understand. It is thus not surprising that the development of computational algorithms to predict protein function from amino acid sequences is now a core aim of bioinformatics[4]. However, even for *Caenorhabditis elegans,* with a much more compact and nonrepetitive genome than the human, as many as half of the proteins may be incorrectly identified[1]—an observation that argues strongly for the need for an independent approach to locate errors in the results obtained from sequence analysis. In any case, it would be a serious mistake to conclude that the work ends with a complete list of the enzymes active in a given organism. Predicting a phenotype requires that the metabolic transformations are stoichiometrically, thermodynamically, and kinetically possible. This problem has often been regarded as if it were trivial, but it is not.

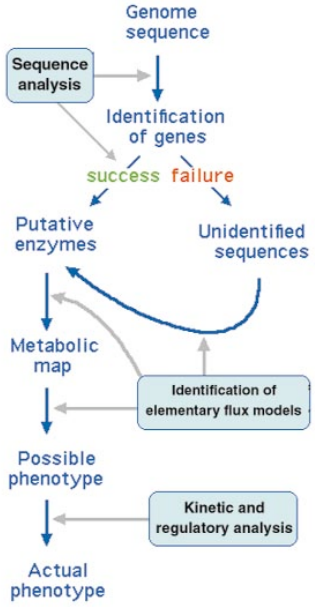

Figure 1. The necessary steps to convert a genome sequence into a phenotypic description include not only the early stages of identifying the genes, but also later ones that have received less attention: assembling the enzyme-catalyzed reactions into a metabolic map, determining what kinds of metabolic activity are stoichiometrically possible, and finally, determining which of these are kinetically possible. Identification of elementary flux modes is especially useful for defining the stoichiometric structure of the map, but it can also aid in identifying the roles of orphan genes remaining after bioinformatics studies have permitted the identification of other genes by comparison with the sequences of known enzymes.

Elementary flux modes can be determined mathematically from the list of reactions in the system, without any kinetic information. Any conceivable set of fluxes becomes a weighted sum of the elementary modes. Glycolysis and the pentose phosphate pathway, for example, can be regarded in the simplest case as the sum of seven elementary modes. This may seem very abstract and only of interest to the theoretical biologist, without any practical implications for biotechnology. Indeed, why should a biotechnologist care whether metabolic pathways are rigorously defined or not?

To understand better why this is important, consider the following case. An enzyme appears to have a key role in controlling a flow of metabolites that we wish to suppress, so as to kill a harmful organism, or to increase the yield of a commercially valuable end product by decreasing a wasteful flow of intermediates into another pathway. How certain can we be that inhibiting the target enzyme or deleting its gene will actually affect the flux in the expected way? If there are no metabolic branchpoints, and if there are no other enzymes that catalyze the reaction, then such treatments should inhibit or eliminate the pathway. However, in reality, metabolic systems achieve robustness as a result of extensive redundancy[5], and there are often several isozymes that catalyze the same reaction. An inhibitor may still work, even if iszymes are

Athel Cornish-Bowden (athel@ibsm.cnrs-mrs.fr) is director of research and María Luz Cárdenas (cardenas@ibsm.cnrs-mrs.fr) is research associate in the laboratory Bioénergétique et Ingéniérie des Protéines of the Institut Fédératif Biologie Structurale et Microbiologie, CNRS, Marseilles, France (http://ir2lcb.cnrs-mrs.fr/~athel/homepage.htm).

present, because isozymes often share a similar structure and catalytic mechanism[6]. In contrast, gene deletion would affect one isozyme only, potentially producing little or no phenotypic effect. Alternatively, a bypass may allow flow of metabolites around the enzyme in question, but it is generally difficult to determine whether a bypass exists within a complex network. Metabolic pathway analysis aims to identify indispensable enzymes, thus predicting the effects of their deletion or inhibition.

Flux analysis should also allow the determination of the effects of adding new activities to a cell, and could thus be useful in engineering new pathways for the creation of valuable products. Knowledge of all the required reactions does not guarantee success, however, as the intended pathway may either consume reactants that cannot be replaced or release by-products that cannot be disposed of.

Many of the enzymes of *E. coli* are also present in *T. pallidum,* with some exceptions: no transaldolase has been identified, for example. Other than by analyzing the genome exhaustively, is there any way to deduce whether the absence of this enzyme reflects incomplete information or a fundamental metabolic difference between the two bacteria? This could be important, for example, if we want to design a strategy to destroy *T. pallidum* without affecting the viability of a physiologically beneficial organism. In fact, the absence of transaldolase from the list of known enzymes is almost certainly due to missing information and not to a real difference between *T. pallidum* and *E. coli,* and there is likely to be a transaldolase gene among the orphans. This follows from the fact that *T. pallidum* contains two distinct transketolases, and all of the elementary modes that involve transketolase steps also involve transaldolase; thus, unless a previously unknown elementary mode operates in *T. pallidum,* the two transketolases cannot be active unless a transaldolase is also present. It follows that pathway analysis can complement sequence analysis as a way of assigning functions to orphan sequences and of checking for inconsistencies in the results. Pathway analysis also identifies sets of enzymes that must necessarily operate together in fixed flux proportions in any steady state of the system[7]. These should correspond to sets of co-expressed genes, thus offering an independent check on other methods for identifying differential and coordinated gene expression[8], such as DNA array analysis of mRNA expression.

For these and other reasons, it is clear that we will be hearing more about elementary flux modes in the coming years, and that pathway analysis will complement comparative sequence analysis as an essential tool for making physiological sense of genome sequences. However, it should not be forgotten that there is a final stage in the progress toward a phenotypic description of an organism that is not dealt with by either system, and that is to take account of both kinetic and regulatory properties. Feedback mechanisms and other biological processes evolved for the good of the organism[9], not for the convenience of biotechnology, and may easily thwart biotechnological objectives if not adequately taken into account.

1. Claverie, J.-M. *Nature* **403**, 12 (2000).
2. Schuster, S., Fell, D. & Dandekar, T. *Nat. Biotechnol*. **18**,326–332 (2000).
3. Skolnick, J., Fetrow, J.S. & Kolinski, A. *Nat. Biotechnol.* **18**, 283–287(2000).
4. Claverie, J.-M., Lopez, F. & Abergel, C. in *Perspectives in structural biology* (ed. Vijayan, M.) 680–700 (The Universities Press, Hyderabad, India; 1999).
5. Bailey, J.E. *Nat. Biotechnol*. **17**, 616–618 (1999).
6. Cárdenas, M.L. "Glucokinase": its regulation and role in liver metabolism. (R.G. Landes, Austin, TX; 1995).
7. Schuster, S., Dandekar, T., Mauch, K., Reuss, M. & Fell, D. in *Technological and medical implications of metabolic control analysis* (ed. Cornish-Bowden, A. & Cárdenas, M.L.) 57–66 (Kluwer, Dordrecht; 2000).
8. Claverie, J.-M. *Hum. Molec. Genet*. **8**, 1821–1832 (1999).
9. Cornish-Bowden, A., Hofmeyr, J.-H.S. & Cárdenas, M.L. *Bioorg. Chem*. **23**, 439–449 (1995).

© 2000 Nature America Inc. • http://biotech.nature.com