



Systemic metabolic reactions are obtained by singular value decomposition of genome-scale stoichiometric matrices

Iman Famili, Bernhard O. Palsson*

Department of Bioengineering, University of California San Diego, 9500 Gilman Drive, Room 418, La Jolla, CA 92093-0412, USA

Received 21 December 2002; received in revised form 7 March 2003; accepted 25 March 2003

Abstract

Genome-scale metabolic networks can be reconstructed. The systemic biochemical properties of these networks can now be studied. Here, genome-scale reconstructed metabolic networks were analysed using singular value decomposition (SVD). All the individual biochemical conversions contained in a reconstructed metabolic network are described by a stoichiometric matrix (S). SVD of S led to the definition of the underlying modes that characterize the overall biochemical conversions that take place in a network and rank-ordered their importance. The modes were shown to correspond to systemic biochemical reactions and they could be used to identify the groups and clusters of individual biochemical reactions that drive them. Comparative analysis of the *Escherichia coli*, *Haemophilus influenzae*, and *Helicobacter pylori* genome-scale metabolic networks showed that the four dominant modes in all three networks correspond to: (1) the conversion of ATP to ADP, (2) redox metabolism of NADP, (3) proton-motive force, and (4) inorganic phosphate metabolism. The sets of individual metabolic reactions deriving these systemic conversions, however, differed among the three organisms. Thus, we can now define systemic metabolic reactions, or eigen-reactions, for the study of systems biology of metabolism and have a basis for comparing the overall properties of genome-specific metabolic networks.

© 2003 Elsevier Ltd. All rights reserved.

Keywords: Singular value decomposition; Stoichiometric matrix; Metabolic network

1. Introduction

Many metabolic reactions have been biochemically characterized and are found to have identical biochemical functions in different organisms. However, every biochemical reaction operates as a part of a larger metabolic network and thus plays a role in the systemic functions of an organism's metabolism. To examine the role of individual reactions in the context of a whole metabolic network, the integrated properties of these networks must therefore be analysed and a comparison of these properties amongst different organisms made. Structural or topological features of metabolic networks can be studied without the knowledge of kinetic properties of individual enzymes and several topological properties of networks have been established. The

connectivity of metabolites in cellular networks has been examined using “small world” models developed in graph theory (Fell and Wagner, 2000; Jeong et al., 2000; Ravasz et al., 2002), and metabolic pathways have been described and analysed using a variety of mathematical tools (Karp, 2001; Schilling and Palsson, 1998).

Topological studies of metabolic networks have led to insightful conclusions about their structural features and their overall organization. We now need to move beyond the study of topological properties and perform studies of the systemic biochemical conversion properties of reconstructed metabolic networks. In this report, we demonstrate the utility of singular value decomposition (SVD) for analysing the systemic biochemical conversion properties of genome-scale metabolic networks of three prokaryotes *Escherichia coli*, *Haemophilus influenzae*, and *Helicobacter pylori* (Covert et al., 2001; Gaasterland and Selkov, 1995; Overbeek et al., 2000; Selkov et al., 1998) by decomposing their stoichiometric matrices. SVD is an objective,

*Corresponding author. Tel.: +1-858-534-5668; fax: +1-858-534-5722.

E-mail address: bpalsson@be-research.ucsd.edu (B.O. Palsson).

non-parameteric, analytical method and thus, its application to metabolic networks of microbial genomes provides a basis for an unbiased and quantitative study of their systemic properties. Here, we show that SVD can be used to identify and characterize the dominant features of biochemical conversions in genome-scale prokaryotic metabolic networks.

2. Methods

2.1. Metabolic network construction

Genome-scale metabolic networks can be constructed using established procedures (Covert et al., 2001; Gaasterland and Selkov, 1995; Overbeek et al., 2000; Selkov et al., 1998). Metabolic network reconstruction is a three-step process using genomic, biochemical, and physiological data. The first step in the reconstruction process of a metabolic network is to gather a comprehensive list of annotated metabolic genes in the genome. Using biochemical information, the appropriate reactions are then assigned to the gene list and additional reactions are added based on biochemical evidence. Finally, physiological information about the organism is used to complete the metabolic network and add reactions to fill in the gaps in the existing pathways and/or to complete the metabolic map by including additional pathways.

2.2. Stoichiometric matrix of metabolic networks

The metabolic networks are mathematically represented in a matrix format with metabolites listed in the rows and biochemical reactions in the columns. The entries in each column corresponded to the stoichiometric coefficients of the substrates (negative numbers) and products (positive numbers) for each reaction. The resulting stoichiometric matrix, $\mathbf{S}_{m \times n}$ of m metabolites and n reactions, is thus a sparse matrix of generally integer elements in which the i th row defines the participation or connectivity of a particular metabolite across all metabolic reactions and the j th column provides the stoichiometry of all metabolites in that reaction. The dynamic mass balance of a metabolic system can then be described (Clarke, 1980; Horn and Jackson, 1972; Reich and Sel'kov, 1981) using the stoichiometric matrix which relates the flux rates of enzymatic reactions, $\mathbf{v}_{n \times 1}$, to time derivatives of metabolic concentrations, $\mathbf{x}_{m \times 1}$, as

$$\frac{d\mathbf{x}}{dt} = \mathbf{S}\mathbf{v}. \quad (1)$$

Although \mathbf{S} has no kinetic parameters, it does contain information about how the network structure affects the overall dynamics, as shown by Eq. (1). The structural

analysis of \mathbf{S} presented here, does not examine the dynamics of the system explicitly but instead provides topological information about the systemic relationships amongst metabolites and reactions. The steady-state reaction pathways (Schilling et al., 1999) and conserved concentration moieties (Schuster and Höfer, 1991) characterized by the null space and the left null space of \mathbf{S} , respectively, will not be discussed here.

2.3. Metabolic networks of *E. coli*, *H. influenzae*, and *H. pylori*

Genome-scale metabolic networks for *E. coli*, *H. influenzae*, and *H. pylori* have been constructed previously (Edwards and Palsson, 1999, 2000; Schilling, 2000). Their stoichiometric matrices contain 739, 461, 381 metabolic reactions, and 442, 367, 332 metabolites, respectively. For this study, the lumped reactions were decomposed into their individual steps to capture the true systemic properties of the network. For example, instead of representing the palmitic acid degradation with one lumped reaction, the degradation process was broken into eight steps. Biomass synthesis was not considered a network reaction and was not included in the analysis. This resulted in matrices of 772, 486, 405 reactions and 472, 389, 355 metabolites for *E. coli*, *H. influenzae*, and *H. pylori*, respectively.

2.4. Universal stoichiometric matrix

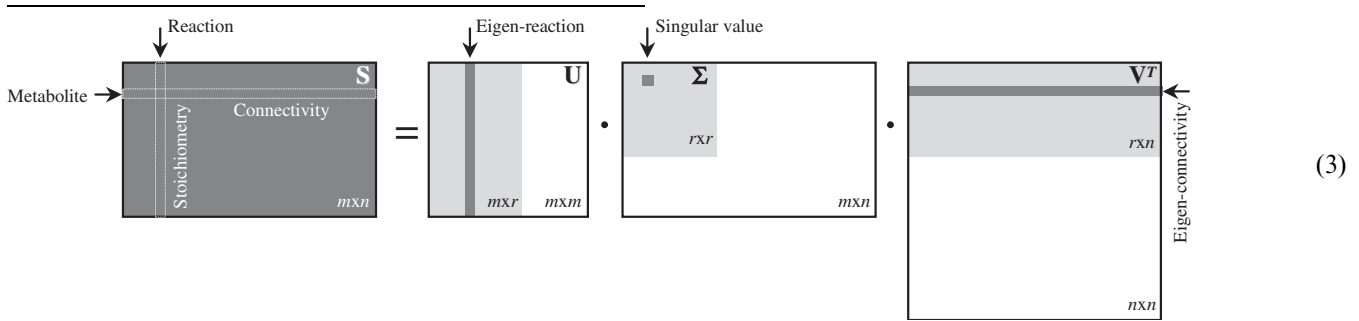
Many of the biochemical reactions are common amongst the three metabolic networks considered here. There also is a subset of reactions unique to each of the three organisms. In the stoichiometric matrices of *E. coli*, *H. influenzae*, and *H. pylori* approximately 20%, 8%, and 10% of the reactions, respectively, were found uniquely in one organism and not the other two. Thus, to unify the dimensionality of the three networks examined, a universal stoichiometric (US) matrix was constructed containing all metabolites and all metabolic reactions present in the three networks listed above. An organism specific post-multiplier matrix (i.e. a diagonal square matrix of 1's and 0's) was then formed to convert the US matrix into an organism-specific matrix (Edwards and Palsson, 1998). SVD was then performed on the resulting organism-specific matrices.

2.5. Control matrices

Two random matrices were generated for control purposes. First, a random matrix was generated by creating a matrix of integer elements with the same sparsity as the US matrix. Randomly chosen sets of columns were then selected from this matrix to form $\mathbf{R}_{m \times n}$. The resulting matrix therefore, represented a random non-biological stoichiometric matrix used as a

control. This random matrix was chosen to have the same dimension as the *H. pylori* network. Random

SVD of matrix \mathbf{S} can be shown schematically as



matrices of different dimensions generated similar characteristics (results are not shown). A second control matrix of the same size as the *H. pylori* network was also generated by randomly choosing a set of metabolic reactions from the US matrix to form a biologically meaningful random matrix.

3. Developing the conceptual framework

SVD is a well-established method used in a wide variety of applications including signal processing, noise reduction, image processing, kinematics, and recently mRNA expression analysis (Alter et al., 2000; Holter et al., 2001, 2000; Moon and Stirling, 2000; Pellegrino, 1993). The usefulness of SVD application for analysing biological data sets has been demonstrated both in analysing mRNA expression data and also predicting future expression levels (Alter et al., 2000; Holter et al., 2001, 2000).

SVD describes the underlying features of the mapping that a matrix represents. The stoichiometric matrix maps the reaction rate vector \mathbf{v} into the time derivatives of the metabolite concentration (Eq. (1)). Thus SVD of \mathbf{S} should lead to the identification of the underlying biochemical transformations of a metabolic reaction network and rank order their importance.

3.1. SVD of stoichiometric matrix

SVD states that for a given matrix $\mathbf{S} \in \mathbb{R}^{m \times n}$ of rank r , there are orthogonal matrices $\mathbf{U}_{m \times m}$ and $\mathbf{V}_{n \times n}$, and a diagonal matrix $\mathbf{\Sigma}_{r \times r} = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_r)$ with $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$ such that (Meyer, 2000):

$$\mathbf{S} = \mathbf{U} \begin{pmatrix} \mathbf{\Sigma} & 0 \\ 0 & 0 \end{pmatrix}_{m \times n} \mathbf{V}^T. \quad (2)$$

The columns of \mathbf{U} and \mathbf{V} are the left and right singular vectors of matrix \mathbf{S} , respectively, and represent its modes.¹ Combining Eq. (2) with the dynamic mass balance equation, Eq. (1), leads to:

$$\mathbf{U}^T \frac{d\mathbf{x}}{dt} = \mathbf{\Sigma} \mathbf{V}^T \mathbf{v} \quad \text{or} \quad \frac{d(\mathbf{u}_k^T \mathbf{x})}{dt} = \sigma_k (\mathbf{v}_k^T \mathbf{v}) \quad (\text{for } k = 1, \dots, r), \quad (4)$$

where \mathbf{u}_k and \mathbf{v}_k are the k th left and right singular vector of \mathbf{S} , respectively, and superscript T indicates the transpose operation. This simple derivation shows that there is a linear combination of metabolites:

$$\mathbf{u}_k^T \cdot \mathbf{x} = u_{k1}x_1 + u_{k2}x_2 + \dots + u_{kr}x_m \quad (5)$$

that is being uniquely moved by a linear combination of metabolic fluxes as

$$\mathbf{v}_k^T \cdot \mathbf{v} = v_{k1}v_1 + v_{k2}v_2 + \dots + v_{kn}v_n \quad (6)$$

and the extent of this motion is given by σ_k .

3.2. Definition of modes as “systemic reactions”

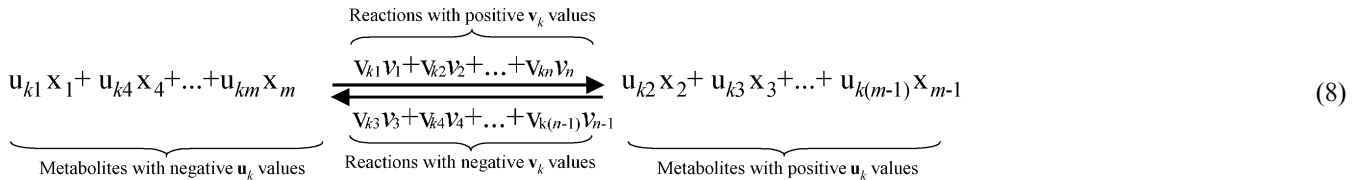
An important feature of SVD is that the singular vectors are orthonormal to each other and consequently, each one of the k th motions in Eq. (4) is decoupled from the others. Thus examining the quantities defined by Eqs. (5) and (6) should delineate these underlying independent factors. Eq. (5) gives a linear combination of concentrations that can be interpreted as a systemic reaction that represents a decoupled process and one that is driven by the linear combination of the reactions defined by Eq. (6). This relationship, therefore, defines

¹Terms “singular vector” and “mode” have been used interchangeably in this manuscript.

an “eigen-reaction” or a “systemic” metabolic reaction as

$$\sum_{\text{for } u_{ki} < 0} u_{ki} X_i \xleftrightarrow{\sum v_{kj} v_j} \sum_{\text{for } u_{ki} > 0} u_{ki} X_i \quad (7)$$

that for example may take the form



where the elements of u_k are equivalent to “systemic” stoichiometric coefficients and the elements of v_k are “systemic” reaction participations. These systemic metabolic reactions can be used to describe the function of the metabolic network as a whole and are useful concepts in studying the systems biology of metabolism. A similar approach has been taken for the analysis of expression arrays to define “eigen-genes” and “eigen-arrays” (Alter et al., 2000).

3.3. Illustrative example

These general concepts can be illustrated via a simple example (Fig. 1). The two reactions, v_1 and v_2 , in Fig. 1A

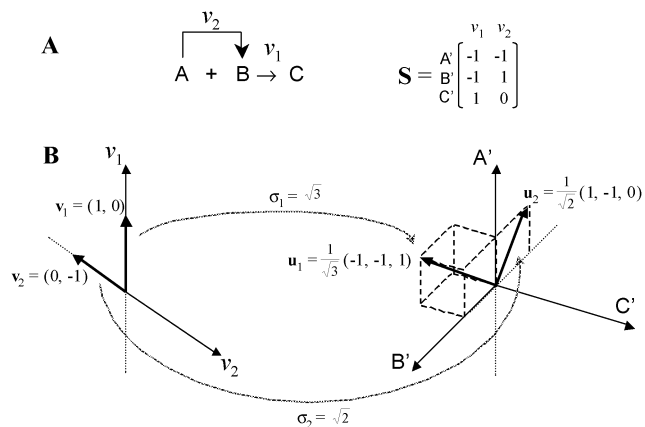


Fig. 1. Simple example of SVD analysis for reacting systems. (A) Schematic of a reaction network with two reactions (v_1 and v_2) that act on metabolites A, B and C, and its corresponding stoichiometric matrix, S. (B) Graphical representation of the mapping between the eigen-connectivities and eigen-reactions. The two singular vectors that correspond to v_1 and v_2 are orthogonal and catalyse each eigen-reaction, u_1 and u_2 , independently. The two singular vectors of time derivatives are orthogonal in a three-dimensional space. The magnitude of the singular values, σ_1 and σ_2 , indicates the relative contribution of its corresponding singular vectors to the overall construct of the biochemical transformation in the network (A' , time derivative of A').

form a two-dimensional flux space and relate metabolites A, B, and C in the space where the time derivatives of concentrations lie. The two singular vectors, v_1 and v_2 , are orthogonal and drive the singular vectors of time derivatives, u_1 and u_2 , independently (Fig. 1B). The singular values in this simple example are $\sigma_1 = 1.73$ and

$\sigma_2 = 1.41$ and indicate the relative contribution of each singular vector to the overall construct of the biochemical motion in the network.

4. Results

Stoichiometric matrices of metabolic networks for three prokaryotes *E. coli*, *H. influenzae*, and *H. pylori* were analysed using SVD. The singular value spectrum of genome-scale stoichiometric matrices is triphasic (Fig. 2A). The first four dominant singular values in the first phase capture about 27% of the networks’ characteristics, as shown by the cumulative singular fractions (Fig. 2B). This phase is found in the spectra of all three genome-scale networks examined. A randomly generated network (see Methods), R_{mtx} , which represents a non-biological network has, in contrast, a more uniform spectrum and does not contain dominant modes. Phase one is followed by a gradual steady decline in a large number of singular values. The second phase constitutes a large portion of the remaining network characteristics (Fig. 2B). Finally, the last phase of the singular value spectra is comprised of a rapidly declining small set of singular values that is absent in the random network. The second randomly generated matrix (see Methods) exhibited the same triphasic shape as the genome-scale matrices, indicating that biochemical reactions exhibit dominant characteristics (Fig. 2B).

If the modes of the stoichiometric matrix from one metabolic network resemble that of another, then the networks share some overall characteristics. The angle between the singular vectors was used as a measure of similarity among the dominant singular vectors of the three genome-scale metabolic networks (Fig. 3). The dominant eigen-reactions were found to be similar amongst the leading modes of the three networks (Fig. 3A). This similarity implies that there are overall

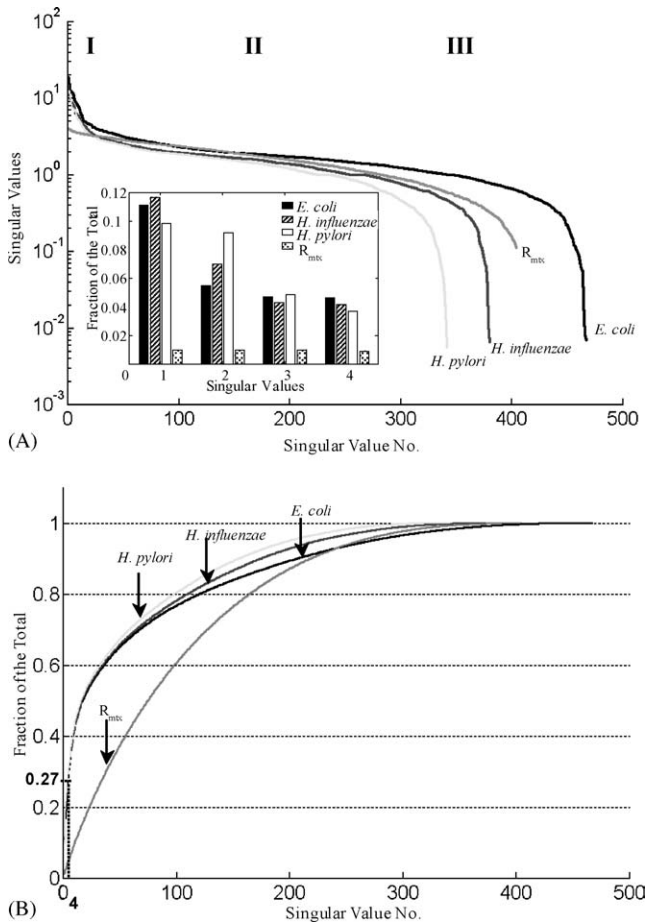


Fig. 2. Singular values spectra of the genome-scale stoichiometric matrices. (A) The singular values of three metabolic networks and a non-biological random set. Unlike for experimental data sets, no statistical preprocessing steps were required for the analysis. The second randomly generated matrix (see Methods) exhibited the same triphasic shape as the genome-scale matrices with a shorter phase III than the genome-scale matrices (results are not shown), indicating that biochemical reactions exhibit dominant characteristics and metabolic networks contain distinct properties at the higher modes. The fractional contributions of the first four singular values are depicted in the inset. None of the dominant modes contributes more than about 12% of the total characteristics of the stoichiometric matrix. (B) The cumulative fractional contributions of singular values (i.e. the sum of fractional contributions up to a singular value number). The fractional contributions in \mathbf{R}_{mtx} add more uniformly to the overall expression and are distinctly different from the other sets (\mathbf{R} , random; \mathbf{mtx} , matrix).

transformations (as shown in Eq. (4)) that are similar in all three networks. The corresponding eigen-connectivities did not exhibit such strong similarities (Fig. 3B). This result implies that although the same overall biochemical transformations are common among all three networks, the combination of metabolic reactions driving them can be different (see Fig. 4). The modes become less similar further down the singular value spectrum. In most cases the closest angles between modes of two genomes are more than 70° (data not shown), with 90° being completely dissimilar and zero degrees being identical.

Several key metabolites appear in the dominant modes. The cofactors participating in energy, redox, and phosphate metabolism emerge with the most significant values in the first four eigen-reactions of all genome-scale networks (Fig. 4A). All three genomes thus share dominant characteristics as described by the first four modes. The dominant eigen-reaction in all the genomes is the conversion of ATP to ADP and P_i . The second eigen-reaction describes the reduction of NADP to NADPH. The third eigen-reaction describes the translocation of a proton and its use in driving various reactions. The fourth mode describes the incorporation and release of inorganic phosphate. In all three networks the dominant biochemical transformations are rank ordered the same way. The overall dominant features of these three genome-scale networks thus represent similar but not identical chemical transformations.

Although the eigen-reactions are similar in the three organisms, the metabolic reactions participating in driving these conversions differ from one network to another (Fig. 4B). This dissimilarity shows that the three organisms accomplish the shared overall biochemical transformations in organism-specific manner. The reaction participation in the four principal conversions is as follows.

1. The ATP-coupled transporters have the highest reaction participations of the first singular vectors of \mathbf{V} , except for *E. coli* where a group of synthetase reactions is present (Fig. 4B). A group of kinases with identical reaction participations follow. These three types of reactions thus define the dominant mode in all three genomes.
2. The second mode corresponds to redox conversions that involve NADPH. In all three genomes fatty acid synthesis reactions have dominant reaction participations in this mode. Then, a group of reductases appears with five to ten-fold lower reaction participations. These two types of redox exchanges dominate NADPH metabolism.
3. The third mode in all genomes corresponds to a translocated proton and thus the proton-motive force. The primary participants in this conversion are the electron transport system (ETS) reactions in all genomes. The secondary participants are sets of proton-coupled transporters.
4. The fourth mode in all three genomes represents transformations involving inorganic phosphate. Unlike the first three modes, there is little overlap amongst the reactions that participate in this transformation in the three genomes studied.

Thus, the dominant modes correspond to important generic biochemical transformations of metabolites in all these organisms but the sets of metabolic reactions that participate in these transformations may differ

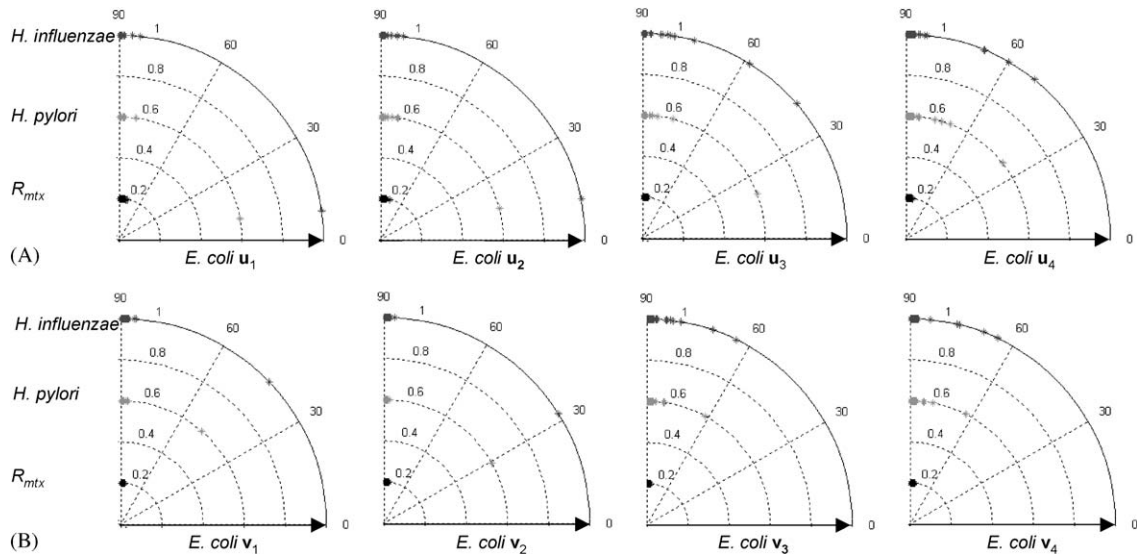


Fig. 3. Angle as a similarity measure between singular vectors from different organisms. The similarity between the singular vectors was evaluated by determining the angle they form with respect to each other. More similar vectors thus form smaller angles than two dissimilar vectors. (A) The first four eigen-reactions of *E. coli* are compared to all eigen-reactions of the other three networks using the angle between u_k 's as a similarity measure. For the first three modes, the closest eigen-reaction is the same in all genomes, e.g. u_1 in *E. coli* is the closest to u_1 in *H. influenzae*. (B) The first four eigen-connectivities of *E. coli* are compared to all eigen-connectivities of the other three networks. As for the eigen-reactions, the first three eigen-connectivities of the genomes are the same, e.g. v_1 in *E. coli* is the most similar vector to v_1 in *H. pylori*. The further one goes down the spectrum (i.e. from v_1 to v_4), the more dissimilar the singular vectors of the three genomes become. The approximate orthogonality of the randomly generated set, R_{mtx} , to all organism-generated singular vectors is consistent with the observations made in Fig. 2. The radial distance on the polar coordinate was used for illustration purposes to separate the networks.

significantly in a number of modes or be similar in others.

Metabolites and biochemical reactions can participate in more than one mode. Within a genome, there are very few cases where a metabolite is a major participant in more than one of the dominant eigen-reactions. For example in *E. coli*, the only significant exception is ATP, ADP, and P_i involvement in modes 1 and 4 (Fig. 5A). Thus, the dominant eigen-reactions are reasonably “decoupled” in this regard. As Fig. 4 suggests, there is a similarity in metabolite and reaction participations in the systemic modes amongst the three organisms (Fig. 5B). The coefficients in two corresponding eigen-reactions in two different organisms can be plotted against one another to show this relationship. The reactions that are shared between corresponding modes in two organisms appear in the quadrants and those that do not appear on the axes. Thus, there are reactions that play similar systemic roles in the metabolism of two different organisms and there are several that are well decoupled.

5. Discussion

Complete genome-sequences have led to the reconstruction of genome-scale metabolic networks. The corresponding stoichiometric matrices that represent

the entire set of biochemical transformations in the network can be studied and interpreted using SVD. SVD of S leads to two principal results. First an elegant interpretation of the dominant systemic metabolic reactions is obtained, where systemic chemical conversions are defined as eigen-reactions with systemically defined stoichiometric coefficients and reaction participations, u_{ik} and v_{jk} , respectively, for the k th mode (Eq. (7)). Second, these systemic reactions can be used to interpret the systems biology of the genome-scale networks. The dominant modes in all three organisms examined are comprised of classes of similar biochemical conversions but they are carried out in organism-specific manner through the use of different set of reactions. Enzyme substrate promiscuity and isozymes may thus play an important role in developing some of the overall features of a metabolic network and play a role in adaptive evolution. Such plasticities in the *E. coli* metabolic network have been described (Ouzounis and Karp, 2000).

Topological features of metabolic networks have been productively studied (Fell and Wagner, 2000; Jeong et al., 2000; Ravasz et al., 2002). SVD of S gives results that lie beyond topological features of metabolic networks and accounts for the biochemical nature of the chemical transformations, not just the links in the network. The results obtained thus have dynamic and biochemical implications. Eigen-metabolites and

eigen-reactions that are associated by non-zero singular values correspond to dynamic features of the metabolic networks where changes in reaction activities cause changes in metabolite concentrations. These in turn correspond to a relationship between the row and column space of S that hereto have not been systematically studied. In contrast, the modes with zero singular values define the null space and left null space of a metabolic network, corresponding to the

steady-state flux maps (Schilling et al., 1999) and conserved concentration moieties (Schuster and Höfer, 1991), respectively, have been well studied. SVD of S thus gives us a network-based analysis of the overall chemical transformation and dynamic properties of a metabolic network without explicit inclusion of kinetic information.

The stoichiometric matrix only contains information about the topological or structural features of the

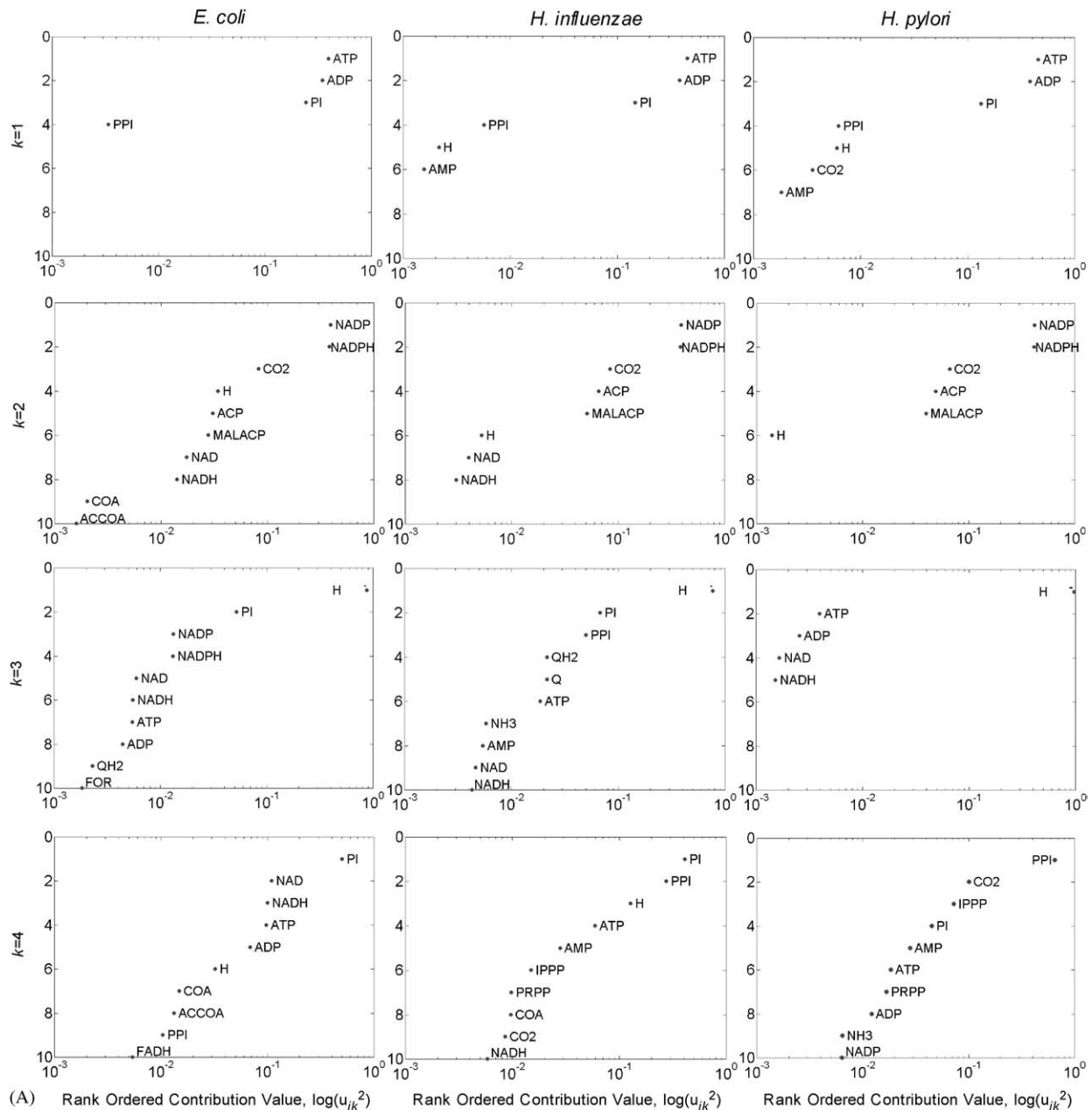


Fig. 4. Eigen-reactions and eigen-connectivities in *E. coli*, *H. influenzae*, and *H. pylori*. (A) The four dominant eigen-reactions in the three organisms. The coefficients given in the figure are the u_{ik} ($i = 1, \dots, m$ and $k = 1, \dots, 4$) shown in Eq. (5) and the corresponding metabolite x_i is indicated. As discussed in the text, the four dominant modes correspond to the conversion of ATP to ADP and P_i , NADPH to NADP conversion, proton-motive force and inorganic phosphate metabolism. (B) The eigen-connectivities for the reactions participating in each eigen-reaction. The coefficients shown correspond to v_{jk} ($j = 1, \dots, n$ and $k = 1, \dots, 4$) in Eq. (6) giving the participation of the reaction indicated in the conversion. The grouping of reaction shown is discussed in the text.

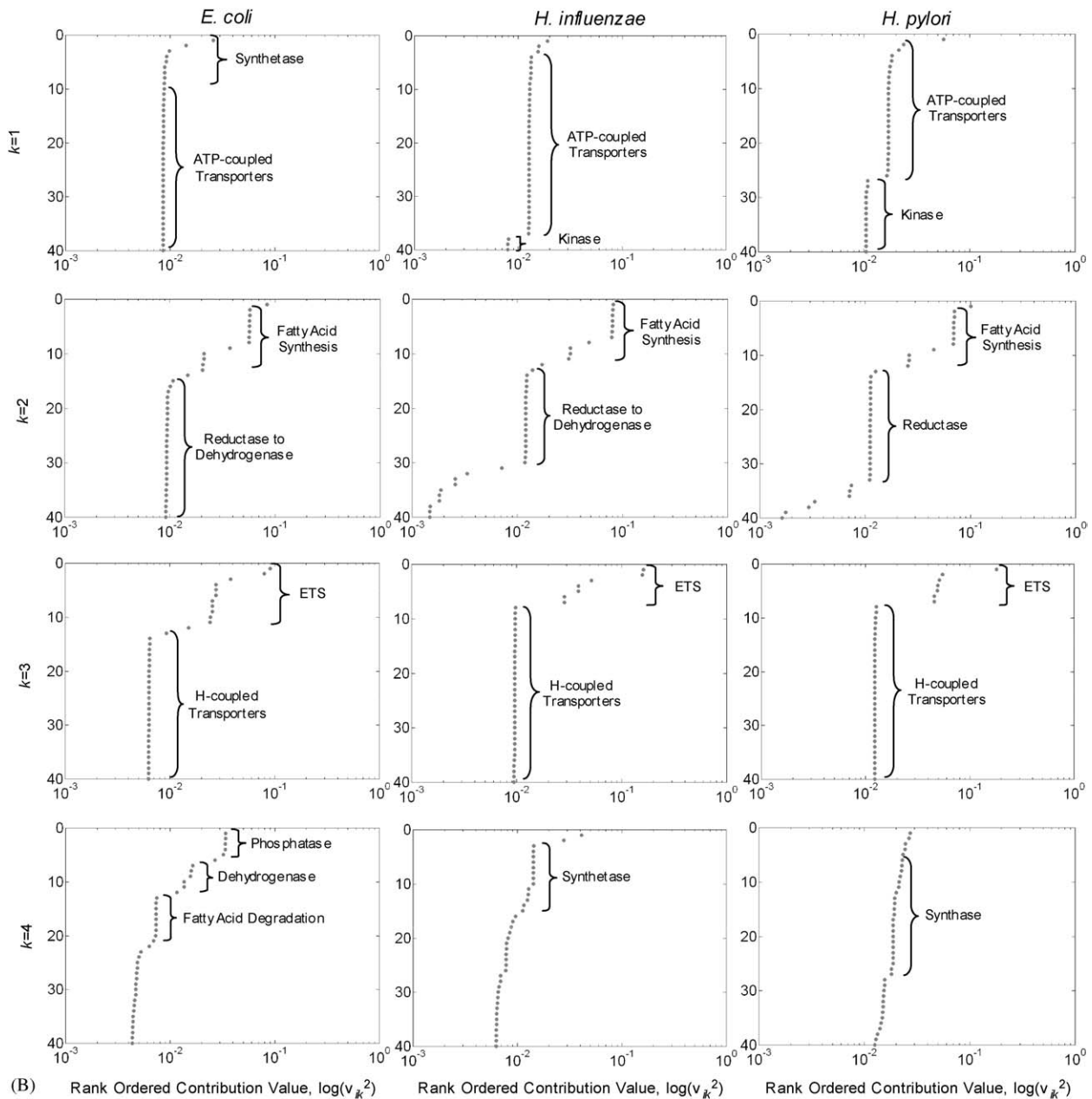


Fig. 4 (continued).

network. No kinetic and thermodynamic information is represented. The systemic metabolic reactions derived from SVD may therefore represent a linear combination of reactions that drive that mode in opposite directions. Since the singular vectors do not have a unique direction, thermodynamic considerations may lead to partitioning of the right singular vectors into two vectors that point into exactly opposite directions. Based on the interesting results obtained here on network properties, a further study of the incorporation of thermodynamic constraints in SVD of S is clearly warranted.

SVD analysis has been used as a clustering method in biological data analysis including mRNA gene

expression in the recent years (Alter et al., 2000; Holter et al., 2001, 2000). However, the application of SVD for studying the stoichiometric matrices differs fundamentally from its common use in data analysis. Unlike a data matrix, the entries in stoichiometric matrices are noise free and thus all the modes which contribute to the network structure can be considered. Thus, the higher modes which contain less prominent features of the network may convey meaningful biological information. A subsequent and more detailed study may elucidate important “small” biochemical differences between organisms that may have biological significance.

Taken together, the commonly used SVD method gives great insight into the biological properties of

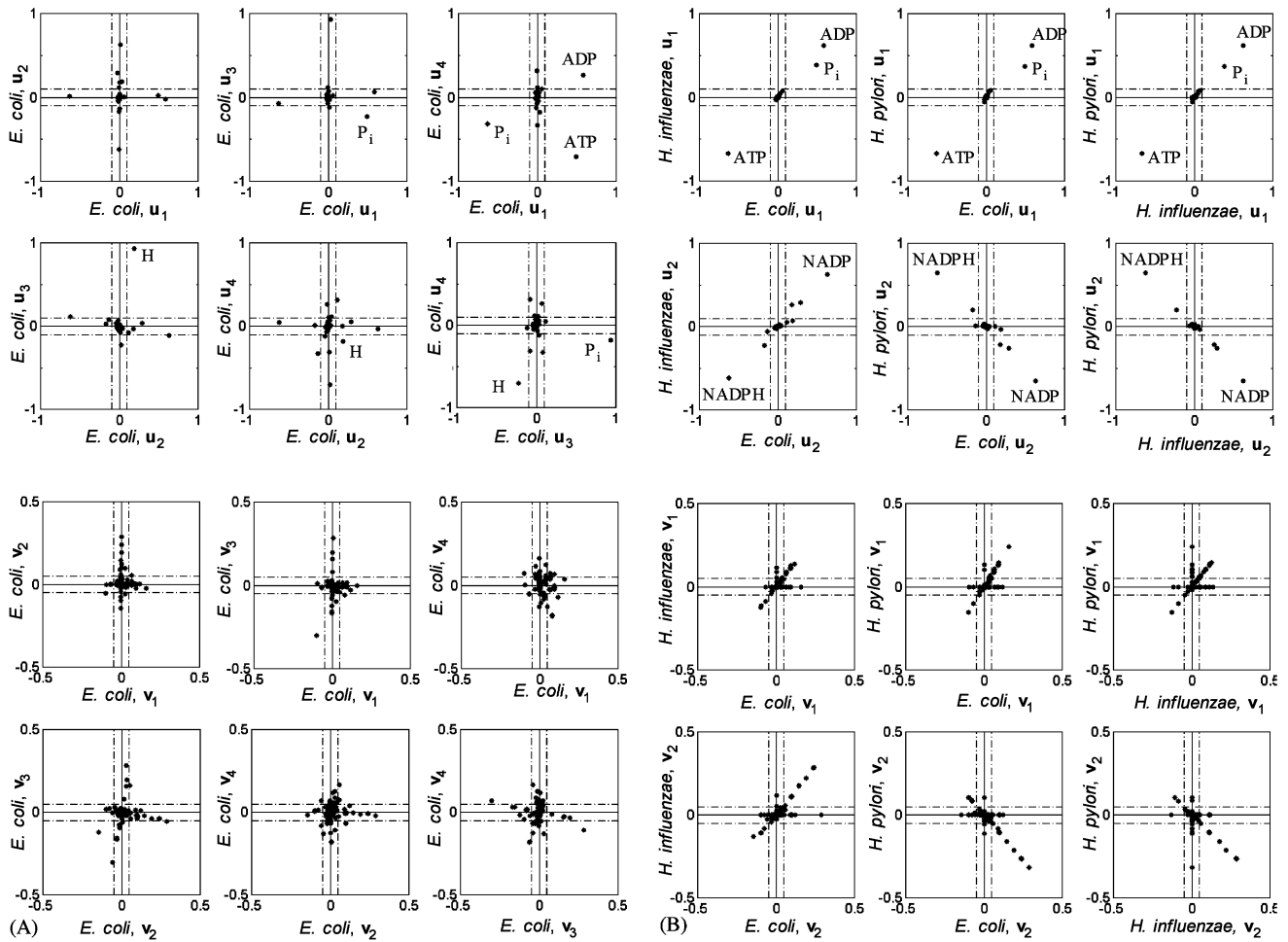


Fig. 5. Singular vector participation amongst genome-scale networks. (A) The first four eigen-reactions and eigen-connectivities for *E. coli* are shown. Dashed lines demarcate 10% of the full scale shown in the diagram. Values outside the 10% region exhibit significant contributions. The panels above the diagonal correspond to u_k ($k = 1, \dots, 4$) and below the diagonal to v_k ($k = 1, \dots, 4$). Above the diagonal, each individual panel graphs two eigen-reactions to show the relative magnitude of the systemic stoichiometric coefficients, u_{ik} , for the same metabolite in two eigen-reactions. Similarly, the reaction participations for individual metabolic reactions in two eigen-reactions are compared below the diagonal line. (B) Dominant pairs of eigen-reactions and eigen-connectivities amongst *E. coli*, *H. influenzae*, and *H. pylori*. The first two eigen-reactions and eigen-connectivities for the three genomes are compared in pairs. The metabolites and reactions that fall on the diagonal are used similarly in the modes of two networks. Note that the singular vectors determined by SVD are only unique up to the sign (\pm). Thus, appearance in the first and third quadrants and in the second and fourth quadrants in these correlation plots is identical. The points that fall on the individual axes correspond to reactions and metabolites that are only found in one of the modes being compared.

genome-scale networks and provide a way to perform a comparative study of network characteristics. Perhaps most importantly, we can now identify systemic metabolic reactions for the study of systems biology of metabolism.

Acknowledgements

We thank the Whitaker Foundation for their support through the Graduate Fellowships in Biomedical Engineering to I.F., the National Science Foundation through grant nos. MCB 9873384, and BES 9814092 and 0120363, and the National Institutes of Health

through grant no. GM57089. We also thank Jennifer L. Reed with help in constructing the universal stoichiometric matrix and David Haussler, Kenneth Kreutz-Delgado and Nagiza Samatova for valuable discussions.

References

- Alter, O., Brown, P.O., Botstein, D., 2000. Singular value decomposition for genome-wide expression data processing and modeling. *Proc. Natl Acad. Sci. USA* 97 (18), 10101–10106.
- Clarke, B.L., 1980. Stability of complex reaction networks. In: Prigogine, I., Rice, S.A. (Eds.), *Advances in Chemical Physics*. John Wiley, New York, pp. 1–215.

- Covert, M.W., Schilling, C.H., Famili, I., Edwards, J.S., Goryanin, I.I., Selkov, E., Palsson, B.O., 2001. Metabolic modeling of microbial strains *in silico*. Trends Biochem. Sci. 26, 179–186.
- Edwards, J.S., Palsson, B.O., 1998. How will bioinformatics influence metabolic engineering? Biotechnol. Bioeng. 58 (2-3), 162–169.
- Edwards, J.S., Palsson, B.O., 1999. Systems properties of the haemophilus influenzae Rd metabolic genotype. J. Biol. Chem. 274 (25), 17410–17416.
- Edwards, J.S., Palsson, B.O., 2000. The *E. coli* MG1655 *in silico* metabolic genotype: its definition, characteristics, and capabilities. Proc. Natl Acad. Sci. 97 (10), 5528–5533.
- Fell, D.A., Wagner, A., 2000. Structural properties of metabolic networks: implications for evolution and modelling of metabolism. In: Hofmeyr, J.-H.S., Rohwer, J.M., Snoep, J.L. (Eds.), Animating the Cellular Map. Stellenbosch University Press, Stellenbosch, pp. 79–85.
- Gaasterland, T., Selkov, E., 1995. Reconstruction of metabolic networks using incomplete information. Intel. Sys. for Mol. Biol. 3 (1), 127–135.
- Holter, N.S., Mitra, M., Maritan, A., Cieplak, M., Banavar, J.R., Fedoroff, N.V.F., 2000. Fundamental patterns underlying gene expression profiles: simplicity from complexity. Proc. Natl Acad. Sci. USA 97 (15), 8409–8414.
- Holter, N.S., Maritan, A., Cieplak, M., Fedoroff, N.V., Banavar, J.R., 2001. Dynamic modeling of gene expression data. Proc. Natl Acad. Sci. USA 98 (4), 1693–1698.
- Horn, F., Jackson, R., 1972. General mass action kinetics. Arch. Ration. Mech. Anal. 47, 81–116.
- Jeong, H., Tombor, B., Albert, R., Oltvai, Z.N., Barabási, A.L., 2000. The large-scale organization of metabolic networks. Nature 407 (6804), 651–654.
- Karp, P.D., 2001. Pathway databases: a case study in computational symbolic theories. Science 293 (5537), 2040–2044.
- Meyer, C.D., 2000. Matrix Analysis and Applied Linear Algebra. Society for Industrial and Applied Mathematics, Philadelphia.
- Moon, T.K., Stirling, W.C., 2000. Mathematical Methods and Algorithms for Signal Processing. Prentice-Hall, Upper Saddle River, NJ.
- Ouzounis, C.A., Karp, P.D., 2000. Global properties of the metabolic map of *E. coli*. Genome Res. 10 (4), 568–576.
- Overbeek, R., Larsen, N., Pusch, G.D., D'Souza, M., Selkov, J.E., Kyrpides, N., Fonstein, M., Maltsev, N., Selkov, E., 2000. WIT: integrated system for high-throughput genome sequence analysis and metabolic reconstruction. Nucl. Acids Res. 28 (1), 123–125.
- Pellegrino, S., 1993. Structural computations with the singular value decomposition of the equilibrium matrix. Int. J. Solids Struct. 30 (21), 3025–3035.
- Ravasz, E., Somera, A.L., Mongru, D.A., Oltvai, Z.N., Barabasi, A.L., 2002. Hierarchical organization of modularity in metabolic networks. Science 297 (5586), 1551–1555.
- Reich, J.G., Sel'kov, E.E., 1981. Energy Metabolism of the Cell: A Theoretical Treatise. Academic Press, London, New York.
- Schilling, C.H., 2000. On systems biology and the pathway analysis of metabolic networks. Ph.D., University of California, San Diego.
- Schilling, C.H., Palsson, B.O., 1998. The underlying pathway structure of biochemical reaction networks. Proc. Natl Acad. Sci. USA 95 (8), 4193–4198.
- Schilling, C.H., Schuster, S., Palsson, B.O., Heinrich, R., 1999. Metabolic pathway analysis: basic concepts and scientific applications in the post-genomic era. Biotechnol. Prog. 15 (3), 296–303.
- Schuster, S., Höfer, T., 1991. Determining all extreme semi-positive conservation relations in chemical reaction networks: a test criterion for conservativity. J. Chem. Soc. Faraday Trans. 87, 2561–2566.
- Selkov Jr., E., Grechkin, Y., Mikhailova, N., Selkov, E., 1998. MPW: the metabolic pathways database. Nucl. Acids Res. 26 (1), 43–45.