9 The Stoichiometric Matrix

Mathematics is the door and key to the sciences – Roger Bacon

The reactions that comprise a biological network can be represented by chemical equations. The stoichiometric matrix is formed from these chemical equations. It has several important attributes. In this chapter we focus on four principal views of the stoichiometric matrix and its content: (i) it is a data matrix, (ii) it is a connectivity matrix, (iii) it is a mathematical mapping operation, and (iv) it is a central part of *in silico* models used to compute steady and dynamic network states. These features are summarized in Figure 9.1.

9.1 The Many Attributes of S

The stoichiometric matrix is formed by the stoichiometric coefficients of the reactions that constitute a reaction network. It is organized such that every column corresponds to a reaction and every row corresponds to a compound. The entries in the matrix are stoichiometric coefficients that are integers. Each column that describes a reaction is constrained by the rules of chemistry, such as elemental balancing. Every row thus describes all the reactions in which the corresponding compound participates, and therefore how the reactions are interconnected. This deceptively simple matrix has many noteworthy attributes that are summarized in Table 9.1.

Informatic attributes. The stoichiometric matrix is a data matrix. The data that go into building a genome-scale stoichiometric matrix come primarily from the annotated genomic sequence and detailed assessment of the literature (bibliomic data) that is available about the target organism. Often, inferences from phylogenetics are used as well. All this information is the basis for the reconstruction process described in Part I.

Physical/chemical attributes. The stoichiometric coefficients represent counts of molecules that are involved in a chemical reaction. Chemical reactions come with conservation relationships of elements, charge, and other properties. These properties must be represented accurately. The cellular location of a reaction is included through the assignment of a metabolite to a cellular compartment.

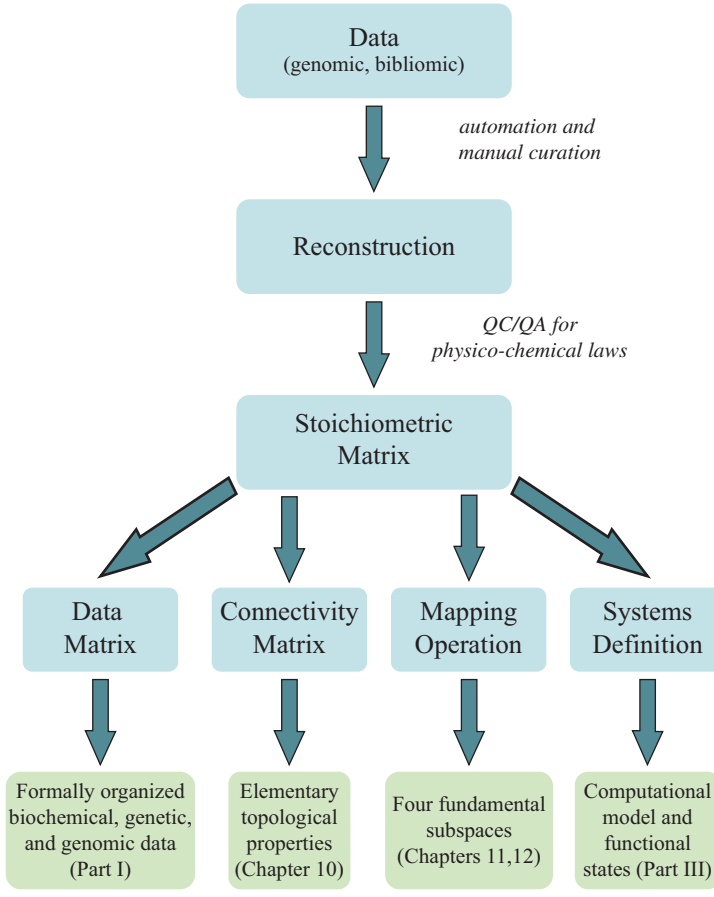


Figure 9.1 Four main features of the stoichiometric matrix and the corresponding portions of this textbook.

Genetic/genomic attributes. A genome-scale network reconstruction effectively represents a two-dimensional annotation of a genome [313]. It contains information not only about the components of a cell, but also about how they are connected. **S** is a species property; all members of a species share this matrix. In principle, there is only one network encoded by the genome of a particular organism.

Biological attributes. Because the stoichiometric matrix is specific to a species, it can be used to study the phenotypic differences and capabilities of various species. As evolution can occur by changing the number of reactions encoded in a genome (e.g., with horizontal gene transfer or gene deletion), the stoichiometric matrix can be used to study distal causation.

Mathematical attributes. The stoichiometric matrix consists of integers. If the reactions represented are elementary, their numerical values are (–2, –1, 0, 1, 2). It therefore has no error associated with it and is thus a *knowable* matrix. In a similar way, a genomic sequence is knowable because each position can only be occupied by one of four bases.

Table 9.1 The many attributes of the stoichiometric matrix.

Attribute	Represents
Informatic	Annotated genome Bibliomic data Comparative genomics
Physico-chemical	Chemistry, cellular location conservations
Genetic	Genomic characteristics Represents a species
Biological	Species differences Distal causation
Mathematical	Integer entries 'Knowable' matrix
Systemic	Pool formation network structure
Numerical	Integers, sparse well-conditioned

Systemic attributes. The stoichiometric matrix is a connectivity matrix and gives the structure (or the topology) of a network. This structural information results in the definition of pools and pathways that are associated with the null spaces of the stoichiometric matrix.

Numerical attributes. For large networks, the stoichiometric matrix has mostly zero elements. It is a sparse matrix and may require sparse matrix representation and computational procedures as it reaches the genome-scale. All the elements of the matrix are of the same order of magnitude, making it a numerically well-conditioned matrix.

Thus, there are many different attributes to the stoichiometric matrix. All are addressed and studied in this book.

9.2 Chemistry: S as a Data Matrix

Every column in **S** represents a chemical reaction. It must be consistent with the chemistry that it represents. For instance, every reaction has to be elementally and charge-balanced. In addition, from a data-mapping standpoint, it is useful to have unique chemical specifiers associated with the compounds that the rows represent, and it can be useful to have EC numbers associated with the columns of the matrix to classify the type of chemical transformation that a column represents. This section illustrates the incorporation and representation of chemical properties in the formulation of **S**.

9.2.1 Elementary biochemical reactions

There is a limited number of elementary types of biochemical reactions that take place in cells. These fall into the three categories below. In the examples of each, derived for metabolic transformations, we used C to denote a primary metabolite, P as a phosphate group, and A as a co-factor such as the adenosine moiety in AMP, ADP, and ATP.

Reversible conversion Transformation between two compounds consisting of the same two chemical moieties C and P can be written as

$$CP \rightleftharpoons PC \tag{9.1}$$

representing two elementary reactions (forward and reverse). Although such reversible conversions are often used to generically describe reactions, they can only represent a simple chemical rearrangement of the molecule without any change in its elemental composition. For instance, isomerases catalyze such reactions. The stoichiometric matrix that describes this reaction is

$$\mathbf{S} = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \begin{matrix} : CP \\ : PC \end{matrix} \tag{9.2}$$

where the first column of the matrix represents the forward reaction and the second column the reverse reaction. These are elementary reactions that have non-negative flux values. The first row represents CP and the second row PC. Concentrations are also non-negative quantities. Under certain circumstances, one may wish to combine the two elementary reactions into a net reaction that can take on positive or negative values.

Bi-molecular association Many biochemical reactions involve the combination of two moieties, C and P, to form a new compound:

$$C + P \rightleftharpoons CP \tag{9.3}$$

Sometimes, such reactions may not involve forming and breaking of covalent bonds, but a series of hydrogen bonds to form a complex, such as the dimerization of two protein molecules, or the initial binding of a substrate to an active site on an enzyme molecule. The stoichiometric matrix that describes a bi-molecular association is

$$\mathbf{S} = \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix} \begin{matrix} : C \\ : P \\ : CP \end{matrix} \tag{9.4}$$

where the rows represent C, P, and CP, respectively, and the single column represents the net reaction.

A co-factor-coupled reaction A frequent reaction in biochemical reaction networks is one in which one compound (AP) donates a moiety (P) to another compound (C):

$$C + AP \rightleftharpoons CP + A \tag{9.5}$$

In reality, such reactions have an intermediate, and can be decomposed into two bi-molecular association reactions. Strictly speaking, this is not an elementary reaction. The stoichiometric matrix that describes the co-factor-coupled (or moiety exchange) reaction is

$$\mathbf{S} = \begin{pmatrix} -1 \\ -1 \\ 1 \\ 1 \end{pmatrix} \begin{matrix} : C \\ : AP \\ : CP \\ : A \end{matrix} \tag{9.6}$$

where the rows represent C, AP, CP, and A, respectively, and the column represents the net reaction. The word 'co-factor' is used synonymously with 'carrier.'

Combining stoichiometric matrices A stoichiometric description of multiple reactions is formed easily by combining the individual matrices. The three types of transformations can be combined into one matrix (Equation (9.7)). Note that the net reaction rate has been used in forming **S** in this equation.

$$\begin{matrix} v_1 & v_2 & v_3 & \\ \begin{pmatrix} 0 & 0 & 1 \\ 0 & -1 & -1 \\ 0 & -1 & 0 \\ -1 & 1 & 1 \\ 0 & 0 & -1 \\ 1 & 0 & 0 \end{pmatrix} & \begin{matrix} A \\ C \\ P \\ CP \\ AP \\ PC \end{matrix} \end{matrix} \tag{9.7}$$

It should be clear to the reader that stoichiometric matrices for two separately reconstructed networks that share compounds can be integrated easily. This easy integration makes the stoichiometric representation of networks highly scalable.

9.2.2 Basic chemistry

The chemical reactions that form the columns in **S** have the basic rules of chemical transformation associated with them. There are conservation quantities (such as elements and charge) and there are non-conserved quantities (such as osmotic pressure and free energy) associated with chemical transformations. These properties must be accounted for in the construction of a biochemically meaningful stoichiometric matrix.

The elemental matrix Metabolites consist of six chemical elements. The elemental matrix, **E**, gives the composition of all the compounds considered in a network. A column of **E** corresponds to a compound, $\mathbf{x}_i$, and the rows correspond to the elements, typically only six of them found in organic compounds: carbon, oxygen, nitrogen, hydrogen, phosphorous, and sulfur. Compounds can be represented as points in a space formed by the elements as the axes (Figure 9.2).

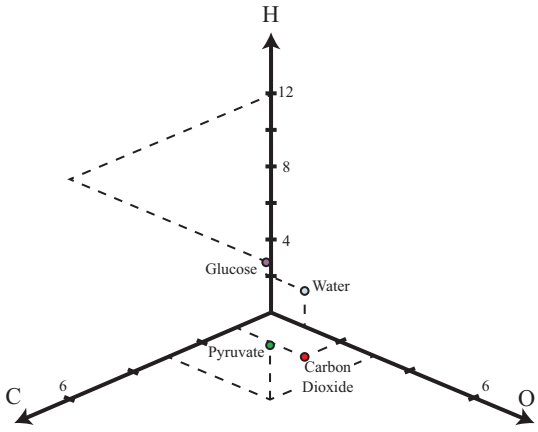


Figure 9.2 The elemental space. Representation of compounds containing carbon, hydrogen, and oxygen in a three-dimensional space. The coordinates are: glucose (6,12,6), pyruvate (3,3,3), water (0,2,1), and carbon dioxide (1,0,2).

It is important to note that the elemental composition of a molecule does not uniquely specify its chemical structure. For instance, glucose and fructose have the same elemental composition. Thus, associating unique chemical identifiers with the columns of **E** is desirable. The elemental composition of common metabolites is shown in Table 9.2.

Example: Consider the simple chemical reaction

$$2H_2 + O_2 \rightarrow 2H_2O \tag{9.8}$$

that involves only two elements, oxygen and hydrogen. The elemental matrix for this chemical reaction is

$$\mathbf{E} = \begin{pmatrix} 0 & 2 & 1 \\ 2 & 0 & 2 \end{pmatrix} \tag{9.9}$$

where the first row corresponds to oxygen, the second row to hydrogen, and the columns correspond to the compounds, in this case ordered as H_2, O_2, and H_2O.

Conserved quantities A chemical reaction cannot create or destroy elements. Thus, the inner product of the rows, $\mathbf{e}_i$, in the elemental matrix and the reaction vectors, and the columns in **S** (called $\mathbf{s}_j^v$) must be zero, or

$$\langle \mathbf{e}_i, \mathbf{s}_j^v \rangle = 0 \tag{9.10}$$

for all the elements found in the compounds that participate in the reaction. This inner product simply adds up the number of elements in the compounds on each side of the reaction. Because the stoichiometric coefficients are negative for the reactants (the

Table 9.2 The elemental composition of some common metabolites. Adapted from [285].

Compound	Elemental composition	Compound	Elemental composition
Glucose	$C_6H_{12}O_6$	Alanine	$C_3H_7NO_2$
Glucose-6-phosphate	$C_6H_{11}O_9P$	Arginine	$C_6H_{14}N_4O_2$
Fructose-6-phosphate	$C_6H_{11}O_9P$	Asparagine	$C_4H_8N_2O_3$
Fructose-1, 6-phosphate	$C_6H_{10}O_{12}P_2$	Cysteine	$C_3H_7O_2NS$
Dihydroxyacetone phosphate	$C_3H_5O_6P$	Glutamic acid	$C_5H_9NO_4$
Glyceraldehyde-3-phosphate	$C_3H_5O_6P$	Glycine	$C_2H_5NO_2$
1,3-Diphosphoglycerate	$C_3H_4O_{10}P_2$	Leucine	$C_6H_{13}NO_2$
2,3-Diphosphoglycerate	$C_3H_3O_{10}P_2$	Isoleucine	$C_6H_{13}NO_2$
3-Phosphoglycerate	$C_3H_4O_7P$	Lysine	$C_6H_{14}N_2O_2$
2-Phosphoglycerate	$C_3H_4O_7P$	Histidine	$C_6H_9N_3O_2$
Phosphoenolpyruvate	$C_3H_2O_6P$	Phenylalanine	$C_9H_{11}NO_2$
Pyruvate	$C_3H_3O_3$	Proline	$C_5H_9NO_2$
Lactate	$C_3H_5O_3$	Serine	$C_3H_7NO_3$
6-Phosphogluco-lactone	$C_6H_9O_9P$	Threonine	$C_4H_9NO_3$
6-Phosphogluconate	$C_6H_{10}O_{10}P$	Tryptophane	$C_{11}H_{12}N_2O_2$
Ribulose-5-phosphate	$C_5H_9O_8P$	Tyrosine	$C_9H_{11}NO_3$
Ribulose-5-phosphate	$C_5H_9O_8P$	Valine	$C_5H_{11}NO_2$
Xylulose-5-phosphate	$C_5H_9O_8P$	Methionine	$C_5H_{11}O_2NS$
Ribose-5-phosphate	$C_5H_9O_8P$	Sedoheptulose-7-phosphate	$C_7\ H_{13}O_{10}P$
Erythrose-4-phosphate	$C_4H_7O_7P$	5-Phosphoribosyl-1-pyrophosphate	$C_5H_8O_{14}P_3$
Inosine monophosphate	$C_{10}N_4H_{12}O_8P$	Ribose-1-phosphate	$C_5H_9O_8P$
Hypoxanthine	$C_5N_4H_4O$	Inosine	$C_{10}H_{12}N_4O_5$

compounds that disappear in the reaction) and positive for the products (the compounds that appear in the reaction), this sum is zero. The number of atoms of an element on each side of the reaction is the same. For the elemental matrix in Equation (9.9) and the reaction vector $\mathbf{s}_i^v = (-2, -1, 2)^T$, we see that

$$\langle (0,2,1), (-2,-1,2)^T \rangle = 0 \quad \text{and} \quad \langle (2,0,2), (-2,-1,2)^T \rangle = 0 \tag{9.11}$$

All elemental balancing equations taken together lead to the simple matrix equation:

$$\mathbf{ES} = \mathbf{0} \tag{9.12}$$

Table 9.3 The elemental composition of the glycolytic intermediates. This table represents the matrix **E**. Note that NAD is treated as one chemical moiety as it never changes in this system.

	Glu	G6P	F6P	FBP	DHAP	GAP	PG13	PG3	PG2	PE	PYR	LAC	NAD	NADH	AMP	ADP	ATP	P_i	H^+	H_2O
C	6	6	6	6	3	3	3	3	3	3	3	3	0	0	10	10	10	0	0	0
H	12	11	11	10	5	5	4	4	4	2	3	5	0	1	13	13	13	1	1	2
O	6	9	9	12	6	6	10	7	7	6	3	3	0	0	7	10	13	4	0	1
P	0	1	1	2	1	1	2	1	1	1	0	0	0	0	1	2	3	1	0	0
N	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	5	5	0	0	0
S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NAD	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0

Although not shown here, the same must be true of compound electrical charge, as it is balanced during a chemical reaction. The rows of **E** are row vectors that are in the left null space of **S**.

9.2.3 Example: glycolysis

We now give an example of how to formulate a stoichiometric matrix. For this purpose, we pick one of the most familiar pathways of all: glycolysis. This example originates from [317].

Defining the system The glycolytic pathway degrades glucose (a six-carbon compound) to form pyruvate or lactate (three-carbon compounds) as end products. During this degradation process, the glycolytic pathway builds redox potential in the form of NADH and high-energy phosphate bonds in the form of ATP via substrate-level phosphorylation. Glycolysis also assimilates an inorganic phosphate group that is converted into a high-energy bond and then hydrolyzed in the ATP use reaction (or the 'load' reaction). Glycolysis as a system is shown in Figure 9.3.

The compounds or the nodes in the network In its simplest form, glycolysis has 12 primary metabolites, 5 co-factor molecules (ATP, ADP, AMP, NAD, NADH), inorganic phosphate (P_i), protons (H^+) and water (H_2O). The system thus has 20 compounds (see Table 2.1). The elemental composition of these compounds is found in Table 9.3.

The reaction or the links in the network The links formed between these compounds are the glycolytic reactions. There are 21 reactions, including all the transport reactions into and out of the system. The reactions are summarized in Table 2.2.

The stoichiometric matrix The stoichiometric matrix, **S**, can be formulated for the glycolytic system (see Table 9.4). Its dimensions are 20×21, representing the 20 metabolites and the 21 fluxes given in Tables 2.1 and 2.2, respectively.

Table 9.4 An annotated stoichiometric matrix for the glycolytic system in Figure 9.3. The matrix is partitioned to show the co-factors separate from the glycolytic intermediates and to separate the exchange reactions and co-factor loads. The last column has the connectivities, ρ_i, for a compound, and the last row has the participation number, π_i, for a reaction. These two quantities are described in Chapter 10. The second block in the table is the product **ES** to evaluate the elemental balancing status of the reactions. All exchange reactions have only one participating compound (i.e., a participation number of unity) and are not balanced elementally.

	Glycolytic reactions											**AMP metabolism**		**Primary export**		**Co-factors**		**Primary inputs**		**Inorganic**		
	v_{HK}	v_{PGI}	v_{PFK}	v_{TPI}	v_{ALD}	v_{GAPDH}	v_{PGK}	v_{PGLM}	v_{ENO}	v_{PK}	v_{LDH}	v_{AMP}	v_{APK}	v_{PYR}	v_{LAC}	v_{ATP}	v_{NADH}	v_{GLUin}	v_{AMPin}	v_{H^+}	v_{H_2O}	ρ_i
Glu	−1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	2
G6P	1	−1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
F6P	0	1	−1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
FBP	0	0	1	0	−1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
DHAP	0	0	0	−1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
GAP	0	0	0	1	1	−1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
PG13	0	0	0	0	0	1	−1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
PG3	0	0	0	0	0	0	1	−1	0	0	0	0	0	0	0	0	0	0	0	0	0	2
PG2	0	0	0	0	0	0	0	1	−1	0	0	0	0	0	0	0	0	0	0	0	0	2
PEP	0	0	0	0	0	0	0	0	1	−1	0	0	0	0	0	0	0	0	0	0	0	2
PYR	0	0	0	0	0	0	0	0	0	1	−1	0	0	−1	0	0	0	0	0	0	0	3
LAC	0	0	0	0	0	0	0	0	0	0	1	0	0	0	−1	0	0	0	0	0	0	2
NAD	0	0	0	0	0	−1	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	3
NADH	0	0	0	0	0	1	0	0	0	0	−1	0	0	0	0	0	−1	0	0	0	0	3
AMP	0	0	0	0	0	0	0	0	0	0	0	−1	1	0	0	0	0	0	1	0	0	3
ADP	1	0	1	0	0	0	−1	0	0	−1	0	0	−2	0	0	1	0	0	0	0	0	6
ATP	−1	0	−1	0	0	0	1	0	0	1	0	0	1	0	0	−1	0	0	0	0	0	6
P_i	0	0	0	0	0	−1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	2
H^+	1	0	1	0	0	1	0	0	0	−1	−1	0	0	0	0	1	1	0	0	−1	0	8
H_2O	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	−1	0	0	0	0	−1	3
π_j	5	2	5	2	3	5	4	2	3	5	5	1	3	1	1	5	3	1	1	1	1	
C	0	0	0	0	0	0	0	0	0	0	0	−10	0	−3	−3	0	0	6	10	0	0	
H	0	0	0	0	0	0	0	0	0	0	0	−13	0	−3	−5	0	0	12	13	−1	−2	
O	0	0	0	0	0	0	0	0	0	0	0	−7	0	−3	−3	0	0	6	7	0	−1	
P	0	0	0	0	0	0	0	0	0	0	0	−1	0	0	0	0	0	0	1	0	0	
N	0	0	0	0	0	0	0	0	0	0	0	−5	0	0	0	0	0	0	5	0	0	
S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
NAD	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	

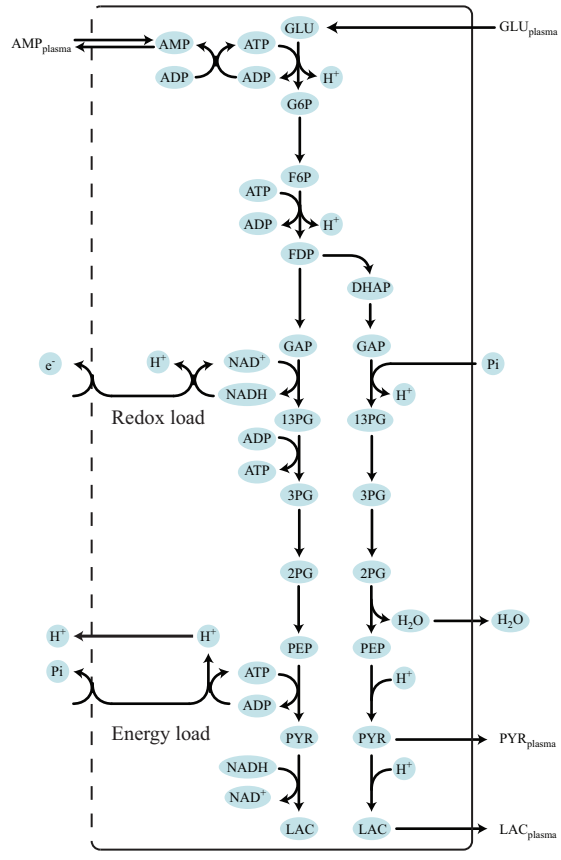


Figure 9.3 Glycolysis as a system: the reaction schema, co-factor interactions (across dashed line), and environmental exchanges (across solid line).

Elemental balancing The stoichiometric matrix needs to be quality-controlled to make sure that the chemical equations are mass-balanced. The elemental compositions of the compounds in the glycolytic system are given in Table 9.3. This table is the elemental matrix, **E**, for this system. We can multiply **ES** to quality control the reconstructed network for elemental balancing properties of the reactions (i.e., verify that $\mathbf{ES} = \mathbf{0}$). The results are shown in Table 9.4. All the internal reactions are balanced elementally. The exchange reactions are not balanced elementally as they represent net addition to or removal from the system as defined.

Charge balancing In this example, we treat the compounds as being uncharged. This assumption is not physiologically accurate, but it will not affect many types of computations. If significant changes in pH are to be considered, then the charged state of the molecules needs to be established. A row can be added into **E** representing the charges of the molecules. Then, charge balance is ensured by making sure that

$\mathbf{ES} = \mathbf{0}$. Charge balancing the whole system for transporters can be difficult, because some of the transport systems, co-transport ions, and individual ions can cross the membrane by themselves. Overall, the system has to be charge-neutral. Accounting for full charge balances and the volume of a system can be quite involved mathematically, see [192, 193].

9.3 Network Structure: S as a Connectivity Matrix

A network can be visually represented as a *map*. Each *node* in the map corresponds to a row in a *connectivity matrix*, and each column corresponds to a *link* in the map.

9.3.1 The maps of S

The reaction map **S** represents a map where a compound is a node and the reactions connect (link) the compounds (Figure 9.4A). This map is the *reaction map* (also called reaction-centered map) and is the standard way of viewing metabolic reactions and pathways in biochemistry textbooks.

The compound map The negative of the transpose of the stoichiometric matrix, $-\mathbf{S}^T$, also represents a map (Figure 9.4B), which we will call the *compound map* (also referred to as the metabolite-centered map). The map that $-\mathbf{S}^T$ represents has the reactions (now the rows in $-\mathbf{S}^T$) as the nodes in the network and the compounds (now the columns of $-\mathbf{S}^T$) as the connections, or the links. This representation of a biochemical reaction network is unconventional, but useful in many circumstances.

Examples: Simple examples of reaction and compound maps are shown in Figure 9.5. The compound map for glycolysis is shown in Figure 9.6. The compound map can be complicated notably by highly connected co-factor molecules.

9.3.2 Biological quantities displayed on maps

It is worth examining the columns ($\mathbf{s}_i^v$) and rows of **S** a bit more closely. Let's examine a reaction:

$$x_1 + x_2 \xrightarrow{v_i} x_3 + x_4 \tag{9.13}$$

with the corresponding column of **S**, $\mathbf{s}_i^v = (-1, -1, 1, 1)^T$. This vector is in the column space of **S**. Moving along this vector is like carrying out this reaction. Note that motion along this vector will conserve the sum $x_1 + x_2 + x_3 + x_4$. Thus, a column in **S** represents a 'tie' between the compounds participating in a particular reaction. If these compounds participate in other reactions, there will be interactions between the motions along the columns of **S**. These vectors, $\mathbf{s}_i^v$, span the column space of **S** and thus give a conceptually useful basis for the column space of **S**. As we will see, certain combinations of the column vectors form pathways through the reaction map.

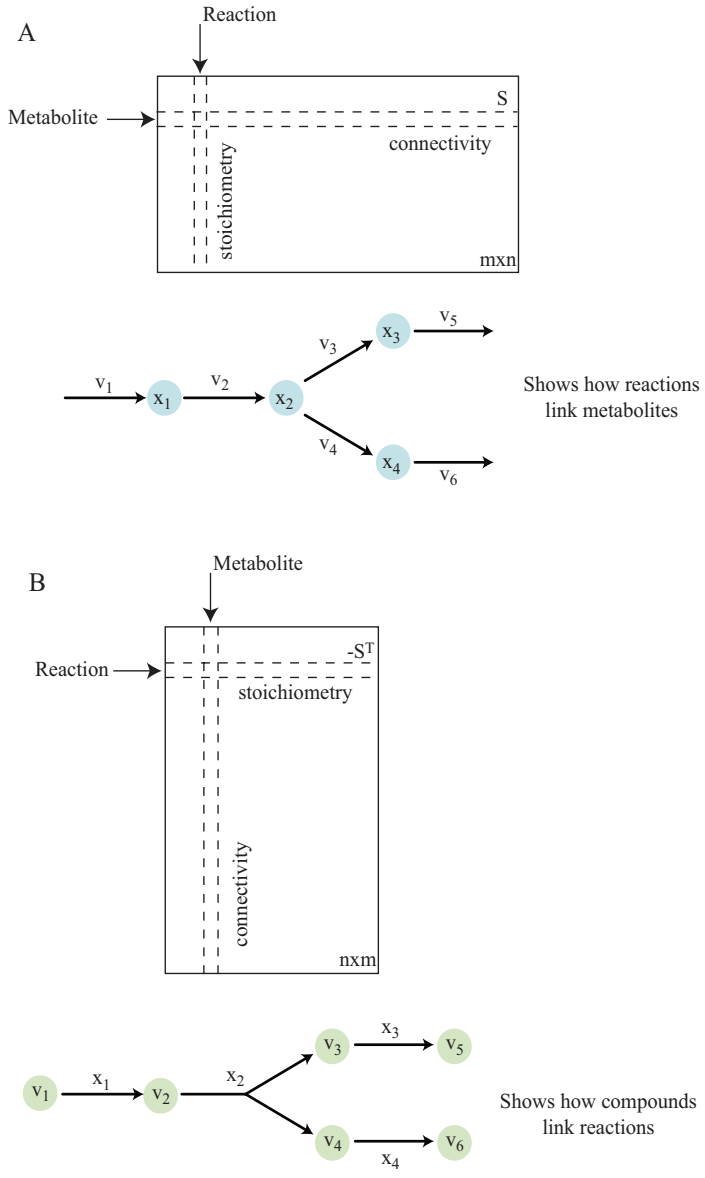


Figure 9.4 The structure of the stoichiometric matrix and how it corresponds to a map. Both the regular (reaction) and transpose (compound) maps are shown in panels A and B, respectively.

Similar observations apply to the rows of **S** (or the columns of $-\mathbf{S}^T$). A column in $-\mathbf{S}^T$ will 'tie' together, or connect, all the reactions in which a metabolite participates. These connections, however, do not imply any particular relationship among reactions, and therefore are not considered 'hard' connections. As will be further discussed in Chapter 11, metabolite pools form, which are the linear combinations of metabolite concentrations. These pools represent a conservation among metabolites that is

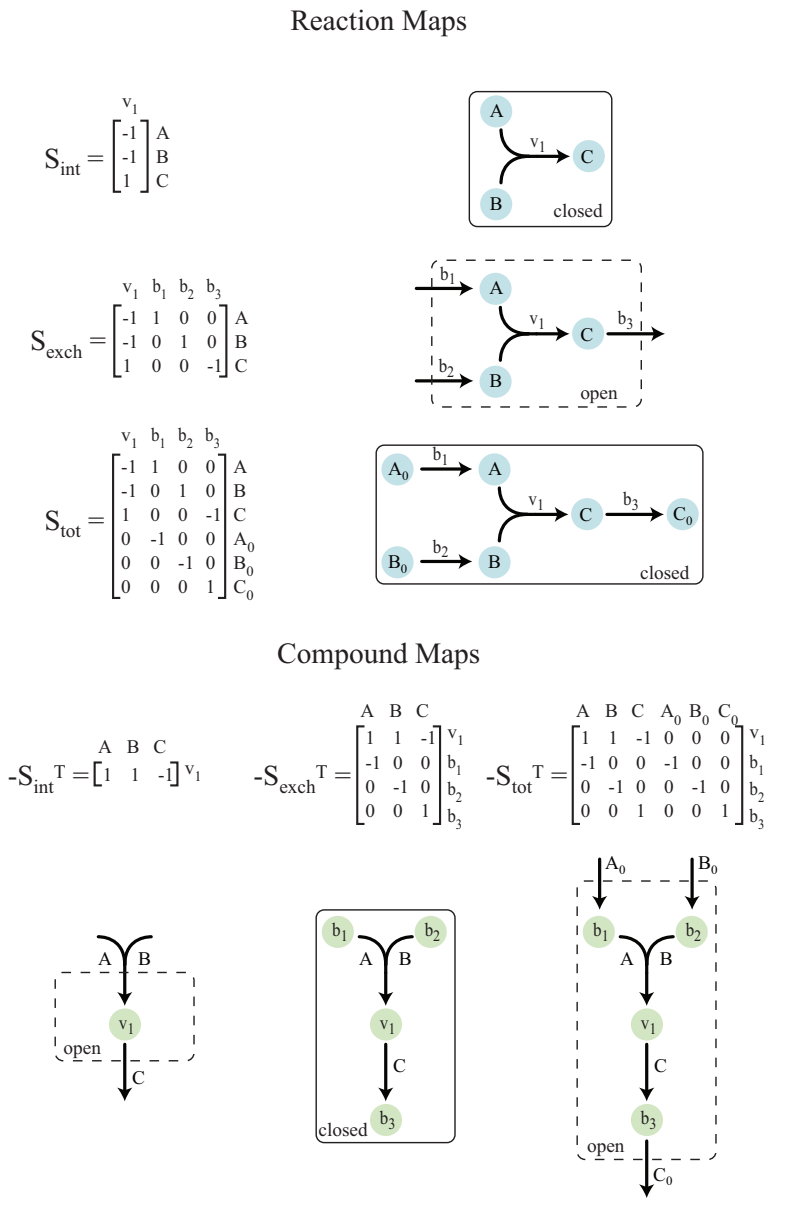


Figure 9.5 Simple examples of reaction maps versus compound maps. Reaction maps (top) show metabolites as nodes and reactions as directed edges. The reaction map includes both the internal and exchange fluxes, if present. In contrast, compound maps of the same systems (bottom) show the reactions as nodes and metabolites as directed edges. A systems boundary that allows for the exchange of the internal nodes is open on a reaction map. The compound map of an open reaction map is closed, and vice versa, as is shown by changing the network from (A) to (B) and to (C).

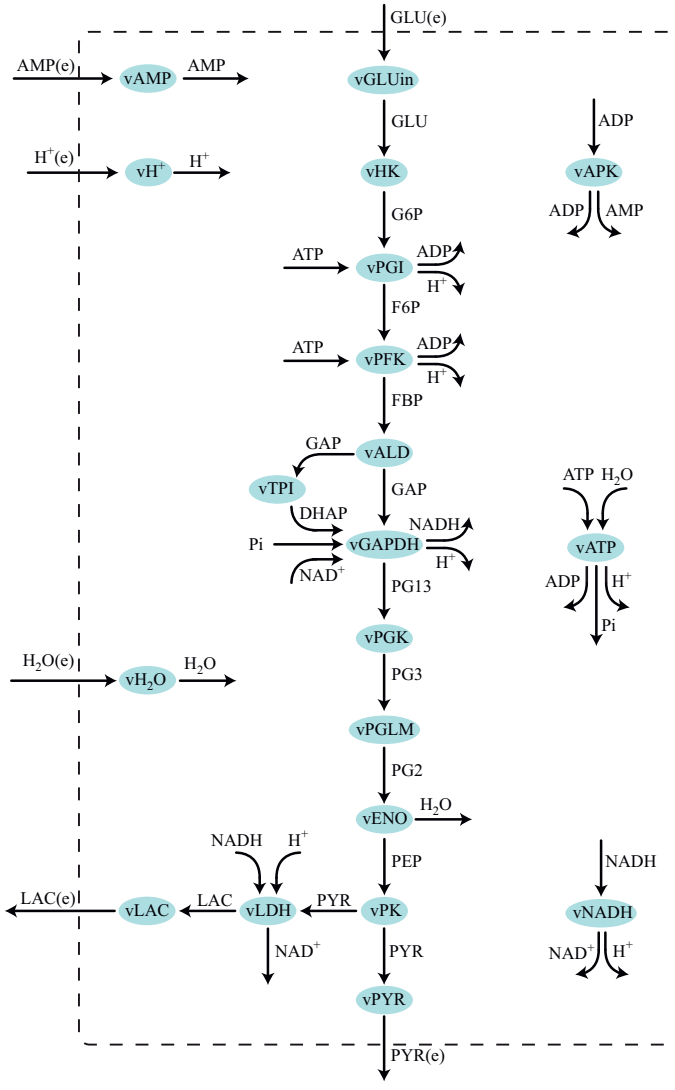


Figure 9.6 The compound map for glycolysis, shown in Figure 9.3. The connections have been broken up to simplify the appearance of the map. Prepared by Addiel De Alba Solis.

mediated by specific reactions, and therefore represents a more meaningful relationship among reactions in which the metabolites participate.

Note that the columns of **S**, in contrast, create a 'hard' connection between the metabolites, as a reaction will simultaneously use and produce the participating compounds. Conversely, the connectivities created between the reactions are 'soft,' as the reactions in which a compound participates can have varying flux levels that may not have fixed ratios. These ratios are determined by the kinetic properties of the reactions.

9.3.3 Linearity of maps

The topological structure of the maps formed by connectivity matrices are very important in determining the properties of the network. The topological properties of maps can be linear and non-linear.

Linear maps are made up of links that have only one input and one output. Thus, the columns of **S** will only have two entries, corresponding to the two nodes (metabolites) that the link (reaction) connects. Similarly for $\mathbf{S}^T$, if only one compound links two reactions, the map is linear. Although frequently used for illustrative purposes, the occurrence of such links in biological reaction networks is rare.

Non-linear maps are made up of links with more than one input or more than one output. The number of compounds that participate in a reaction can be found by adding up the non-zero elements in the corresponding column of **S**. In genome-scale metabolic models, the most common number of metabolites participating in a reaction is four, as in reaction (9.5). Thus, metabolic co-factors create non-linearity in the map of **S**.

Metabolites that participate in more than two reactions create a non-linearity in the map of $\mathbf{S}^T$. The participation number of a metabolite in genome-scale models can be as high as 150 (for ATP), but is 2 for most compounds found in the cell [105]. The metabolites that participate in many reactions thus create strong non-linearities in the compound map. The co-factors lead to strong non-linear topological features of metabolic networks.

9.4 Mathematics: S as a Linear Transformation

9.4.1 Mapping fluxes onto concentration time derivatives

Mathematically, the stoichiometric matrix **S** is a *linear transformation* (Figure 9.7) of the flux vector,

$$\mathbf{v} = (v_1, v_2, \ldots, v_n) \tag{9.14}$$

to a vector of time derivatives of the concentration vector

$$\mathbf{x} = (x_1, x_2, \ldots, x_m) \tag{9.15}$$

as:

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathbf{S}\mathbf{v} \tag{9.16}$$

The reader may also be familiar with other notations of time derivatives

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathbf{x}' = \dot{\mathbf{x}} \tag{9.17}$$

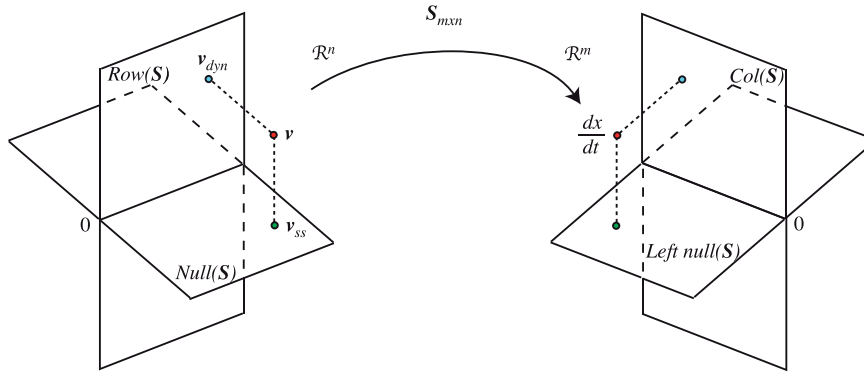


Figure 9.7 The stoichiometric matrix as a linear transformation. The four fundamental subspaces of **S** are shown. Prepared by Iman Famili.

which perhaps makes it clearer that the $\mathrm{d}\mathbf{x}/\mathrm{d}t$ is a vector, and that

$$\dot{\mathbf{x}} = \mathbf{S}\mathbf{v} \tag{9.18}$$

is a linear transformation.

Dimensions There are m metabolites (x_i) found in the network and n reactions (v_i), thus:

$$\dim(\mathbf{x}) = m, \;\; \dim(\mathbf{v}) = n, \;\; \dim(\mathbf{S}) = m \times n \tag{9.19}$$

For a typical biological network there are more reactions than compounds, or $n > m$. The matrix **S** may not be full rank, and therefore $\mathrm{rank}(\mathbf{S}) = r < m$.

9.4.2 The four fundamental subspaces

There are four fundamental subspaces associated with a matrix. Figure 9.7 shows the four fundamental subspaces of **S**, which have important roles in the analysis of biochemical reaction networks, as detailed in the following chapters. The vector produced by a linear transformation is in two orthogonal spaces (the column and left null spaces), and the vector being mapped is also in two orthogonal spaces (the row and null spaces).

Dimensions of the fundamental subspaces The mapping that the stoichiometric matrix represents is illustrated in Figure 9.7. The stoichiometric matrix is typically rank-deficient. The rank r of a matrix denotes the number of *linearly independent* rows and columns that the matrix contains. Rows are linearly dependent if any one row can be computed as a linear combination of the other rows. Linear dependency between the compounds and reactions determines the dimensionality of each of the four fundamental subspaces.

The dimensions of both the column and row space is r:

$$\dim(\text{Col}(\mathbf{S})) = \dim(\text{Row}(\mathbf{S})) = r.$$

Because the dimension of the concentration vector is m, we have

$$\dim(\text{Left Null}(\mathbf{S})) = m - r.$$

Similarly, the flux vector is n-dimensional, thus

$$\dim(\text{Null}(\mathbf{S})) = n - r.$$

Contents of the fundamental subspaces The four fundamental subspaces contain important information about a reaction network. Their contents are as follows.

Null space: the null space of **S** contains all the steady-state flux distributions allowable in the network. The steady state is of much interest as most homeostatic states are close to being steady states.

Row space: the row space of **S** contains all the dynamic flux distributions of a network, and thus the thermodynamic driving forces that change the rate of reaction activity.

Left null space: the left null space of **S** contains all the conservation relationships, or *time-invariants*, that a network contains. The sum of conserved metabolites or conserved metabolic pools do not change with time and are combinations of concentration variables.

Column space: the column space of **S** contains all the possible time derivatives of the concentration vector, and thus how the thermodynamic driving forces move the concentration state of the network.

Basis for vector spaces A *basis* for a space can be used to *span* the space. Thus, a basis describes all of the contents of a space. Different bases can be used for this purpose, including a linear basis like the commonly used orthonormal basis, and a convex basis for finite linear spaces. The choice of basis for the four fundamental subspaces becomes important because it influences the interpretation of the contents of a space. Singular value decomposition gives simultaneous orthonormal bases for all the four fundamental subspaces (Chapter 11). Chapter 12 gives alternative and more biologically meaningful sets of basis vectors.

9.4.3 Looking into the four fundamental subspaces

The column and left null spaces The time derivative is in the column space of **S** (denoted by Col(**S**)) as can be seen from the expansion of **Sv**:

$$\frac{d\mathbf{x}}{dt} = \mathbf{s}_1^v v_1 + \mathbf{s}_2^v v_2 + \cdots + \mathbf{s}_n^v v_n \tag{9.20}$$

where the $\mathbf{s}_i^v$ are the *reaction vectors* that form the columns of $\mathbf{S}$. The Col($\mathbf{S}$) is therefore spanned by the reaction vectors, $\mathbf{s}_i^v$. The reaction vectors are structural features of the network, and are fixed. However, the fluxes v_i are scalar quantities and represent the flux through reaction i. The fluxes are variables. We do note that each flux has a maximal value, $v_i \leq v_{i,\max}$, and this limits the size of the time derivatives. Thus, only a portion of the column space is explored, i.e., we can cap the size of the column space of $\mathbf{S}$. The vectors in the left null space ($\mathbf{l}_i$) of $\mathbf{S}$ are orthogonal to the column space, i.e., $\langle \mathbf{l}_j \cdot \mathbf{s}_i^v \rangle = 0$. The vectors $\mathbf{l}_i$ represent a mass conservation (see Chapter 11).

The row and null spaces The flux vector can be decomposed into a dynamic component and a steady-state component,

$$\mathbf{v} = \mathbf{v}_{\text{dyn}} + \mathbf{v}_{\text{ss}} \tag{9.21}$$

The steady-state component satisfies

$$\mathbf{S}\mathbf{v}_{\text{ss}} = 0 \tag{9.22}$$

and $\mathbf{v}_{\text{ss}}$ is thus in the null space of $\mathbf{S}$ (see Chapter 12). The dynamic component of the flux vector, $\mathbf{v}_{\text{dyn}}$, is orthogonal to the null space and consequently is in the row space of $\mathbf{S}$.

Recapitulation Each pair of subspaces, where $\mathbf{v}$ and $\dot{\mathbf{x}}$ reside, form orthogonal sets to each other, and their dimensions sum up to the dimension of their corresponding vectors, i.e., dim(Null($\mathbf{S}$)) + dim(Row($\mathbf{S}$)) = n and dim(Left null($\mathbf{S}$)) + dim(Col($\mathbf{S}$))= m. These are introductory observations about $\mathbf{S}$ and its fundamental subspaces. In Chapter 11, we will study the individual fundamental subspaces in more detail.

9.5 Systems Science: S and Network Models

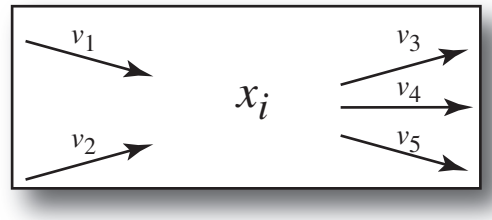
Dynamic mass balances A dynamic mass balance on a compound is formed by summing up the fluxes through all the reactions that form the compound and subtracting those that degrade it (see Figure 9.8). In general, such a dynamic mass balance is described by the ordinary differential equation:

$$\frac{\mathrm{d}x_i}{\mathrm{dt}} = \sum_k s_{ik} v_k = \langle \mathbf{s}_i^x \cdot \mathbf{v} \rangle \tag{9.23}$$

representing a summation of all fluxes v_k that form compound x_i, and those that degrade it. s_{ik} is the corresponding stoichiometric coefficient. The set of all such differential equations that describe the dynamic mass balance of every compound in a network is represented by the matrix equation

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{dt}} = \mathbf{S}\mathbf{v} \tag{9.24}$$

A



B

$$\frac{dx_i}{dt} = \underbrace{v_1 + v_2}_{formation} - \underbrace{v_3 - v_4 - v_5}_{degradation}$$
$$= \left\langle (1,1,-1,-1,-1)\ ,\ (v_1,v_2,v_3,v_4,v_5)^{\mathrm{T}} \right\rangle$$
$$= \left\langle \mathbf{s}_i^x \cdot \mathbf{v} \right\rangle$$

Figure 9.8 The dynamic mass balance on a single compound. Panel A shows all the rates of formation and degradation of a compound x_i (a graphical representation called a node map). Panel B shows the corresponding dynamic mass balance equation that simply states that the rate of change of the concentrations x_i is equal to the sum of the rates of formation minus the sum of the rates of degradation. This summation can be represented as an inner product between a row vector, $\mathbf{s}_i^x$ and the flux vector, $\mathbf{v}$. This row vector becomes a row in the stoichiometric matrix in Equation (9.16).

and thus $\mathbf{s}_i^x$ is a row in $\mathbf{S}$. Equation (9.24) represents the fundamental equation of the *dynamic mass balances* that characterizes all functional states of a reconstructed biochemical reaction network. The stoichiometric matrix is a key component of this relationship.

Systems boundary The boundaries around a network can be drawn in different ways (see Figure 9.9). When defining a network, a *systems boundary* is drawn. The reactions are then partitioned into internal and exchange reactions. Exchange, or *boundary*, fluxes are denoted with b_i and internal fluxes with v_i. Similarly, the concentration vector is partitioned into internal (x_i) and external (c_i) concentrations. There are several different versions of $\mathbf{S}$ depending on what is encompassed by a network. A specific example is provided in Figure 9.10.

Defining the systems boundary Note that in the above consideration we have drawn a systems boundary around the cell. Such definition is common because it is consistent with physical realities. However, because the definition of a systems boundary can be chosen, we can segment any network into subnetworks by drawing 'virtual' boundaries. This property is useful in defining subsystems that may be 'fast' (i.e.,

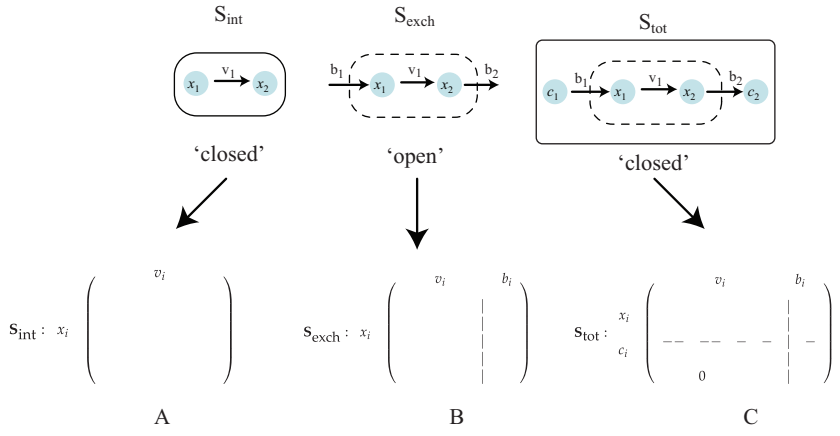


Figure 9.9 Schematic illustration of open and closed networks. (A) The internal stoichiometric matrix: considers the cell as a closed system; we focus just on the internal fluxes. This form is useful to define pools of compounds that are conserved (Chapter 11) and closed loop pathways (Chapter 12). (B) The exchange stoichiometric matrix: does not consider the external compounds, and contains the internal fluxes and the exchange fluxes with the environment. This form of the matrix is frequently used in pathway analysis of a network (see Chapter 12). (C) The total stoichiometric matrix. This form of **S** is the most general one. The dashed lines show the partitioning of the internal elements in the matrix. This form accounts for the internal reactions (v_i), the exchange reactions (b_i), the internal compounds (x_i), and the external compounds (c_i).

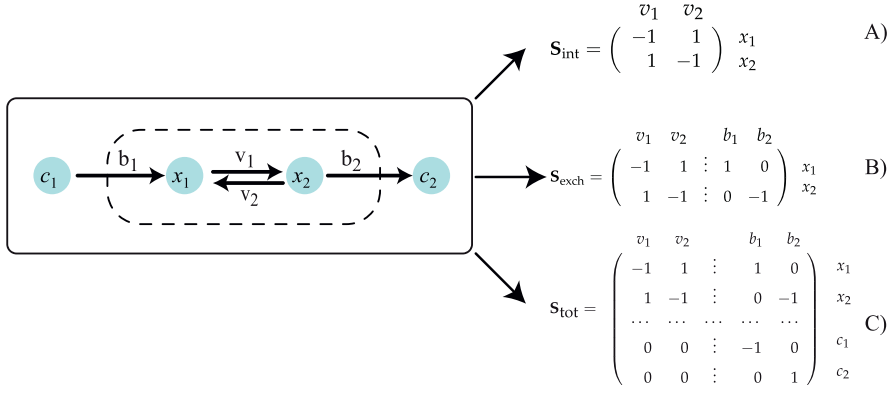


Figure 9.10 An example network. (A) The internal stoichiometric matrix has $m = 2$, $n = 2$, and $r = 1$. Thus, all the fundamental subspaces have a dimension of 1. The internal stoichiometric matrix can be further partitioned. The compounds that cannot be exchanged with the environment form one group, and those that can, form another. (B) The exchange stoichiometric matrix has $m = 2$, $n = 4$, and $r = 2$. It is full rank. The null space has a dimension of 2 ($= 4 - 2$), while the left null space has a dimension of 0 ($= 2 - 2$). (C) The total stoichiometric matrix has $m = 4$, $n = 4$, and $r = 3$. Thus, both of the null spaces are one-dimensional. In later chapters, we will learn how the dimensions of the null spaces relate to pools and pathways.

have rapid dynamics) and lead to temporal decomposition, and subsystems that have biochemical relevance (e.g., fatty acid biosynthesis).

9.6 Summary

- The stoichiometric matrix is a mathematical representation of a reconstructed network. It represents the biochemical, genomic, and genetic information on which the reconstruction is based.
- The stoichiometric matrix consists of stoichiometric coefficients that are integer numbers. The columns of the stoichiometric matrix represent chemical reactions while the rows represent compounds.
- The stoichiometric matrix includes informatic, chemical, physical, genetic, genomic, mathematical, numerical, and systemic attributes of the biological system that it represents.
- The stoichiometric matrix has many important features: (1) it is a data matrix, (2) it gives the structure of a network, (3) it is a mathematical mapping operation with four fundamental subspaces, and (4) it forms a key part of *in silico* models describing the functional states of networks.
- The stoichiometric matrix is a data matrix. The columns, $\mathbf{s}_i^v$, represent chemical transformations and thus come with chemical information. They are reaction vectors, that imply elemental and charge balance. The reaction vectors are thus orthogonal to the rows of the elemental matrix. These conservation quantities are in the left null space of the stoichiometric matrix. Some quantities, such as free energy, are non-conserved during a chemical reaction. These quantities will be in the row space of the stoichiometric matrix.
- The stoichiometric matrix is a connectivity matrix; it represents a reaction map. The transpose of the stoichiometric matrix represents a compound map. Both maps are topologically non-linear, as they contain joint edges between nodes.
- Mathematically, the stoichiometric matrix represents a transformation, or a mapping, of one vector (the flux vector) to another (the vector of time derivatives of the concentrations). Such a mapping operation comes with four fundamental subspaces (the row, the null, the column, and the left null spaces). Each one of these spaces contains chemically and physically meaningful quantities.
- The stoichiometric matrix is a key component of a systems model of network functions that is in the form of dynamic mass balances. The boundaries of a reaction network can be drawn in different ways and lead to three fundamental forms of **S**.