J Math Chem (2010) 47:1287–1302
DOI 10.1007/s10910-009-9654-x

ORIGINAL PAPER

Torsion in biochemical reaction networks

Peter H. Sellers

Received: 29 September 2009 / Accepted: 26 November 2009 / Published online: 3 January 2010

Abstract This article starts in Part I with a simple example of two biochemical reaction networks that are indistinguishable at the macroscopic level but are different at the molecular level and are shown to have significantly different kinetic properties. So, if one completely ignores the fact that reactions advance in discrete steps at the molecular level, then one can fail to distinguish between networks with widely different kinetics. In part II biochemical reaction networks are treated in a general way to discover what property of a network, only seen at the molecular level, affects its kinetics. It is shown that every such network has a unique *torsion group,* which can be described numerically and readily determined by a programmable computation. If the group is found to be the singleton $\{0\}$ (as is most often the case in practice), then the network is said to be *torsion-free* and its kinetic properties unaffected by ignoring its discrete character. A chemical reaction network has to be represented algebraically to calculate its torsion group. If the network is to be understood only at the macroscopic level, it can be placed in the context of real vector spaces, but to recognize its discrete character and its torsion group, each vector space is replaced by a discrete subset of that space, where each molecule can be recognized as a distinct and indivisible entity. Next, the process of calculating a torsion group is shown in several cases, including the example in part I. In this particular case it is shown to have the torsion group with 2 elements, reflecting the fact that the substrate molecules become product molecules 2 at a time, with the result that the overall macroscopic reaction is $R \Leftrightarrow T$, whereas at the molecular level it is $2R \Leftrightarrow 2T$. In general, however, the torsion group of a biochemical reaction network can be any finite additive group, which is a property of the network that can only be seen at the molecular level. Finally, this fact is demonstrated by

P. H. Sellers (✉)
The Rockefeller University, 1230 York Avenue, New York, NY 10065, USA
e-mail: sellers@mail.rockefeller.edu

showing how to construct a hypothetical, but plausible, biochemical reaction network that has any given finite additive group as its torsion group.

1 PART I

1.1 A sample biochemical reaction network

The purpose of Part I is to use a hypothetical biochemical reaction network that exhibits its torsion to demonstrate the importance of that concept in chemical kinetics, before proceeding to Part II, where an algebraic method is given to determine the unique *torsion coefficients* that characterize the torsion of any reaction network. In the case of an explicitly given network the said method reduces to a program using matrix operations to determine the numerical values of its torsion coefficients.

The example that follows involves two versions of an enzyme-catalyzed biochemical reaction network that are indistinguishable at the macroscopic level but have significantly different kinetics.

1.2 Reaction rates

At the macroscopic level our example is as follows:

$$\mathrm{R} + \mathrm{ES} \Leftrightarrow \mathrm{ESS} \Leftrightarrow \mathrm{ES} + \mathrm{T}$$

The substrate R and the product T consist of small molecules of the same species in compartments that are separated by a membrane, and the two intermediates ES and ESS are large molecules embedded in the membrane. There are two elementary collision processes that are described above by balanced macroscopic chemical equations. By virtue of chemical balance it is obvious that the letter S represents a fragment of each intermediate that has the same empirical formula as both R and T. At the molecular level this example is capable of two interpretations, which will be seen to have significantly different kinetic properties.

Let ES_1 and ES_2 denote separate molecules of the same species, then the two following cases differ only at the molecular level.

Case 1

$$\mathrm{R} + \mathrm{ES}_1 \Leftrightarrow \mathrm{ES}_2\mathrm{S}_1 \Leftrightarrow \mathrm{ES}_1 + \mathrm{T}$$

Case 2

$$\mathrm{R} + \mathrm{ES}_1 \Leftrightarrow \mathrm{ES}_2\mathrm{S}_1 \Leftrightarrow \mathrm{ES}_2 + \mathrm{T}$$
$$\mathrm{R} + \mathrm{ES}_2 \Leftrightarrow \mathrm{ES}_1\mathrm{S}_2 \Leftrightarrow \mathrm{ES}_1 + \mathrm{T}$$

In Case 1 the molecule ES_1 is an ordinary catalyst, whereas in Case 2. a pair of distinct molecules, ES_1 and ES_2, of the same species are needed to catalyze $2\mathrm{R} \Leftrightarrow 2\mathrm{T}$ at the

molecular level. To show to what extent they differ kinetically, it is necessary place them in a physical context. So, let us make the following assumptions:

(1) Assume that the reaction R ⇔T takes place on a membrane that separates molecules of the same species into an R-pool and a T-pool at different concentrations, whose values –say in moles per liter– are held constant at r and t, respectively.
(2) Assume that the whole system is maintained at constant temperature and pressure with R ⇔T in a steady-state, so that its net rate of advancement is constant.
(3) A constant finite total number n of molecules of ES and ESS are fixed on the membrane. Let e and f denote the numbers of molecules of ES and ESS, respectively, so that $n = e + f$.
(4) The reactions are governed by the Law of Mass action, where k and k' are the rate constants applicable to each of the reactions R + ES⇔ESS and T + ES⇔ESS. In other words, their forward rates (left-to right) are ker and ket, and their reverse rates (right-to-left) are both $k'f$. This law holds at the macroscopic level, where e and f are expressible as functions of the 5 given parameters r, t, k, k' and n.

With this information the net steady-state rate u will be shown to be equal in both cases to the same function of the 5 given parameters. However, suppose u is separated as follows into *unidirectional rates*:

$$u = u^+ - u_1^-$$

Here, u^+ is the rate at which R-molecules become T-molecules, and it is less than ker, because some molecules of R go back to the R-pool without reaching the T-pool. Likewise, u^- is the rate at which T-molecules become R-molecules, and it is less than ket, because some molecules of T go back to the T-pool without reaching the R-pool.

Show that the unidirectional rates differ in cases 1 and 2, and that u^+ takes the following values in cases 1 and 2, respectively:

$$1/2\,[\mathrm{ker}] \qquad \text{and} \qquad 1/2\,[\mathrm{ker}]\,[r/(r+t)]$$

That is, the second case has a unidirectional rate u^+ which is smaller than that of the first case by a factor of r/(r+t). Similarly, u^- takes the following values in cases 1 and 2, respectively:

$$1/2\,[\mathrm{ket}] \qquad \text{and} \qquad 1/2\,[\mathrm{ket}]\,[t/(r+t)]$$

In other words, the second case has a unidirectional rate u^- which is smaller than that of the first case by a factor of t/(r+t).

Thus it is seen that the two cases can be told apart by their unidirectional rates, even though their net rates and their intermediate concentrations can be seen to be equal in the two cases.

$$\begin{aligned} u^+ - u^- &= (1/2)\ ke\,(r-t) \\ e &= \left[2\,k'n\right] / \left[k\,(r+t) + 2k'\right] \\ f &= n - e \end{aligned}$$

1.2.1 Rate computation in case 1

Expression (1) gives symbols for the forward and reverse rates of each of the four elementary reactions in Case 1. If a letter has the same subscript on two sides of a reaction, then it refers, not only to the same species, but the same molecule.

$$\begin{aligned} R + ES_1 &\overset{w}{\rightarrow} ES_1S_2 \overset{v}{\rightarrow} ES_1 + T \\ R + ES_1 &\overset{v}{\leftarrow} ES_1S_2 \overset{x}{\leftarrow} ES_1 + T \end{aligned} \tag{1}$$

Each will be expressed in terms of the parameters n, r, t, k and k′, which were introduced under assumptions (i) through (iv) to define the chemical kinetic context, where the parameters can be any positive real numbers. The system is in equilibrium if, r = s.

From assumptions (i.) through (iv.) we have the following equation:

$$n = e + f$$

where e and f are the numbers of molecules of ES_2 and ES_2S_1, respectively, and also:

$$w = ker \qquad v = k'f \qquad x = ket$$

The net rate of advancement u of the reaction –R + T is as follows:

$$u = w - v = v - x = ker - k'f = k'f - ket$$

Solve the last row for e and u:

$$\begin{aligned} e &= \left[2k'n\right] / \left[k\,(r+t) + 2k'\right] \\ u &= (1/2)\,(ke)\,(r-t) \end{aligned}$$

The unidirectional rate u^+ at which molecules go from the R-pool to the T-pool excludes those which return to R. Of the molecules leaving the R-pool at rate w half return and the others reach the T–pool.

$$u^+ = (1/2)\ w = (1/2)\ ker$$

Similarly we get the unidirectional rate u^- from the T-pool to the R-pool.

$$u^- = (1/2)\ x = (1/2)\ ket$$

1.2.2 Rate computation in case 2

Expression (2) gives symbols for the forward and reverse rates for each of the eight elementary reactions in Case 2.

$$\begin{aligned} &R_1 + ES_2 \overset{w}{\rightarrow} ES_1S2 \overset{v}{\rightarrow} ES_1 + T_2 \\ &R_1 + ES_2 \overset{v}{\leftarrow} ES_1S_2 \overset{x}{\leftarrow} ES_1 + T_2 \\ &T_1 + ES_2 \overset{x}{\rightarrow} ES_2S_1 \overset{v}{\rightarrow} ES_1 + R_2 \\ &T_1 + ES_2 \overset{v}{\leftarrow} ES_2S_1 \overset{w}{\leftarrow} ES_1 + R_2 \end{aligned} \tag{2}$$

The unidirectional rate u^+, at which R-molecules become T-molecules, will be computed by considering all paths in (2), starting at R_1, following the arrows through any sequence of intermediates containing S_1, and ending at T_1. For each path the rate at which molecules take that path from R_1 to T_1 is computed. The sum of all these rates is $1/2(u^+)$, because every molecule R_1 that becomes T_1 is matched by a molecule R_2 that becomes T_2.

The shortest path, as defined above, from R_1 to T_1 takes 4 steps via ES_1S_2, ES_1 and ES_2S_1. Let this path be denoted by PQ, where P goes from R_1 to ES_1 and Q goes from ES_1 to T_1, each in 2 steps.

Of the R_1 molecules combining with ES_2 to form ES_1S_2 at rate w only ½ of them go on from ES_1S2 to ES_1, of which only the fraction r/(r + s) of them go from ES_1 to ES_1S_2,and of those only ½ of them go on to T_1. Therefore, the rate at which R_1-molecules become T_1 molecules over path PQ is as follows:

$$w\,(1/2)\,(r/\,(r+t))\,(1/2)$$

Next, determine the rate at which R_1-molecules become T_1- molecules in 6 steps. This happens over 2 paths, PMQ and PNQ, where M is the two-step path, ES_1 to ES_2S_1 and ES_2S_1 to ES_1, and, similarly, N is the 2-step path, ES_1 to ES_1S_2 and ES_1S_2 to ES_1. The rate of molecules over path M is reduced by a factor of (1/2)(r/(r+t)) and over path N by a factor of (1/2)(t/(r+t)). Therefore, the rate that R1-molecules become T1-molecules over path PMQ + PNQ, which may be written as P(M + N)Q, is as follows:

$$w\,(1/2)\,[(r/\,(2r+2t)) + (t/\,(2r+2t))]\,(r/\,(r+t))\,(1/2) = (w/4)\left[1/2\right](r/\,(r+t))$$

Every path is of the form PM^mN^nQ for all pairs (m,n) of non-negative integers. Accordingly, all paths can be expressed as follows:

$$\begin{aligned} PQ \;+\; &P\,(M+N)\,Q + P\,(M+N)^2\,Q + P\,(M+N)^3\,Q + \cdots \\ &= P\left[1 + (M+N) + (M+N)^2 + (M+N)^3 + \cdots\right]Q \end{aligned}$$

Therefore the rate that R1-molecules become T1-molecules over all paths is as follows:

$$(w/4)\left[1 + (1/2) + (1/2)^2 + (1/2)^3 + \ldots\right](r/(r+t)) = (w/2)\,(r/(r+t))$$

This is $1/2$ of the unidirectional rate u^+, but w is $1/2$ the rate ker that molecules of type either R_1 or R_2 are entering. Therefore,

$$u^+ = (1/2)\,\ker\,(r/(r+t))$$

and by symmetry

$$u^- = (1/2)\,\mathrm{ket}\,(t/(r+t))\,.$$

Hence the net rate of R to T is

$$u = u^+ - u^- = (1/2)\,\mathrm{ke}\,(r-t).$$

This completes the demonstration that a chemical reaction network has a characteristic at the molecular level that affects its kinetics, which will be described in algebraic terms in Part II as its *torsion group.*

2 PART II

2.1 Discrete additive groups

Aspects of the study of biochemical reaction networks that require some familiarity with discrete additive groups have been deferred to this part. It will be shown that pairs of integer matrices with product zero can be used to specify and examine explicitly given biochemical reaction networks one at a time, but, to examine the properties of such networks in general, the pairs of integer matrices will be replaced by pairs of linear transformations from one discrete subgroup of a real vector space to another, where the composition of each pair of transformations is the zero function. This algebraic structure is known as a *2-dimensional chain complex.* The discrete subgroups of vector spaces are not themselves vector spaces but technically known as *finitely generated free additive (or Abelian) groups.*

The idea of using a chain complexes to describe network-like structures is not new. For instance, an n-dimensional chain complex, which is a sequence of n linear transformations such that the product of every successive pair is zero, has been used to describe the networklike structure of an n-dimensional polyhedron embedded in a (2n+1)-dimensional Euclidean space as in Pontryagin [1]. A graph from this point of view is a 1-dimensional chain complex that can be used to describe the combinatorial structure of a 1-dimensional polyhedron (a set of line segments that meet only at certain of their end-points) embedded in 3-space. Higher dimensional chain complexes have been used to characterize networks of instructions as in computer science [2].

In algebra and topology the use of chain complexes and their homology groups is well established, but in applied mathematics, especially in the study of networks, they have been largely overlooked. M. E. J. Newman has written a review [3] of the ways networks have been treated mathematically and concludes, "We do not yet, as we do in some other fields, have a systematic program for characterizing network structure." The present article may be seen as a step in the direction that Newman had in mind, based on the idea of putting networks in the context of finitely generated additive groups. This mathematical context is well covered in a self-contained chapter VI that was added to later editions of a book by W. Lederman [4]. It includes the classification of all finite additive groups, which will be relevant here.

The absence of an accepted way to characterize networks in general has also been observed in the context of chemical reaction networks. D. Angeli and E. O. Sontag [5] have written that, " … a unified theory encompassing networks of arbitrary topology as well as reactions with arbitrary kinetics is presently not available." Such a theory is introduced here. It is to be expressed in algebraic language, which requires a brief comment on how the words "topology" and "kinetics" must be understood in this context.

2.1.1 Topology

It cannot be assumed that the word *topology*, as used in the above quotation, is a unique attribute of a given chemical reaction network. From a mathematical point of view one is bound to assume that a network can be modeled by many different topological spaces, each of which can bring with it some extraneous network property that does not hold in general. Accordingly, what is needed is some common property of all topological models of any given network. Such a property is the *homology group* of the network, which is unique to the network and can be found by readily programmable matrix operations. Algebraically the homology group of a chemical reaction network is described by a finitely generated additive group, based on a formal mathematical definition of a network.

2.1.2 Kinetics

It has been shown in Part I that the *kinetics* of a chemical reaction network cannot be determined in general without taking into consideration the fact that chemical processes advance in discrete steps at the molecular level. Therefore, chemical kinetics, as studied here, will be confined to discrete subsets of real vector spaces. Every such subset is generated by the additive operations of vector analysis, but scalar multiplication by real numbers other than integers is not a permissible operation. This change in the familiar rules of chemical kinetics is required by the example in Part I, case 2, to place it in a suitable mathematical context. In that case there is no chemical mechanism at the molecular level that could change one R-molecule into a T-molecule. So the chemical equation 2R $\Leftrightarrow$2T is valid, but R $\Leftrightarrow$T is not.

Starting with the kind of information that is the focus of molecular biology and describing biochemical systems at the molecular level, one can construct the stoichiometric matrix and the empirical formula matrix, which together define a biochemical

reaction network at that level. The chemical principles that are inherent in this definition are as follows:

(1) Stoichiometry: There is an (e × m) *stoichiometric matrix* that has one row for each of the network's e elementary collision processes and one column for each of the network's m molecules. Each row is made up of the coefficients in the chemical equation that describes the collision process of that row with negative integers for reactants and positive integers for products. In other words, if R ⇔S is the conventional chemical reaction equation for an elementary collision process, where R and S are linear combinations of molecules with positive coefficients, then −R + S will be called a *reaction* and its coefficients are the entries in the row of the stoichiometric matrix corresponding to the collision process in question.

(2) Empirical Formulas: There is an (m × a) *empirical formula matrix* that has one row for each of the network's m molecules and one column for each of the network's a atoms (or stable radicals). Each row is made up of the integers that appear as subscripts in the traditional empirical formula. To be consistent with algebraic practice, it will be necessary here to express every empirical formula as a linear combination of atoms (or stable radicals) with non-negative integer coefficients.

(3) Balance: Chemical reactions are *balanced,* which means in algebraic language that the product of the matrices in (1) and (2) equals zero, that is, the (e × a) matrix of zeros.

The two integer matrices used to define a particular biochemical reaction network at the molecular level can be replaced by linear transformations, which allow the general properties of any such network to be studied. The stoichiometric matrix is replaced by the function Δ_2 of the set of all elementary collision processes in the network, such that for each such process E the reaction it produces is given by $\Delta_2(E)$, which is an integral linear combination of molecules. Similarly, the empirical formula matrix is replaced by the function Δ_1 of the set of all molecules in the network, such that for each such molecule M its empirical formula is given by $\Delta_1(M)$, which is an integral linear combination of atoms or stable radicals. These two functions extend uniquely the following linear functions, whose composition is zero:

$$\mathbf{E} \overset{\Delta_2}{\rightarrow} \mathbf{M} \overset{\Delta_1}{\rightarrow} \mathbf{A}$$

The symbols **E, M** and **A** in bold print are the sets of every integral linear combination E of elementary collision processes, every integral linear combination M of molecules, and every integral linear combination A of atoms or stable radicals, respectively, in the network. These three domains are discrete subgroups of real vector spaces and are distinct from one another. The emphasis on distinctness is to avoid confusion between a collision process and the reaction it produces and between a molecule and its empirical formula. If the elements of the domains had been defined with real coefficients, as a chemical kineticist might have done, then these domains would have been *real vector* spaces and would have represented the macroscopic view of the network. However, if they are defined, as here, using integers rather than real numbers, as coefficients, then

they are correctly described as *finitely generated free additive groups* and represent the network at the molecular level. In the additive group context their two transformations are correctly described as *homomorphisms*. These mathematical terms are consistent with the discrete nature of chemical processes and make mathematical sense of the accepted use of the + sign in chemical reactions.

A simple example is presented next to clarify, how Δ_2 and Δ_1 are related to the stoichiometric matrix and the empirical formula matrix.

2.2 A simple example

Suppose we have a biochemical reaction network with E_1, E_2, E_3 as its elementary collision processes, involving three small molecules A, B, AB and three large intermediates E, EA, EAB. In addition, let us denote by [E], [A], [B] all the stable radicals that appear in the empirical formulas of the above six molecules. Then, instead of defining the network by two matrices, let us use the two functions Δ_1 and Δ_2 such that Δ_1 assigns to each molecule its empirical formula, written in additive notation, and Δ_2 assigns to each collision process the reaction it produces. The network is characterized explicitly as follows:

$$\begin{array}{lll}
\Delta_1(\mathrm{A}) = [\mathrm{A}] & \Delta_1(\mathrm{E}) = [\mathrm{E}] & \Delta_2(\mathrm{E}_1) = -\mathrm{E} - \mathrm{A} + \mathrm{EA} \\
\Delta_1(\mathrm{B}) = [\mathrm{B}] & \Delta_1(\mathrm{EA}) = [\mathrm{E}] + [\mathrm{A}] & \Delta_2(\mathrm{E}_2) = -\mathrm{EA} - \mathrm{B} + \mathrm{EAB} \\
\Delta_1(\mathrm{AB}) = [\mathrm{A}] + [\mathrm{B}] & \Delta_1(\mathrm{EAB}) = [\mathrm{E}] + [\mathrm{A}] + [\mathrm{B}] & \Delta_2(\mathrm{E}_3) = -\mathrm{E} - \mathrm{AB} + \mathrm{EAB}
\end{array}$$

Notice that the reactions in the last column are written in transposed form, thus arbitrarily assigning a positive direction to each of the three elementary collision processes in the network.

Every possible mechanism M in this network is an integral linear combination of elementary collision processes expressible as follows:

$$\mathrm{M} = \mathrm{m}_1\mathrm{E}_1 + \mathrm{m}_2\mathrm{E}_2 + \mathrm{m}_3\mathrm{E}_3$$

The reaction produced by M is as follows:

$$\begin{aligned}
\Delta_2(\mathrm{M}) &= \mathrm{m}_1\Delta_2(\mathrm{E}_1) + \mathrm{m}_2\Delta_2(\mathrm{E}_2) + \mathrm{m}_3\Delta_2(\mathrm{E}_3) \\
&= \mathrm{m}_1(-\mathrm{E} - \mathrm{A} + \mathrm{EA}) + \mathrm{m}_2(-\mathrm{EA} - \mathrm{B} \\
&\quad + \mathrm{EAB}) + \mathrm{m}_3(-\mathrm{E} - \mathrm{AB} + \mathrm{EAB}) \\
&= -\mathrm{m}_1\mathrm{A} - \mathrm{m}_2\mathrm{B} - \mathrm{m}_3\mathrm{AB} - (\mathrm{m}_1 + \mathrm{m}_3)\mathrm{E} \\
&\quad + (\mathrm{m}_1 - \mathrm{m}_2)\mathrm{EA} + (\mathrm{m}_2 + \mathrm{m}_3)\mathrm{EAB}
\end{aligned}$$

The above equation is as follows in matrix notation:

$$\begin{array}{l}
(\mathrm{m}_1, \mathrm{m}_2, \mathrm{m}_3)\ (-1\ 0\ \ 0\ \ -1\ 1\ \ 0) = (-\mathrm{m}_1, -\mathrm{m}_2, -\mathrm{m}_3, (-\mathrm{m}_1 - \mathrm{m}_3), (\mathrm{m}_1 - \mathrm{m}_2), \\
\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (\mathrm{m}_2 + \mathrm{m}_3)) \\
\qquad\quad (0\ \ -1\ 0\ \ 0\ \ -1\ 1) \\
\qquad\quad (0\ \ 0\ \ -1\ -1\ 0\ \ 1)
\end{array}$$

The (3×6) matrix in the above matrix equation is the *stoichiometric matrix* of the network. The empirical formula matrix is found similarly by applying Δ_1 to any linear combination of molecules:

$$\begin{aligned}
&\Delta_1 (n_1 A + n_2 B + n_3 AB + n_4 E + n_5 EA + n_6 EAB) \\
&\quad = n_1 [A] + n_2 [B] + n_3 ([A] + [B]) + n_4 [E] + n_5 ([E] + [A]) + n_6 ([E] \\
&\qquad + [A] + [B]) \\
&\quad = (n_1 + n_3 + n_5 + n_6) [A] + (n_2 + n_3 + n_6) [B] + (n_4 + n_5 + n_6) [E]
\end{aligned}$$

The above equation is as follows in matrix notation:

$$\begin{array}{ll}
(n_1, n_2, n_3, n_4, n_5, n_6) & (1\ 0\ 0) = ((n_1 + n_3 + n_5 + n_6), (n_2 + n_3 + n_6), (n_4 + n_5 + n_6)) \\
& (0\ 1\ 0) \\
& (1\ 1\ 0) \\
& (0\ 0\ 1) \\
& (1\ 0\ 1) \\
& (1\ 1\ 1)
\end{array}$$

The (6×3) matrix in the above matrix equation is the *empirical formula matrix* of the network.

The equivalence of defining a network either in matrix language or the language of additive groups is apparent. The matrix notation is useful when studying the properties of a single network rather than networks in general. Therefore, before turning to the general question of *torsion*, it must be acknowledged that in kinetic studies torsion very often has negligible effects or can be ignored completely. Therefore, when studying a particular network, one would like to know in advance whether the network is *torsion-free* and have no further concern about the effects of torsion in the network under consideration. The matter can be settled by using elementary row operations on the stoichiometric matrix to put it in row echelon form. Then, if the first non-zero entry in every row is $+1$ or -1, the network is torsion-free and not subject to the precautions that are the principal concern of this article. Notice that the (3×6) stoichiometric matrix above is in row echelon form already with -1 in the first non-zero position in every row, so it is torsion-free.

2.3 Torsion

Returning to the general definition in Sect. 2.1 of a biochemical reaction network, as described by the 2-dimensional chain complex

$$\mathbf{E} \xrightarrow{\Delta_2} \mathbf{M} \xrightarrow{\Delta 1} \mathbf{A}$$

Let us define the unique finite *torsion group* associated with this network. If the network is defined at the macroscopic level, its torsion group will be $\{0\}$, but at the

molecular level it could be any finite additive group. What this means from a chemical point of view is that there can be a chemical equation of the form

$$kM_1 \Leftrightarrow kM_2$$

with k an integer greater than 1, which is valid and remains valid only if k is replaced by an integral multiple of k, while the equation

$$M_1 \Leftrightarrow M_2$$

is not valid, in the sense that no chemical mechanism in **E** produces this second equation.

In algebraic language these two equations are stated as follows:

$$(kM_1 - kM_2)\,\varepsilon\,\mathrm{im}\Delta_2 \text{ and } (M_1 - M_2)\,\varepsilon\,(\ker\Delta_1 - \mathrm{im}\Delta_2)$$

where $\mathrm{im}\Delta_2$ is the set of all possible reactions, which is a subset of $\ker\Delta_1$ that includes all balanced elements of **M** whether or not they are reactions. It is by looking at the set of balanced elements that are not reactions in a given network that we measure how much the molecular view of the network differs from the less discerning macroscopic view. They are characterized algebraically by the *factor group*

$$\ker\Delta_1/\mathrm{im}\Delta_2.$$

Additive groups of this kind are discussed in reference [4]. What is important in the present context is that a factor group differs from all the additive groups considered so far in the fact that it can have finite subgroups other than $\{0\}$.

An additive group is said to be *free* if it contains no finite subgroup other than $\{0\}$. All real vector spaces and their subgroups are free. Factoring one vector space by another also produces a free group. The groups **E, M** and **A** are free by definition and all their subgroups are free, including the set $\mathrm{im}\Delta_2$ of all reactions in **M** and the set $\ker\Delta_1$ of all balanced elements of **M**. These groups, which have been used to define a biochemical reaction network at the molecular level, differ from vector spaces in that their factor groups are not necessarily free. Accordingly, the above factor group can contain a non-zero finite subgroup, which, if it exists in a given network, distinguishes it from the many commonly occurring networks whose kinetic properties are the same from the macroscopic or molecular viewpoint. The largest finite subgroup of the above factor group is the *torsion group* of the network.

A network whose torsion group is $\{0\}$ is said to be *torsion-free.* Otherwise it is one of the exceptional networks that are the entire focus of this article.

Since the torsion group of a network can be calculated by a programmable procedure, based on information at the molecular level, as developed in molecular biology, the theory of finitely additive groups [4] has a relevance to programmers and others concerned with implementing discoveries in molecular biology.

2.4 Minimal subnetworks

Having seen that any biochemical reaction network, as described by complex (1), has a unique torsion group **T**, it remains to be shown, conversely, that for any finite additive group **T** there is a biochemically plausible reaction network with **T** as its torsion group. Let us begin by showing that for any network (1)

$$\mathbf{E} \overset{\Delta_2}{\to} \mathbf{M} \overset{\Delta 1}{\to} \mathbf{A} \tag{1}$$

with **T** as its torsion group, there is a subnetwork

$$\mathbf{D} \overset{\Delta_2}{\to} \mathbf{M} \overset{\Delta_2}{\to} \text{ im } \Delta_1 \tag{2}$$

with the same torsion group, which is *minimal* in the sense that Δ_2 is restricted to a minimal subgroup **D** of **E**, such that $\text{im}\Delta_2$ and $\ker\Delta_1$ are the same, respectively, in complex (2) as in complex (1). Consequently, in looking for a complex with any specified torsion group **T**, we shall only need to consider minimal subnetworks.

It is obvious that complexes (1) and (2) have the same $\ker\Delta_1$. They will also have the same $\text{im}\Delta_2$ if **D** is constructed as follows:

Since **M** is a free additive group, its subgroup $\text{im}\Delta_2$ is free. Therefore, $\mathbf{E}/\ker\Delta_2$ is also free, because it is isomorphic to $\text{im}\Delta_2$. Let

$$\{(e_1 + \ker\Delta_2), (e_2 + \ker\Delta_2), \ldots\}$$

be a basis for $\mathbf{E}/\ker\Delta_2$, and define **D** as the subgroup of **E** whose basis is

$$\{e_1, e_2, \ldots\}.$$

In complex (2) the function Δ_2 is one-to-one, and so **D** is the smallest subgroup of **E** such that the groups $\ker\Delta_1$, $\text{im}\Delta_2$ and the homology group $\ker\Delta_1/\text{im}\Delta_2$ are the same in complexes (1) and (2). So the torsion groups are the same in both cases.

The minimal subnetworks of a given network play a role in the theory of higher dimensional networks that is analogous to the role of trees in a graph. To be more precise, the minimal subnetworks of a given graph are the ways that one or more trees can span the graph – spanning forests, so to speak. Their role in chemistry was recognized by Milner [8], and it was sufficiently important that an algorithm [6,7,9,10] was introduced and a program written to list all the minimal subnetworks, as defined above, in a given network of any dimension. The application to chemistry hinges on the fact that every reaction in a minimal subnetwork has a unique mechanism, just as in a tree every two vertices are connected by a unique succession of linked edges. Conversely, every mechanism so constructed belongs to a unique minimal subnetwork and therefore has a unique torsion group. Such a mechanism is characterized by the fact that its elementary collision processes have linearly independent reactions, and it has been described by Milner [8] as a *direct mechanism*. This terminology may come from

the fact that in a graph-theoretic network direct mechanisms reduce to direct (non-self-intersecting) paths from one vertex to another, but in 2-dimensional biochemical reaction networks there is no such simple way to model direct mechanisms. However, a complete list of the direct mechanisms in a given higher dimensional network can be listed by computer, and, using the method described here, they can be classified according to their torsion groups by a programmable procedure.

2.5 Examples of torsion

In this section we consider some hypothetical biochemical reaction networks whose torsion groups reflect the fact that any finite additive group can be the torsion group of some plausible network. Every example will contain the reaction $-\mathrm{R}+\mathrm{T}$ described as the *overall reaction*, and Examples 1, 2 and 3, when viewed at the macroscopic level, will involve the two elementary reactions $-\mathrm{R}-\mathrm{E}+\mathrm{ES}$ and $-\mathrm{ES}+\mathrm{E}+\mathrm{T}$, whose chemical equations are written traditionally in the form (3).

$$\mathrm{R}+\ \mathrm{E} \Leftrightarrow \mathrm{ES} \Leftrightarrow \mathrm{E}+\ \mathrm{T} \tag{3}$$

The examples are based on alternative ways in which these reactions can be understood at the molecular level. R and T will be described as the *terminal species* and all other species as *intermediates.*

Since all equations are assumed to be balanced, R and T are isomers. For simplicity, let us assume that R and T are of the same species, but held in distinct pools, separated by a membrane and held at constant concentrations—say in moles per liter. The membrane is an intermediate region, where a molecule of R or T can enter by combining with an intermediate E to produce another intermediate ES, and the process is reversible. The examples that follow will illustrate different ways these processes can be implemented at the molecular level. With every example it is assumed that there are two reversible elementary collision processes whose macroscopic chemical equations are of the form (3), but whose equations will differ from these at the molecular level, where it is possible to distinguish between two molecules of the same species in the same region. In Example 1 the torsion group will be $\{0\}$, which means the network is torsion-free.

Example 1 Prove that a biochemical reaction network, represented by chain complex (1), is torsion-free, if it is defined as follows:

$\{\mathrm{U}, \mathrm{V}\}$, $\{\mathrm{R}, \mathrm{E}, \mathrm{ES}, \mathrm{T}\}$, $\{\mathrm{E}_0, \mathrm{S}_0\}$ are the sets of basis elements for the free additive groups **E, M, A** in (1), respectively, with Δ_2 and Δ_1 defined thus:

$$\begin{array}{llll} \Delta_2(\mathrm{U}) = -\mathrm{E}-\mathrm{R}+\mathrm{ES} & \Delta_1(\mathrm{R}) = \mathrm{S}_0 & \Delta_1(\mathrm{ES}) = \mathrm{E}_0+\mathrm{S}_0 \\ \Delta_2(\mathrm{V}) = -\mathrm{ES}+\mathrm{E}+\mathrm{T} & \Delta_1(\mathrm{E}) = \mathrm{E}_0 & \Delta_1(\mathrm{T}) = \mathrm{S}_0 \end{array}$$

A basis for $\mathrm{im}\Delta_2$ is $\{-\mathrm{E}-\mathrm{R}+\mathrm{ES}, -\mathrm{ES}+\mathrm{E}+\mathrm{T}\}$. By doing appropriate row operations on the (4×2) empirical formula matrix that defines Δ_1, the same basis is found for $\ker\Delta_1$. Therefore, the factor group $\mathrm{im}\Delta_2/\ker\Delta_1$ is $\{0\}$. So the network is torsion-free.

Example 2 In Example1 replace E by ES:

$$\mathrm{R} + \mathrm{ES} \Leftrightarrow \mathrm{ESS} \Leftrightarrow \mathrm{ES} + \mathrm{T} \tag{4}$$

The resulting expression (4) is ambiguous at the molecular level, because it has not been specified yet whether the molecule that binds to ES in the first step remains bound to E in the second step or becomes T in the second step. Let us resolve this by looking at two molecules of S and labeling them S_1 and S_2. Then let the elementary reactions (4) be rewritten as follows:

$$\begin{array}{c} \overset{U_2}{\Leftrightarrow} \quad \overset{V_2}{\Leftrightarrow} \\ \mathrm{R} + \mathrm{ES}_2 \quad \mathrm{ES}_1\mathrm{S}_2 \quad \mathrm{ES}_1 + \mathrm{T} \\ \underset{U_1}{\Leftrightarrow} \quad \underset{V_1}{\Leftrightarrow} \\ \mathrm{R} + \mathrm{ES}_1 \quad \mathrm{ES}_2\mathrm{S}_1 \quad \mathrm{ES}_2 + \mathrm{T} \end{array}$$

Show that the biochemical reaction network with these elementary reactions has the cyclic group $\mathbf{C}_2$ of order two as its torsion group:

The network is defined by the chain complex (1), in which **E, M** and **A** are the free additive groups whose bases are

$$\{\mathrm{U}_1, \mathrm{V}_1, \mathrm{U}_2, \mathrm{V}_2\}, \{\mathrm{ES}_1, \mathrm{R}, \mathrm{ES}_1\mathrm{S}_2, \mathrm{ES}_2, \mathrm{ES}_2\mathrm{S}_1, \mathrm{T}\} \text{ and } \{\mathrm{E}_0, \mathrm{S}_0\},$$

respectively, with Δ_2 and Δ_1 defined thus:

$$\begin{array}{lll} \Delta_2(\mathrm{U}_1) = -\mathrm{ES}_1 - \mathrm{R} + \mathrm{ES}_2\mathrm{S}_1 & \Delta_1(\mathrm{ES}_1) = \mathrm{E}_0 + \mathrm{S}_0 & \Delta_1(\mathrm{ES}_2) = \mathrm{E}_0 + \mathrm{S}_0 \\ \Delta_2(\mathrm{V}_1) = -\mathrm{ES}_2\mathrm{S}_1 + \mathrm{ES}_2 + \mathrm{T} & \Delta_1\mathrm{R} = \mathrm{S}_0 & \Delta_1(\mathrm{ES}_2\mathrm{S}_1) = \mathrm{E}_0 + 2\mathrm{S}_0 \\ \Delta_2(\mathrm{U}_2) = -\mathrm{ES}_2 - \mathrm{R} + \mathrm{ES}_1\mathrm{S}_2 & \Delta_1(\mathrm{ES}_1\mathrm{S}_2) = \mathrm{E}_0 + 2\mathrm{S}_0 & \Delta_1(\mathrm{T}) = \mathrm{S}_0 \\ \Delta_2(\mathrm{V}_2) = -\mathrm{ES}_1\mathrm{S}_2 + \mathrm{ES}_1 + \mathrm{T} & & \end{array}$$

Δ_1 and Δ_2 in matrix form are as follows:

$$\begin{array}{c} \quad \mathrm{ES}_1 \ \mathrm{ES}_2\mathrm{S}_1 \ \mathrm{ES}_2 \ \mathrm{ES}_1\mathrm{S}_2 \ \mathrm{R} \ \mathrm{T} \\ \Delta_2(\mathrm{U}_1) \begin{pmatrix} -1 & 1 & 0 & 0 & -1 & 0 \end{pmatrix} \\ \Delta_2(\mathrm{V}_1) \begin{pmatrix} 0 & -1 & 1 & 0 & 0 & 1 \end{pmatrix} \\ \Delta_2(\mathrm{U}_2) \begin{pmatrix} 0 & 0 & -1 & 1 & -1 & 0 \end{pmatrix} \\ \Delta_2(\mathrm{V}_2) \begin{pmatrix} 1 & 0 & 0 & -1 & 0 & 1 \end{pmatrix} \end{array} \qquad \begin{array}{cc} & \mathrm{E}_0 \ \mathrm{S}_0 \\ \mathrm{ES}_1 & (1 \ 1) \\ \mathrm{ES}_2\mathrm{S}_1 & (1 \ 2) \\ \mathrm{ES}_2 & (1 \ 1) \\ \mathrm{ES}_1\mathrm{S}_2 & (1 \ 2) \\ \mathrm{R} & (0 \ 1) \\ \mathrm{S} & (0 \ 1) \end{array}$$

These two matrices, the *stoichiometric* and *empirical formula* matrices, define a chemical reaction network in terms of its given basis elements. The fact that the product of the matrices is zero confirms that all reactions in the network are balanced.

By elementary row operations on the stoichiometric matrix a basis for $\mathrm{im}\Delta_2$ is found to be the following:

$$\{(-\mathrm{ES}_1 + \mathrm{ES}_2\mathrm{S}_1 - \mathrm{R}), (-\mathrm{ES}_2\mathrm{S}_1 + \mathrm{ES}_2 + \mathrm{T}), (-\mathrm{ES}_2 + \mathrm{ES}_1\mathrm{S}_2 - \mathrm{R}), (-2\mathrm{R} + 2\mathrm{T})\}$$

Likewise, from the empirical formula matrix a basis is found for $\ker\Delta_1$ that differs from the above only in the last entry, which becomes $(-\mathrm{R}+\mathrm{T})$. So, the factor group $\ker\Delta_1/\mathrm{im}\Delta_2$ is the cyclic group of order two, and it is the torsion group of the network.

The above network is hypothetical, but the following description gives an idea of how it might look in nature. Imagine an R-pool and a T-pool separated by a membrane, which has one or more tubular enzymes attached to and passing through it. Each of these passages is a structure E that allows molecules to pass through it single file in either direction. In the present example it is assumed that the passage is short and always contains at most two molecules and at least one. Suppose R_1 and R_2 are a pair of R-molecules, then it takes at least 4 steps for them to be transformed into a pair of S-molecules, S_1 and S_2. A single R-molecule, say R_2, takes on the five states in the sequence $(\mathrm{R}_2, \mathrm{ES}_2\mathrm{S}_1, \mathrm{ES}_2, \mathrm{ES}_1\mathrm{S}_2, \mathrm{T}_2)$, ending up as T_2.

The next example is based on the idea of an enzyme E shaped like a tube through the membrane that always contains n or $(\mathrm{n}-1)$ molecules of the terminal species as they move through the membrane single file in either direction.

Example 3 The biochemical reaction network with the following 2n elementary reactions has the cyclic group of order n as its torsion group.

$$\begin{array}{lll}
\mathrm{R} + \mathrm{ES}_1\ldots\mathrm{S}_{(n-1)} & \Leftrightarrow \mathrm{ES}_n\mathrm{S}_1\ldots\mathrm{S}_{(n-1)} & \Leftrightarrow \mathrm{ES}_n\mathrm{S}_1\ldots\mathrm{S}(n-2)+\mathrm{T} \\
\mathrm{R} + \mathrm{ES}_n\mathrm{S}_1\ldots S_{(n-2)} & \Leftrightarrow \mathrm{ES}_{(n-1)}\mathrm{S}_n\mathrm{S}_1\ldots\mathrm{S}_{(n-2)} & \Leftrightarrow \mathrm{ES}_{(n-1)}\mathrm{S}_n\mathrm{S}_1\ldots\mathrm{S}_{(n-3)}+\mathrm{T} \\
\mathrm{R} + \mathrm{ES}_{(n-1)}\mathrm{S}_n\mathrm{S}_1\ldots\mathrm{S}_{(n-3)} & \Leftrightarrow \mathrm{ES}_{(n-2)}\ldots\mathrm{S}_n\mathrm{S}_1\ldots\mathrm{S}_{(n-3)} & \Leftrightarrow \mathrm{ES}_{(n-2)}\ldots S_n\mathrm{S}_1\ldots\mathrm{S}_{(n-4)}+\mathrm{T} \\
 & \ldots & \ldots \quad \ldots \\
\mathrm{R} + \mathrm{ES}_2\ldots\mathrm{S}_n & \Leftrightarrow \mathrm{ES}_1\ldots\mathrm{S}_n & \Leftrightarrow \mathrm{ES}_1\ldots\mathrm{S}_{(n-1)}
\end{array}$$

This is a generalization of examples 1 and 2, in which cases n was 1and 2.

Example 4 A biochemical reaction network can be defined by putting two networks of the type in example 3 in tandem. Let the first have 2m elementary steps that take molecules from an R-pool to an intermediate I-pool, and let the second have 2n steps that take molecules from the I-pool to a T-pool. The torsion group of this network is the direct sum of the cyclic groups $\mathbf{C}_\mathrm{m}$ and $\mathbf{C}_\mathrm{n}$ of orders m and n.

This is demonstrated as before by applying appropriate matrix operations to stoichiometric and empirical formula matrices.

2.6 Conclusion

Example 4. tells how two networks can be combined so that the torsion group of the resulting network is the direct sum of the torsion groups of the individual networks. Since every finite additive group is the direct sum of finitely many cyclic groups, it

follows that a plausible biochemical reaction network can be constructed with any desired finite additive group as its torsion group. This fact leaves us with a classification system for biochemical reaction networks by their torsion groups. In particular if the torsion group is $\{0\}$, as is commonly the case in practice, then the network is not ambiguous at the macroscopic level. This case can be predicted without the need to determine the torsion group, as noted at the end of Sect. 2.2.

Acknowledgments I acknowledge with gratitude that the late distinguished chemical engineer John Happel, his collaborator in the areas of mathematics and computer science, Masood Otarod, and my colleague Jon Wagg [11] opened my eyes to the need for a mathematical approach to the problems in their areas of research, which are addressed in this paper.

References

1. L.S. Pontryagin, *Foundations of Combinatorial Topology* (Graylock Press, Rochester, New York, 1952)
2. P.H. Sellers, *Combinatorial Complexes, a Mathematical Theory of Algorithms* (D. Reidel Publishing Company, Dordrecht: Holland, 1997)
3. M.E.J. Newman, The structure and function of complex networks. SIAM Rev. **45**(2), 240 (1998)
4. W. Lederman, *Introduction to the Theory of Finite Groups,* (fourth edition Chapter 6 Interscience Publishers, New York, 1961)
5. D. Angeli, E.O. Sontag, Monotone chemical reaction networks. J. Math. Chem. **41**(3), 295 (2007)
6. P.H. Sellers, Chemical reaction networks, treatise in preparation on the applications of homology in chemistry (2007)
7. P.H. Sellers, Mathematical tools for a reaction database in biology. Graph Theory Notes of New York **XXXV**, 22–31 (1998)
8. P.C. Milner, The possible mechanisms of complex reactions involving consecutive steps. J. Electrochem. Soc. **3**, 228–232 (1964)
9. J. Happel, P.H. Sellers, Analysis of the possible mechanisms for a catalytic reaction system. Adv. Catal. **32**(4), 273–323 (1983)
10. P.H. Sellers, Combinatorial classification of chemical mechanisms. SIAM J. Appl. Math. **44**(4), 784–792 (1984)
11. J. Wagg, P.H. Sellers, Enumeration of flux routes through complex chemical reactions. Biocomputing: Proceedings of the 1997 Symposium, ed. by R.B. Altman, A. Keith Dunker, L. Hunter, T.E. Klein, World Scientific Publishing Co., Singapore (1997)