A Linear-Programming Reformulation of Chemical Stoichiometry and Catalysis

Raúl E. Valdés-Pérez

September 1990

CMU-CS-90-172$_3$

School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

Abstract

We have reformulated the concept of stoichiometry of a reaction mechanism in chemistry in terms of an optimization problem in linear programming.

This reformulation has two advantages. First, the concept has been formalized so that it can be carried out easily by computer program without human intervention. Second, the new formulation relates stoichiometry explicitly to reaction yield, so that the calculated stoichiometry corresponds to the ideal yield of a specified target product. This relation is not made by a previous published characterization of mechanism stoichiometry.

The new formulation has been deployed to formalize also the concept of a species playing a catalytic role in a mechanism. This formalization serves within our automated pathway-elucidation project to check pathway hypotheses for compatibility with evidence of catalysis.

The author is an Air Force Laboratory Graduate Fellow, as administered by the Southeastern Center for Electrical Engineering Education (SCEEE). The views and conclusions contained in this document are those of the author and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. government, or of the SCEEE.

Keywords: linear programming, chemistry, stoichiometry, catalysis.

1. Introduction

The concept of stoichiometry is used in chemistry in different contexts. One use refers to an abstract, balanced transformation of a set of species (reactants) into another set of species (products), characterized well by [Smith and Missen, 1979]. Another use refers to the stoichiometry of a reaction mechanism, and is concerned roughly with a balanced transformation of starting materials into final products *that is implied by the mechanism.* This note examines the concept of stoichiometry only in this sense of a balanced transformation implied by a mechanism.

The stoichiometric coefficients of starting materials are often construed as the ideal proportions of starting materials, in the sense of not providing more initial concentration than needed to obtain, in the idealized case, a certain yield of a target product.

The purpose of this note is to formalize the concept of reaction-mechanism stoichiometry by relating the concept explicitly to ideal yield. The current, semi-formal notion of mechanism stoichiometry, as described by Corio [1989], has an ambiguity due to an uncertain, somewhat arbitrary decision of which mechanism species to regard as intermediates. This ambiguity makes the concept of stoichiometry less valuable for the discussion of ideal mechanism yields. We shall illustrate by example how our formalization in terms of linear programming removes this ambiguity, while gaining a clearer understanding of the relation between stoichiometry and ideal yield.

2. Concept of reaction extent

The formal development of stoichiometry relies on the concept of reaction molar extent, or hereafter simply extent. A reaction extent is a number associated with each mechanistic step that represents the net flow in moles from left-side species to right-side species. A step $A + B = C$ of extent 1.5 means that *via this step* 1.5 moles of A, B are transformed to C.

3. Stoichiometry and reaction extent

One procedure to determine stoichiometry is sketched by [Corio, 1989]. This procedure is often followed implicitly when finding stoichiometry by hand on simple cases. First, one arranges the mechanistic steps in order from starting materials to target product. Next, one decides which species are intermediates, and which are final products, i.e., not intermediates nor starting materials. Then, one selects values for the extents such that the net formation of intermediates across the steps is zero (the extent contributes negatively to the net formation of a step's left-side species, and positively to the net formation of the right-side species). The species of negative net formation are the mechanism's starting materials, and those of positive net formation are the final products. The stoichiometry is derived from these non-zero net formations, adopting their magnitudes as stoichiometric coefficients.

This procedure is illustrated on the following schematic mechanism I, where the e_i above the arrows are the extents.

$$\begin{aligned} A+B &\xrightarrow{e_1} X+2Y \\ B+X &\xrightarrow{e_2} 2Z \\ Y+Z &\xrightarrow{e_3} T \end{aligned}$$

The intermediates are X, Y, Z, and the target product is T. The three equations that enforce zero net formation of the intermediates are these:

$$\begin{aligned} X: \quad e_1 - e_2 &= 0 \\ Y: \quad 2e_1 - e_3 &= 0 \\ Z: \quad 2e_2 - e_3 &= 0 \end{aligned}$$

which possess the solutions $e_1 = e_2 = \alpha$, $e_3 = 2\alpha$. Summing over the steps the species formations determined by the extents, and equating α to 1 for convenience, we obtain the stoichiometry $A + 2B \rightarrow 2T$.

The above semi-formal procedure has some drawbacks, which we illustrate by the following mechanism II, differing from I only in the coefficient of Y in the first step; the mechanism is quite ordinary, e.g., it exhibits no feedback.[1]

$$\begin{aligned} A+B &\rightarrow X+Y \\ B+X &\rightarrow 2Z \\ Y+Z &\rightarrow T \end{aligned}$$

The species X, Y, Z again are intermediates. We derive equations as before to force their net formations to zero:

$$\begin{aligned} X: \quad e_1 - e_2 &= 0 \\ Y: \quad e_1 - e_3 &= 0 \\ Z: \quad 2e_2 - e_3 &= 0 \end{aligned}$$

The only solution to these equations has the extents e_i identically zero, meaning that the intermediates cannot all have zero net formation, unless no reactions occur at all. That is, there exists no ratio of starting materials that yields T with no by-product. Therefore, one of these intermediates will appear as a stoichiometric product, in which case the net formation of the remaining two can be forced to zero. A plausible candidate for status as stoichiometric product is species Z, which appears from the last two steps to be formed in excess. Proceeding in this way, the derived stoichiometry would be $A + 2B \rightarrow T + Z$.

There are two lessons from mechanism II. First, the concept of intermediate as a species formed and later consumed, which is on the path to a target product, cannot be identified with the

[1] It is important to realize that mechanism II is consistent. One set of consistent instances has A, B, X, Y, Z all sharing the same molecular formula, and T double that formula. Whether such instances are empirically plausible is another matter.

set of mechanism species that disappear from the stoichiometry. We saw above that not all the intermediates could be eliminated. Second, the stoichiometry of a mechanism need not be unique, since it depends in some quite ordinary cases on an arbitrary decision of which intermediate shall be regarded as a stoichiometric product. This raises the question of how best to define stoichiometry in a manner compatible with results on non-problematic cases such as mechanism I, while gaining on problematic mechanisms such as II a correspondence to ideal yield.

4. Stoichiometry via minimization of starting materials

Among the possible mechanism stoichiometries, we shall single out that which affords the best yield of a target product. Our formalization of stoichiometry shall take the form of a linear minimization problem. In this section, we formulate the equations, inequalities, and minimization criterion for mechanism II. Then the stoichiometry determined by our procedure is contrasted with the stoichiometry seen above.

The intuitive idea is to select, as before, values of the reaction extents, but without any preconceived decision to force the net formation of certain species to zero; we have no need for the concept of intermediate. In addition, minimal values are selected for the initial concentrations of starting materials, such that there exist reaction extents resulting in an arbitrary yield of 1 mole for the target product.

To illustrate, we write the following equation derived from mechanism II.

$$\begin{pmatrix} A(t) \\ B(t) \\ X(t) \\ Y(t) \\ Z(t) \\ T(t) \end{pmatrix} = \begin{pmatrix} A(0) \\ B(0) \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} -1 & 0 & 0 \\ -1 & -1 & 0 \\ 1 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 2 & -1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} e_1 \\ e_2 \\ e_3 \end{pmatrix} \tag{1}$$

This states that a transition from concentrations at time zero to those at time t is induced by the reaction extents. The matrix pre-multiplying the extents is derived from the stoichiometries of the individual mechanistic steps. The next equation,

$$T(t) = 1 \tag{2}$$

arbitrarily fixes the yield of T at 1 mole. All the other variables receive no fixed value, but are constrained to be non-negative:

$$A(0), A(t), B(0), B(t), X(t), Y(t), Z(t), e_1, e_2, e_3 \geq 0 \tag{3}$$

The objective is to find the minimum amounts of starting materials that yield 1 mole of T via mechanism II. This objective is formalized by the following statement:

Minimize the quantity $A(0) + B(0)$ subject to the constraints of equations 1, 2 and inequalities 3.

We discuss how to solve this minimization problem shortly; its solution is this:

$$A(0) = 1.0,\ B(0) = 1.5,\ X(t) = 0.5,\ A(t) = B(t) = Y(t) = Z(t) = 0$$

Multiplying all values by 2, we obtain a stoichiometry of

$$2A + 3B \rightarrow 2T + X$$

which improves on the yield of T implied by the previous stoichiometry $A + 2B \rightarrow T + Z$, derived by arbitrarily forcing the net formations of intermediates X, Y to zero.

Following the same procedure, we calculate for the unproblematic mechanism I a stoichiometry of $A + 2B \rightarrow 2T$, the same as obtained before with the usual procedure.

5. The general case and its solution

In section 4 we formulated a specific minimization problem to obtain a stoichiometry for mechanism II. The general case is treated here.

Given a mechanism, we derive from it a matrix $R_{r \times s}$ of r rows, one for each mechanistic step, and s columns, one for each species. The entries of a row are the stoichiometric coefficients of an individual step, using negative coefficients for the left-side species, positive coefficients for the right-side species, and zero coefficients for any species not appearing in the step. If a species appears both on the left and right sides, then its row entry is simply the signed sum of the two coefficients.

$$c(t)_{s \times 1} = c(0)_{s \times 1} + (R^T)_{s \times r}\, E_{r \times 1}$$

This equation says that the species concentrations at time t equal the concentrations at time 0 plus the contributions to the concentrations from the reaction extents, represented by the vector E, having r rows, one for each step. This equation is the general case of equation 1 above.

The other constraints needed are equation 2 and a generalization of the inequalities 3:

1. The target product has concentration of 1 mole (or any convenient value) at time t.
2. All variables (i.e., species concentrations, extents) are non-negative.

The generalized minimization problem is that of minimizing at $t = 0$ the sum of the concentrations of all starting materials.

There is one subtlety in the procedure of this section. If any of the mechanistic steps is reversible, one should include the reversed step explicitly in the mechanism. Otherwise, since extents are constrained to be non-negative, a net reaction in the reverse direction would be precluded.

6. Solving the minimization problem

The above minimization problem is an instance of *linear programming.* A linear programming (LP) problem consists partly of a set of constraints, which can be linear equations or linear inequalities. By convention, all variables appearing in the constraints assume only non-negative values. The remaining part of an LP problem is the minimization criterion, required to be a linear expression (e.g., $A(0) + B(0)$ in the example above).

An LP problem is solved by reporting one of the following: 1. that the constraints are inconsistent; 2. that the quantity to be minimized is unbounded; or 3. values for all variables, as well as the value of the expression to be minimized. [Papadimitriou and Steiglitz, 1982] have a detailed discussion of linear programming.

These reports are interpreted in our case as follows.

1. Inconsistent constraints mean that the mechanism is inconsistent, in one of two ways. If the mechanism is schematic, then inconsistency means that no assignment of molecular formulas to the variables can result in balanced steps. If the mechanism is a specific instance, then inconsistency means it is unbalanced.

2. The quantity that we minimize *cannot* be unbounded, because concentrations cannot be negative. Hence the lower bound on the summed quantity of starting materials is zero.

3. If the minimization succeeds, then a stoichiometry maximizing the yield of target product exists.

We have chosen a minimization criterion rather arbitrarily. In general, the criterion can be any suitable, weighted sum of the starting materials, i.e., $\alpha A + \beta B + \gamma C \ldots$, of which our choice of unit weights is a special case. The weights might reflect relative costs, for example. In general, the resulting stoichiometry depends on the weights. For example, a zero weight for starting material B of mechanism II would result in the otherwise inferior stoichiometry of $A + 2B \rightarrow T + Z$). We return in section 8 to the issue of weights in the context of catalysis.

Computer programs to solve LP problems are readily available; the most common are based on the Simplex algorithm. [Press et al., 1986] list a program for one implementation of Simplex. We have used this program as a foundation for another program that finds stoichiometry from the symbolic inputs of mechanism, starting materials, and target product.

6.1. Comparison with usual procedure

The remaining comments of this section shall assume good familiarity with the Simplex algorithm. Below, we let $\#sp$ be the number of mechanistic species, $\#st$ the number of mechanistic steps, and $\#sm$ the number of starting materials.

We have already remarked that the usual method to find stoichiometry consists of selecting certain

intermediates X_i, and forcing their net formations $X_i(t)$ to zero. The Simplex-based procedure is a generalization of the same idea, as will be seen shortly.

The linear-programming formulation builds a constraints matrix of rank $\#sp + 1$, since there is a constraint for each species (as in equation 1), plus the constraint $T(t) = 1$ that fixes the yield. The number of variables *nvars* in the constraints matrix equals $\#sp + \#st + \#sm$, corresponding to the species concentrations at time t, the reaction extents, and the initial concentrations of starting materials.

The Simplex algorithm in effect optimizes the objective function over all of these cases: find a subset of the *nvars* variables of size $nvars - rank$, force the variables in the subset to zero, and calculate the values of the non-subset variables (some of these may turn out zero also). Hence, any solution to the linear program has at least

$$nvars - rank = (\#sp + \#st + \#sm) - (\#sp + 1) = \#st + \#sm - 1$$

zeroed variables. If we assume that for the calculated stoichiometry the final concentrations of the $\#sm$ starting materials are zero, and that the reaction extents are positive (all mechanistic steps have reacted to some degree), then we obtain that at least $\#st - 1$ (non-starting-material) species have final values of zero, i.e., are *stoichiometric intermediates.*

We can illustrate this result with a mechanism consisting of steps $X_k \rightarrow X_{k+1}$, $k = 1, \ldots, N$. The stoichiometry is $X_1 \rightarrow X_{N+1}$, the final concentration of X_1 is zero, and the extents are identically 1. The number of *stoichiometric intermediates* is seen to be $N - 1$, one less than the number of steps, consistent with the formula $\#st - 1$.

7. Empirical Interpretation

The stoichiometry found by the LP formulation is an idealized input/output relation between starting materials and target product. The reaction extents are freely chosen to optimize this relation, although physically a choice of extents might correspond to "freezing" a reaction step while its reactants are present. In addition, empirical reaction yields depend on energetic and kinetic considerations. Hence, it may not be feasible actually to achieve the ideal yield determined by the stoichiometry.

Another point is that a certain species might not appear at all in the stoichiometry, although the reaction could not proceed without it. An example is a species that is consumed by an initial step, but is re-generated completely by a later step. Hence, the stoichiometric reactants may be only a subset of the necessary starting materials.

These empirical issues are quite separate from the concept of mechanism stoichiometry, which it is our purpose here to clarify.

8. Applications to Catalysis

The linear-programming reformulation of chemical stoichiometry can easily be deployed to define formally the notion of 'catalyst.'

One type of catalyst accelerates the progress of a reaction, but is not regarded as participating chemically in the reaction. Another type of catalyst includes species that do react chemically, but are re-generated later in the mechanism in amounts equal to the amount reacted. With regard to this latter meaning of catalyst, this section shall develop a formal answer to the following question: Is a given mechanism compatible with a catalytic role for a particular starting material?

To answer this question, first we observe that if a mechanism stoichiometry omits a starting material B, then the mechanism can account for a catalytic role for B.[2] So the question is reformulated as: Is there a mechanism stoichiometry having a zero coefficient for the starting material?

This new question is answered simply by changing slightly the objective function (shown on page 3) of the linear program. Rather than minimize the sum of initial concentrations of all starting materials $\sum_{i=1}^{\#sm} SM_i(0)$, instead we minimize the sum

$$\sum_{i=1}^{\#sm} SM_i(0) + c \times SM_k(0), \quad c \gg 1$$

where SM_k is the species whose possible catalytic role is in question. This formulation has the effect of minimizing most crucially $SM_k(0)$, so that if a stoichiometry corresponding to $SM_k(0) = 0$ exists, it will be chosen.

We shall illustrate with two examples derived from the cyclic pathway for urea synthesis discovered by Hans Krebs [Holmes, 1980, Kulkarni and Simon, 1988], for which the starting materials are *ornithine*, NH_3, and CO_2 (a schematic version of the mechanism appears on the right):

$$\begin{aligned} \textit{ornithine} + NH_3 + CO_2 &\rightleftharpoons \textit{water} + C_6H_{13}N_3O_3 \\ NH_3 + C_6H_{13}N_3O_3 &\rightleftharpoons \textit{arginine} + \textit{water} \\ \textit{arginine} + \textit{water} &\rightleftharpoons \textit{ornithine} + \textit{urea} \end{aligned} \qquad \begin{aligned} A + B + C &\rightleftharpoons Y + M_1 \\ B + M_1 &\rightleftharpoons Z + Y \\ Z + Y &\rightleftharpoons A + T \end{aligned}$$

Our linear programming formulation, with a large multiplier for the initial concentration of ornithine, yields the stoichiometry

$$2(NH_3) + CO_2 \rightarrow \textit{urea} + \textit{water}$$

which means that the pathway is compatible with the observed catalytic action of ornithine (the "ornithine effect"). Clearly the formation of ornithine in the third step suggests the possibility of complete re-generation.

---

[2] Whether the starting material in fact behaves catalytically depends on more details, such as the reaction kinetics. We are treating the compatibility question assuming only a mechanism, with no information on reaction speeds.

A more subtle example is the following, alternative explanation of the ornithine effect, found by a pathway generation algorithm [Valdes-Perez, 1990]:

$$\begin{array}{rcl} ornithine + CO_2 & \rightleftharpoons & water + C_6H_{10}N_2O_3 \\ 2(NH_3) + C_6H_{10}N_2O_3 & \rightleftharpoons & arginine + water \\ CO_2 + arginine & \rightleftharpoons & C_6H_{10}N_2O_3 + urea \end{array} \qquad \begin{array}{rcl} A + C & \rightleftharpoons & Y + M_2 \\ 2B + M_2 & \rightleftharpoons & Z + Y \\ C + Z & \rightleftharpoons & M_2 + T \end{array}$$

In this case, ornithine does not appear on the right side of any step. Nevertheless, a linear programming formulation with the same objective function as the previous example yields the same stoichiometry as before: $2(NH_3) + CO_2 \rightarrow urea + water$.

This surprising result is explained as follows. "Summing" the three steps, we obtain the net reaction

$$ornithine + 2(NH_3) + 2(CO_2) \rightarrow urea + 2(water) + C_6H_{10}N_2O_3.$$

If the first step then reacts in the reverse direction, the resulting net reaction is $2(NH_3) + CO_2 \rightarrow urea + water$, which is the derived stoichiometry. This stoichiometry implies a *net* extent of zero across the first step. [3]

We are using this formalization of catalysis in a project for automated pathway-elucidation to rule out generated mechanisms (or pathways) that are incompatible with a presumed catalytic role for a starting material.

9. Conclusion

We have clarified the concept of mechanism stoichiometry, and its relation to ideal yield, by formalizing the derivation of stoichiometry. A derived stoichiometry corresponds to the ideal yield of a target product, in the sense of minimizing the summed quantity of starting materials. The formalization is in terms of linear algebraic equations and inequalities, using the concept of reaction extent, but without using the concept of reaction intermediate. Finding the stoichiometry involves solving a simple, linear-programming problem.

The usual method of finding stoichiometry, as described by [Corio, 1989], contains an ambiguity due to its reliance on the concept of reaction intermediate. Our example above has illustrated that one cannot identify the mechanistic intermediates with the set of species not appearing stoichiometrically. Hence, an arbitrary decision is usually made regarding which species to omit from the stoichiometry. A wrong decision, from the viewpoint of ideal yield, will lead to an inferior stoichiometry.

This new formulation in terms of linear programming also permits a new formalization of the concept of a species having a catalytic role, which has seen application in our project on automated pathway elucidation.

[3] We emphasize that whether the mechanism kinetics will allow such a net reaction is another issue.

Acknowledgment. Prof. Herbert A. Simon commented valuably on a draft of this report.

References

[Corio, 1989] Corio, P. L. (1989). Theory of reaction mechanisms. In *Topics in Current Chemistry: Relationships and Mechanisms in the Periodic Table*, volume 150. Springer-Verlag.

[Holmes, 1980] Holmes, F. L. (1980). Hans krebs and the discovery of the ornithine cycle. In *Proc. of 63rd Annual Meeting of Federation of American Societies for Experimental Biology.* Symposium on Aspects of the History of Biochemistry.

[Kulkarni and Simon, 1988] Kulkarni, D. and Simon, H. (1988). The processes of scientific discovery: The strategy of experimentation. *Cognitive Science*, 12:139–175.

[Papadimitriou and Steiglitz, 1982] Papadimitriou, C. and Steiglitz, K. (1982). *Combinatorial Optimization: Algorithms and Complexity.* Prentice-Hall.

[Press et al., 1986] Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. (1986). *Numerical Recipes: The Art of Scientific Computing.* Cambridge University Press.

[Smith and Missen, 1979] Smith, W. R. and Missen, R. W. (1979). What is chemical stoichiometry? *Chemical Engineering Education*, 13(1):26–32.

[Valdes-Perez, 1990] Valdes-Perez, R. E. (1990). *Machine Discovery of Chemical Reaction Pathways and their Experimental Confirmation.* PhD thesis, Carnegie Mellon University. (School of Computer Science, forthcoming).