

Section: Chemistry

DATA INPUT ORGANIZATION FOR STOICHIOMETRIC CALCULATIONS

Yilmaz Nataliia

PhD, scientist EPFL

the Swiss Federal Institute of Technology in Lausanne, Switzerland

Kozub Pavlo

PhD (Technical Sciences), Associate Professor

Kharkiv National University of Radio Electronics, Ukraine

Kozub Svetlana

PhD (Technical Sciences), Associate Professor

Kharkiv National Medical University, Ukraine

Balancing chemical reactions is a fundamental and integral part of stoichiometric calculations in chemistry. Any further calculations inevitably require this step. To many, it appears to be merely a mathematical problem solvable by various computational methods, often perceived as loosely related to the chemical nature of the task, and requiring only mathematical software packages.

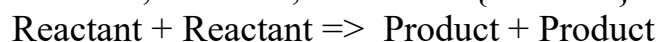
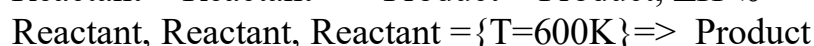
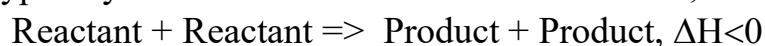
In fact, methods of data structuring and presentation for such calculations are equally important. Effective human-computer interaction should maintain the traditional form of chemical equations and transformation schemes. To ensure accurate and convenient research management, it is essential to use a unified, user-friendly, and information-rich interface for presenting data both to human researchers and software algorithms.

The rules and algorithms of such an interface must preserve conventional representations (formulas, reaction schemes) while accommodating limitations in input environments – such as lack of formatting in browsers and mobile apps.

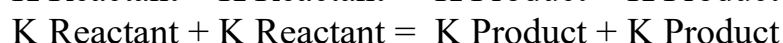
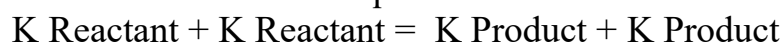
The system should support flexible output formatting under clear rules, and be extensible for future capabilities.

The result of data processing should be a mathematical structure optimized for computation, with well-defined data organization, full preservation of the original chemical information, and the ability to reconstruct the input in its original form.

A review of literature and online resources shows that chemical transformations are typically described as lists of reaction schemes, for example:



or as lists of chemical equations:



K Reactant + K Reactant = K Product + K Product

In many cases—especially in organic chemistry—transformation schemes use multiple transformation symbols, indicating that some substances act as both reactants and products across several stages. For example:

Reagent => Reagent + Reagent => Reagent, Reagent => Reagent

Which is a collapsed form of a more expanded sequence:

Reagent => Reagent + Reagent

Reagent + Reagent => Reagent, Reagent

Reagent, Reagent => Reagent

In these notations, a "Reagent" is a reactant if placed left of the delimiter (=, =>, etc.) and a product if placed on the right. Thus, a transformation path in symbolic form consists of listed reagents separated by special delimiters.

- **Transformation symbol** (=, =>, <=>, ...) separates groups of reagents. "=" is used for chemical reactions; others may be used in transformation schemes.

- **Reagent delimiter** (+ or ,) separates reagents within each group. "+" is used for defined reactions, while "," often indicates uncertainty in quantity.

In chemical equations, stoichiometric coefficients are required. They may be omitted if equal to 1, but must be non-negative rational or fractional values.

Extended equations may include:

- transformation energy, volume/mass changes
- reaction conditions (above/below the delimiter)
- reaction direction (arrows, crossed equals signs)

Though reagents may be referred to by trivial names (especially in organic chemistry), they must be reduced to a formal chemical formula for balancing—where elemental composition, ion charge, and functional groups are computable.

The simplest formula type is the **gross formula**: a string that lists elementary components (elements, radicals) with their quantities, e.g., CO₂, N₂O₃, H₂SO₄, FeO_{0.8}, NO_x.

One challenge is that the quantity is omitted if equal to 1, and may be fractional or undefined, complicating automated parsing.

Conventionally, subscript is used for quantities, but plain text lacks formatting—so quantities must follow the element name directly.

Groupings are indicated with parentheses or multiplication signs: – Na₂CO₃·10H₂O, CuSO₄*5H₂O.

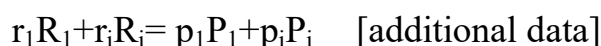
Ion charges and oxidation states use superscripts (e.g., Na⁺, Ca²⁺, NO₃⁻, SO₃²⁻, V⁺⁵₂O⁻²₅).

In plain text, these conflict with subscripts, so formats like NO₃⁻¹ or NO₃Z-1, Z-1NO₃ may be used to distinguish charge.

Modifiers and attributes (radical notation, phase state, crystal type, excitation state) are inconsistently standardized. Common suffixes include: (l), (g), (s), (aq), (cr), etc.

To ensure consistent parsing, a system of rules has been defined to decode such symbolic input reliably.

The standard reaction format:

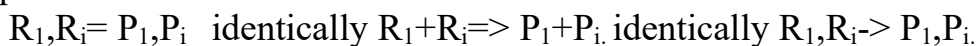


$R_i P_i$ – Reagent formulas (text strings)

$r_i p_i$ – Stoichiometric coefficients (positive rational numbers; defaults to 1 if omitted; 0 if missing from one side).

An equation without coefficients may indicate imbalance.

Symbols like "+" or "," and transformation signs like "<=>", "-->", or "<-" are accepted variations:



If no product side is provided, the assumption is that all species may act as both reactants and products:



Gross formula format: string representation of composition as:



Where E_i – is an elemental symbol,

a, b, n and i – subscripts in the formula that reflect the amounts of elements in the reactant.

Missing elements are assigned zero counts for uniform vector structure. Restricting element names to Latin letters with an uppercase initial helps with parsing.

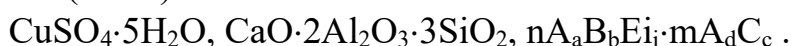
Unlike stoichiometric coefficients, element counts may be negative (e.g., ions, radicals). Example representations:

The charge of particles in the gross formula is conveniently represented as a fraction (for example, Z) with a sign value: $A_aB_bZ_{+z}$ (ions), $C_6H_6H_{-1}$ (phenyl).

For the user, in text format, the charge can be conveniently represented as $A_aB_b^{-z}$ or $A_aB_b\{-z\}$.

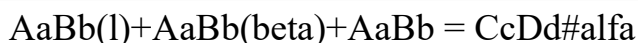
For convenience, scientists use a collapsed formula - a formula in which repeating particles are written in parentheses: $(A_aB_b)_i(A_aB_b)_j$. The type of brackets is determined solely by user convenience, so brackets $[]$ are identical to brackets $()$. So in inorganic chemistry you can very often see $K_3[Fe(CN)_6]$, and in organic $CH_3(CH_2)_{17}COOH$.

Full formulas consist of condensed groups with coefficients and connection symbols (\cdot or $*$):



This notation is common for writing the formulas of crystal hydrates, mixed salts, and oxides.

Suffix metadata does not affect stoichiometric calculations but can be used to identify specific features of the reagent, such as its physical state: (l), (g), (s), (aq), (a), (i), (vit), (cr), etc., type of crystal lattice, method of synthesis, or purity. This information always follows the final index in the formula and does not contain elemental symbols. However, it is currently not standardized and may include arbitrary characters.



With this approach, the decoded structure provides a sufficient basis for both mathematical data processing and restoring the textual chemical representation into a human-readable form.

The resulting decoded data is represented as a multi-dimensional nested array with the following structure:

Array of chemical transformation schemes (size = number of reactions):

[0] – Additional metadata for the transformation list

[1]-[n] – Decoded data for each individual scheme

This array allows for an unambiguous description of the sequence of chemical transformations and includes both detailed information on each reaction and shared metadata in the [0] element.

[i] Individual transformation scheme (size = number of reagents):

[0] – Additional data for the transformation

[1]-[n] – Decoded data for each reagent involved

This structure enables precise computation of material balances for each transformation and contains specific metadata for the transformation in the [0] position.

[i] Reagents array (size = number of elements):

[0] – Additional data about the reagent

[1]-[n] – Quantities of each element

This array allows for the accurate calculation of the behavior and properties of each reagent, the status of each transformation, and the characteristics of the overall reaction mixture, as each [i] element holds the elemental counts, and [0] holds metadata for the reagent as a whole.

References

1. M. M. Shaikh, M. Yousaf On mathematical methods to balance equations of chemical reactions – a comparison and way forward // Journal of mechanics of continua and mathematical sciences. Vol.-18, No.-01, January (2023) pp 1-20 <https://doi.org/10.26782/jmcms.2023.01.00001>
2. J. Aleksejeva, S. Guseynov To the issue of finding the stoichiometric coefficients in the chemical reactions // Integration. education Proceedings of the International Scientific Conference. Volume II, May 28th-29th, 2021. pp. 19-48 // <https://doi.org/10.17770/sie2021vol2.6457>
3. Козуб П. А., Мірошніченко Н. М., Лук'янова В. А., Козуб С. М., Мігунов В. Л. Використання векторного підходу для балансування хімічних рівнянь // Modern directions of scientific research development. Proceedings of the 15th International scientific and practical conference. BoScience Publisher. Chicago, USA. 2022. Pp. 65-73.
4. Козуб П. А., Мірошніченко Н. М., Лук'янова В. А., Козуб С. М., Гуріна Г. І. Використання векторного підходу для задач хімічної стехіометрії // Modern scientific research: achievements, innovations and development prospects.

Proceedings of the 15th International scientific and practical conference. MDPC Publishing. Berlin, Germany. 2022. Pp. 80-87.

5. Козуб П. А., Мірошніченко Н. М., Лук'янова В. А., Козуб С. М., Мартинюк М. М. Математичні аспекти використання векторного підходу для балансування хімічних реакцій. // Modern science: innovations and prospects. Proceedings of the 12th International scientific and practical conference. SSPG Publish. Stockholm, Sweden. 2022. Pp. 65-74.