

Solving Incomplete Inorganic Chemical Systems through a Fuzzy Knowledge Frame

I. Luque Ruiz,* C. Martínez Pedrajas, and M. A. Gómez-Nieto†

Department of Computing and Numerical Analysis, University of Córdoba, Campus Universitario de Rabanales, Edificio C2, Planta-3, E-14071 Córdoba, Spain

Received January 30, 2000

A model for the completion and balancing of inconsistent inorganic reactions is presented. A series of fuzzy parameters is proposed. These parameters are considered within a knowledge frame representing inorganic reactions via a semantic/functional network and, through the calculation of possibility measures, allow the completion and solution of such reactions by considering new inorganic species to be added to the reaction. Species to be added are formulated on the basis of atoms present and/or absent in the reaction, along with valence, system features, and a series of flags that determines the cardinality of the possible solution set. Incorporation of the proposed model into an inorganic chemistry formulation system would contribute to the development of powerful computer-assisted learning systems, given the extent of information extracted from a reaction.

1. INTRODUCTION

The balancing of inorganic reactions is a relatively simple process using analytical techniques, and has attracted much attention in articles aimed at the teaching of this discipline. These studies have examined the advantages of mathematical methods and software for balancing complex reactions, while sometimes proposing techniques and models for ensuring the success of the process by considering new equations obtained from analysis of the original reaction.^{1–6}

The importance of balancing reactions lies in the fact that such reactions will be involved in any and all inorganic chemistry problems and will thus feature in any software aimed at assisting and/or teaching the user.

In a previous study the authors proposed a model for integrating the balancing process into a formulation system such that by employing complex data structures the user may be provided with much more substantial information than just the mathematical equations representing the laws of balancing matter, charge, and valence.⁶

However, in the development of powerful software systems it is not enough that these be able to solve just simple reactions. They should also be capable of solving inconsistent reactions—those reactions which cannot be balanced without the introduction of new inorganic species into the equation.

Over the past 20 years or so, several models and systems have been proposed, particularly in organic chemistry, to assist in synthesis. These studies have been centered on proposing possible reactant sets for the synthesis of a known product.^{7–15} Such processes are mainly based on measures of similarity, thus ensuring that the simplest possible reaction will be found, and use huge databases of chemical species and standard organic reactions (hydrogenation, addition, decomposition, etc.) and/or algorithms for structural breakdown and matching.^{15–20}

Inorganic reactions, however, are vastly different from organic ones. The former generally display greater atomic diversity and, except for certain families of inorganic compounds such as those of silicon or boron, do not normally involve problems concerning the synthesis of particular species.

If we focus on assisted learning systems, whether tutorials be classical or “smart”, such as in virtual laboratories, the problem of manipulating inorganic reactions is clearly different, since here we are seeking to solve a reaction where there is matter imbalance.

Accordingly, the task in hand involves the proposal of inorganic species to ensure the consistency of the reaction, which can then be balanced and solved. This process may be performed in different ways:

1. using large databases storing known reactions and matching the problem system with entries in the database in order to obtain the exact or most probable solution
2. employing a knowledge frame that, taking into account the features of the problem system and being aware of the basic principles of inorganic formulation, can formulate species which, when added to the reaction, will ensure its consistency

Most previous studies have opted for the former approach, relying on high computational power to build a sufficiently complete system. Where such power is not available, however, such systems (usually for assisted learning) are limited to the consideration of a small, finite set of problem systems and will not, therefore, allow an acceptable degree of user intervention either in the proposal of new reactions or in the manipulation of existing ones.^{21–23}

The aim of the present study is to overcome these disadvantages by removing such limitations, thus affording the user free intervention at the same time as obviating the need for major computational power.

To achieve this, our approach is focused on the following points:

* Corresponding author. Phone: +34 957 212082. FAX: +34 957 218630. E-mail: mallurui@uco.es.

† E-mail: mangel@uco.es.

1. the proposal of an open model that may be integrated into any system, regardless of its objectives
2. integration of the model with formulator systems that allow the formulation of new inorganic species without the need for user intervention
3. avoidance of the use of databases storing known reactions to be used for proposing a solution; instead, the model should only work with the information needed for the formulation of inorganic species, information extracted from the problem system, and a series of basic rules for inorganic reactions
4. the proposal and use of parameters to allow the ordering of the set of solutions by a fuzzy measure of possibility, taking into account the complexity of the species taking part in the solution
5. the use of data structures capable of suitably representing not only the species but also the problem system and the actual solving process, in such a way that users may be provided with a great deal of useful information output

Bearing the above in mind, the present article describes the model proposed for the solution of both complete and incomplete inorganic reactions. The model is based on structural disparity and molecular feasibility measures, as described in section 2. These measures reflect the atomic diversity and molecular complexity of inorganic species, while ignoring their physicochemical properties. The cardinality and diversity of atoms and subcompounds present in the inorganic species are used to order the proposed solutions according to their measure of possibility, so as to prioritize those solutions that will introduce the least complexity into the reaction.

Section 3 describes the proposed knowledge frame on which the model is based. This semantic/functional network will, on analysis of the automatically extracted features of the problem reaction, assign it a fixed category, and a series of linked procedures is called to produce a set of possible solutions.

Within this framework, the process of solving a problem reaction involves passing through a series of states, each reached by satisfying a set of conditions dealing with reaction features and calling a series of procedures to carry the reaction through to the next state, until eventually a consistent, solvable state is reached.

Finally, section 4 includes a selection of example reactions to illustrate the working of the model; the article concludes with a discussion and briefly describes current studies being carried out for improvement and completion of the model.

2. STRUCTURAL DISPARITY

A reaction whose two sides do not contain the same types of atomic elements is said to be inconsistent with regard to matter. Determination of the elements required on one or both sides to produce a matter-consistent equation is very simple, but this is not the case for the chemical species, be they ionic or neutral, in which these elements are to be found.

Indeed, the proposal of chemical species to be considered in order to make an incomplete reaction consistent is a complex process involving consideration of the physicochemical features and properties of the reaction, that is, of the chemical species present and their components (subcompounds and atoms).

If this process is to be performed automatically in software, a database is required for the storage of information about chemical reactions and the physicochemical properties of known species, such that on accessing this information a solution may be proposed for an incomplete reaction, either by finding the most similar known reaction or through reference to the stability properties and frequency of the most probable chemical species.

Use of a database of this nature not only involves expensive development and high computational costs for small software systems run on personal computers, but also fails to guarantee the correctness of any proposed solution. An alternative to the above solution is to consider nothing but the information contained within the actual structure of a chemical species, derived directly from its formula, and then using these data to propose different solutions for the reaction, based on possibility measures.

The proposed model includes the following considerations:

Definition 1: Let the disparity of a chemical species, be it an element, ion, radical, or molecule, be a measure of its atomic multiplicity and diversity.

Postulate 1: The disparity of a chemical element in its atomic state is 1.

Postulate 2: The lower the disparity value, the greater the chance of a chemical species being present in a reaction.

The possibility of atoms combining to form more complex structures is inversely related to the disparity value of such structures; in other words, the greater the number and diversity of atoms in a chemical structure, the lower the possibility of such a structure being present in a reaction.

The above principles form the basis of the present model, meaning that the solutions most likely to be produced by the model will be those formed by the most structurally simple species, even though in certain cases these may not be the most correct, chemically.

According to this model, the most likely solution for a matter-inconsistent reaction will always be one that includes chemical species in their native state, but such a solution will often be incorrect due to violation of the laws of chemical combination.

Definition 2: The disparity of an ion or radical, called its *structural disparity (SD)*, is obtained from the summation of the product of the multiplicity of a component element and its presence index, divided by the total number of component elements, as shown in the following expression:

$$SD = \frac{\sum_{i=1}^{i=m} i \min(O_i)}{n}$$

where: n is the total number of atoms present in the species, m is the number of different atoms, being at all times $m \leq n$, and $\min(O_i)$ is the minimum value for the occurrence of the element i in the chemical species.

It may be seen that

- SD will take on values ranging from 1 to the number of atoms in the chemical species plus 1, divided by 2.
- For an identical number of atoms, structural disparity will increase as their own diversity rises.
- For an identical number of atoms and the same diversity, SD rises with the difference or proportion of their occurrence.

Table 1. Measures of Structural Disparity for Species of up to Seven Atoms

elements		species	structural disparity $[1, (N + 1)/2]$	SD
total	different			
1	1	X	$(1 \times 1)/1 = 1/1$	1.0000
2	1	X ₂	$(1 \times 2)/2 = 2/2$	1.0000
2	2	XY	$(1 \times 1 + 2 \times 1)/2 = 3/2$	1.5000
3	1	X ₃	$(1 \times 3)/3 = 3/3$	1.0000
3	2	X ₂ Y	$(1 \times 1 + 2 \times 2)/3 = 5/3$	1.6667
3	3	XYZ	$(1 \times 1 + 2 \times 1 + 3 \times 1)/3 = 6/3$	2.0000
4	1	X ₄	$(1 \times 4)/4 = 4/4$	1.0000
4	2	X ₂ Y ₂	$(1 \times 2 + 2 \times 2)/4 = 6/4$	1.5000
4	2	X ₃ Y	$(1 \times 1 + 2 \times 3)/4 = 7/4$	1.7500
4	3	X ₂ YZ	$(1 \times 1 + 2 \times 1 + 3 \times 2)/4 = 9/4$	2.2500
4	4	XYZW	$(1 \times 1 + 2 \times 1 + 3 \times 1 + 4 \times 1)/4 = 10/4$	2.5000
5	1	X ₅	$(1 \times 5)/5 = 5/5$	1.0000
5	2	X ₃ Y ₂	$(1 \times 2 + 2 \times 3)/5 = 8/5$	1.6000
5	2	X ₄ Y	$(1 \times 1 + 2 \times 4)/5 = 9/5$	1.8000
5	3	X ₂ Y ₂ Z	$(1 \times 1 + 2 \times 2 + 3 \times 2)/5 = 11/5$	2.2000
5	3	X ₃ YZ	$(1 \times 1 + 2 \times 1 + 3 \times 3)/5 = 12/5$	2.4000
5	4	X ₂ YZW	$(1 \times 1 + 2 \times 1 + 3 \times 1 + 4 \times 2)/5 = 14/5$	2.8000
5	5	XYZWV	$(1 \times 1 + 2 \times 1 + 3 \times 1 + 4 \times 1 + 5 \times 1)/5$	3.0000
6	1	X ₆	$(1 \times 6)/6 = 6/6$	1.0000
6	2	X ₃ Y ₃	$(1 \times 3 + 2 \times 3)/6 = 9/6$	1.5000
6	2	X ₄ Y ₂	$(1 \times 2 + 2 \times 4)/6 = 10/6$	1.6667
6	2	X ₅ Y	$(1 \times 1 + 2 \times 5)/6 = 11/6$	1.8333
6	3	X ₂ Y ₂ Z ₂	$(1 \times 2 + 2 \times 2 + 3 \times 2)/6 = 12/6$	2.0000
6	3	X ₃ Y ₂ Z	$(1 \times 1 + 2 \times 2 + 3 \times 3)/6 = 14/6$	2.3333
6	3	X ₄ YZ	$(1 \times 1 + 2 \times 1 + 3 \times 4)/6 = 15/6$	2.5000
6	4	X ₂ Y ₂ ZW	$(1 \times 1 + 2 \times 1 + 3 \times 2 + 4 \times 2)/6 = 17/6$	2.8333
6	4	X ₃ YZW	$(1 \times 1 + 2 \times 1 + 3 \times 1 + 4 \times 3)/6 = 18/6$	3.0000
6	5	X ₂ YZWV	$(1 \times 1 + 2 \times 1 + 3 \times 1 + 4 \times 1 + 5 \times 2)/6 = 20/6$	3.3333
6	6	XYZWVT	$(1 \times 1 + 2 \times 1 + 3 \times 1 + 4 \times 1 + 5 \times 1 + 6 \times 1)/6 = 21/6$	3.5000
7	1	X ₇	$(1 \times 7)/7 = 7/7$	1.0000
7	2	X ₄ Y ₃	$(1 \times 3 + 2 \times 4)/7 = 11/7$	1.5714
7	2	X ₅ Y ₂	$(1 \times 2 + 2 \times 5)/7 = 12/7$	1.7143
7	2	X ₆ Y	$(1 \times 1 + 2 \times 6)/7 = 13/7$	1.8571
7	3	X ₃ Y ₂ Z ₂	$(1 \times 2 + 2 \times 2 + 3 \times 3)/7 = 15/7$	2.1429
7	3	X ₃ Y ₃ Z	$(1 \times 1 + 2 \times 3 + 3 \times 3)/7 = 16/7$	2.2857
7	3	X ₄ Y ₂ Z	$(1 \times 1 + 2 \times 2 + 3 \times 4)/7 = 17/7$	2.4286
7	3	X ₅ YZ	$(1 \times 1 + 2 \times 1 + 3 \times 5)/7 = 18/7$	2.5714
7	4	X ₂ Y ₂ Z ₂ W	$(1 \times 1 + 2 \times 2 + 3 \times 2 + 4 \times 2)/7 = 19/7$	2.7143
7	4	X ₃ Y ₂ ZW	$(1 \times 1 + 2 \times 1 + 3 \times 2 + 4 \times 3)/7 = 21/7$	3.0000
7	4	X ₄ YZW	$(1 \times 1 + 2 \times 1 + 3 \times 1 + 4 \times 4)/7 = 22/7$	3.1429
7	5	X ₂ Y ₂ ZWV	$(1 \times 1 + 2 \times 1 + 3 \times 1 + 4 \times 2 + 5 \times 2)/7 = 24/7$	3.4286
7	5	X ₃ YZWV	$(1 \times 1 + 2 \times 1 + 3 \times 1 + 4 \times 1 + 5 \times 3)/7 = 25/7$	3.5714
7	6	X ₂ YZWVT	$(1 \times 1 + 2 \times 1 + 3 \times 1 + 4 \times 1 + 5 \times 1 + 6 \times 2)/7 = 27/7$	3.8571
7	7	XYZWVTS	$(1 \times 1 + 2 \times 1 + 3 \times 1 + 4 \times 1 + 5 \times 1 + 6 \times 1 + 7 \times 1)/7 = 28/7$	4.0000

• Structural disparity does not take into account the charge of chemical elements.

• Structural disparity intrinsically takes into account the coordination valences of chemical elements, since this is responsible for the multiplicity with which two or more elements combine.

• SD is a suitable measure of the complexity of subcompounds or ions forming molecular species.

Table 1 shows the calculation of SD for species of up to seven atoms. It can be seen that structural disparity is independent of the relationships between components or elements of the species. In fact, this calculation supposes the existence of one or more central elements, those with the lowest occurrence number, to which other atoms are linked in an order determined by their occurrence number, producing a hierarchical structure to represent the species, as described in previous articles.⁶

Here, the weighting of each element or group of repeated elements in the final calculation is determined by its distance from the initial element or elements selected in the structure.

Clearly this calculation does not take into account other possible structures and thus does not deal with the internal structure of chemical species to which this disparity parameter would be applied. Hence, the same complexity value would be obtained for NaHSO₄ and Al(OH)SO₃, species which are totally distinct and of obviously different complexities, since their atoms are of equal number and diversity.

The SD measure has to be applied to each subcompound forming part of an inorganic species. These subcompounds are intercoordinated to form a more complex species, and so on, eventually forming a molecular species. Viewed thus, the structure of an inorganic species may be seen as a hierarchical structure where the root of the tree is the inorganic molecule, the nodes are the various component subcompounds, and the leaf nodes are the component atoms of these subcompounds. This hierarchical structure is represented in tabular form by the CST and CTCS;^{24,25} from these abstract structures, disparity measures may be easily derived.

Table 2. Calculation of MD and MF for a Selection of Inorganic Species

species	subspecies	SD/MDS _i	O _i	calculation	MD	MF
Na	Na ⁰	1.000	1	(1 × 1 × 1)/1 = 1/1	1.0000	1.0000
O ₂	O ²⁻	1.000	2	(1 × 2 × 1)/2 = 2/2	1.0000	1.0000
NaH	Na ⁺	1.000	1	(1 × 1 × 1 + 2 × 1 × 1)/2 = 3/2	1.5000	0.6667
	H ⁻	1.000	1			
Na ₂ O	Na ⁺	1.000	2	(1 × 1 × 1 + 2 × 2 × 1)/3 = 5/3	1.6667	0.6000
	O ²⁻	1.000	1			
Na(OH)	Na ⁺	1.000	1	(1 × 1 × 1 + 2 × 1 × 1.5)/2 = 4/2	2.0000	0.5714
	OH ⁻	1.500	1			
CaH ₂	Ca ²⁺	1.000	1	(1 × 1 × 1 + 2 × 2 × 1)/3 = 5/3	1.6667	0.6000
	H ⁻	1.000	2			
CaO	Ca ²⁺	1.000	1	(1 × 1 × 1 + 2 × 1 × 1)/2 = 3/2	1.5000	0.6667
	O ²⁻	1.000	1			
Ca(OH) ₂	Ca ²⁺	1.000	1	(1 × 1 × 1 + 2 × 2 × 1.5)/3 = 7/3	2.3333	0.4286
	OH ⁻	1.500	2			
Fe ₂ O ₃	Fe ³⁺	1.000	2	(1 × 2 × 1 + 2 × 3 × 1)/5 = 8/5	1.6000	0.6250
	O ²⁻	1.000	3			
Fe ₃ O ₄	FeO	1.500	1	(1 × 1 × 1.5 + 2 × 1 × 1.6)/2 = 4.7/2	2.3500	0.4255
	Fe ₂ O ₃	1.600	1			
H ₂ SO ₄	H ⁺	1.000	2	(1 × 1 × 1.8 + 2 × 2 × 1)/3 = 5.8/3	1.9333	0.5172
	SO ₄ ²⁻	1.800	1			
NaHSO ₄	Na ⁺	1.000	1	(1 × 1 × 1 + 2 × 1 × 2.3)/2 = 5.6/2	2.8000	0.3571
	H ⁺	1.000	1			
	HSO ₄ ⁻	2.300	1			
	SO ₄ ²⁻	1.800	1			
NaNOSO ₄	Na ⁺	1.000	1	(1 × 1 × 1 + 2 × 1 × 1.5 + 3 × 1 × 1.8)/3 = 9.4/3	3.1333	0.3191
	NO ⁺	1.500	1			
	SO ₄ ²⁻	1.800	1			
Al(OH)SO ₃	Al ³⁺	1.000	1	(1 × 1 × 1.75 + 2 × 1 × 2)/2 = 5.75/2	2.8750	0.3478
	Al(OH) ²⁺	2.000	1			
	OH ⁻	1.500	1			
	SO ₃ ²⁻	1.750	1			

Definition 3: The disparity of molecular structures, referred to as *molecular disparity* (MD), may be obtained through the iterative calculation of the disparities of the subcompounds forming the molecule; thus

$$MD = \frac{\sum_{i=1}^{i=s} i \min(O_i \text{ MDS}_i)}{c}$$

where s is the number of different subcompounds, c is the total number of subcompounds, and $\min(O_i \text{ MDS}_i)$ is the minimum of the product of the occurrences and the disparity of the subcompound i in the chemical species.

By examining the values of MD, the following may be deduced:

1. The value of $\text{MDS}_i = \text{SD}$ for a subcompound composed of atomic species, and when a subcompound is formed of other, lower-lying subcompounds, MDS_i is calculated using the previous expression.

2. When a molecular species comprises just one species with a multiplicity of 1 or n , the outcome is $\text{MD} = \text{SD}$.

3. MD is based on SD and will thus take into account the diversity and multiplicity of the atoms found in the chemical species.

4. MD takes into account the diversity and multiplicity of the subcompounds forming a chemical species.

5. MD does not take into account the charge of the species in question, with the result that $\text{MD}(\text{HIO}_4) = \text{MD}(\text{HSO}_4^-)$.

6. MD takes into account the structural relationships existing between the atoms forming a species. These atoms are interrelated by forming subspecies (subcompounds) which then join up to form either other subcompounds or a

more complex chemical species. Thus, as may be seen in Table 2, the molecular disparity of a species of the form XYZ is different from that of another species of the form $X(YZ)$; in the former, there is only one subcompound formed by the three elements, whereas in the latter there are two subcompounds, one formed by a single element and the other by two.

2.1. Molecular Feasibility. Here we wish to standardize molecular disparity measures for use in the proposal of species that may be present in an incomplete reaction.

Definition 4: *Molecular feasibility* (MF) is defined as a measure of the possibility of an inorganic species being present in a reaction, and is obtained as the inverse of molecular disparity, namely

$$MF = 1/MD$$

Table 2 shows the values for MD and MF for certain inorganic species. It may be seen that MF varies through the interval [0,1], its value being consistent with those postulated in the model. Thus, a species in its native state will take on the value $MF = 1$, and be most likely to exist in a reaction, from the point of view of complexity.

3. DESCRIPTION OF THE KNOWLEDGE FRAME

A knowledge frame is an abstract representation of knowledge concerning a specific problem. Its main components are thus^{26,27}

- a knowledge/fact base representing information about the problem, derived from experience and observation of the problem by an expert
- a set of rules representing the reasoning processes used to tackle and solve the problem

- one or a set of decision parameters representing the measure in which both final and partial solutions to the problem may be accepted for the making of decisions

- a set of procedures or methods representing the deductive and inductive processes inferred when a rule is satisfied through evaluating data corresponding to a specific, determined problem, with respect to the knowledge/fact base.

In the previous section, description was given of the decision-making parameters used in the proposed knowledge frame, while the current section will describe not only the parameter used as a measurement of the possibility of the final solution, but also the remaining components of the proposed knowledge frame.

The knowledge frame considers states, conditions, and procedures.

A reaction to be examined and solved will pass through various states. A state refers to when the reaction satisfies a set of conditions.

A condition is a predicate whose premises consider the features of the reaction and which is dealt with by the procedures, giving rise to a transition of the reaction to a new state.

A procedure is one or a set of methods brought into play when a reaction reaches a certain state. Each state is assigned one and only one procedure, and this is named the same as the state. A procedure assigned to one state may call procedures from other states, meaning that certain procedures will necessarily involve others, or states that are hierarchically dependent on others, when viewed from the abstraction.

The process of solving a reaction will therefore involve the application of a set of conditions, from the initial state to the point at which an analytical solution is reached. During the process the problem reaction will advance through a semantic/functional network until it reaches a leaf node; this terminal node means that either a solution has been found or the state is declared as inconsistent; that is, no solution for the reaction has been defined in the knowledge frame.

Figure 1 shows the semantic/functional network; states are represented by unshaded circles, final states or nodes containing solutions by shaded circles, procedures by round-cornered squares, and connections between states by pentagons.

As the reaction reaches a given state, the associated procedure is automatically executed. These procedures may consist of one of the following actions:

Type PI: The reaction is solved, as in the case of final states or nodes containing solutions (e.g., *S19*, *S25*, and *S35*).

Type PII: The reaction is examined at this state without alteration of its features, and a new condition is applied (e.g., *System*, *S1*, *S2*, *S3*, *S4*, etc.).

Type PIII: The reaction is examined and its features are altered, producing one or a set of new reactions, each of which is considered as a new state, and the procedures corresponding to its branch on the semantic/functional network are applied (e.g., *S10*, *S20*, *S36*, *S37*, etc.).

When a type PIII procedure alters the state of a reaction and the new reaction or set of reactions is submitted to the procedures for a new state, two situations may arise:

The new reaction may reach its final state, thus ending the process.

The new reaction may reach the same state again, producing an exception. An exception to a type PIII

procedure may be either a new procedure of the same type or an inconsistency, the latter determining that a reaction or given new reaction has no solution available in the defined knowledge frame.

An inconsistency is produced when the reaction, at a given state, is acted upon by a type PIII procedure and this does not alter its state.

3.1. Description of the Conditions. The knowledge frame is formed by nine conditions (A–I, Figure 1). Each condition is considered by the procedure associated to a state, although any one condition may be considered by more than one procedure.

The conditions are simple feature assertions of the system. There are two kinds of conditions:

- The first kind is those which will be satisfied or not by a reaction in a given state, as is the case for conditions A, B, C, D, E, F and H. If the features of a reaction do not allow complete evaluation of these conditions due to a lack of relevant information, the reaction is said to have failed to satisfy the condition; an instance of this would be where a reaction contained no species whatsoever on one side of the equation.

- The second kind is those which may or may not be satisfied by the reaction at the same time; that is, their consideration will only determine the cardinality of the set of possible solutions depending on whether they are found to be true or false, as is the case for conditions G and I.

Condition A: checks for the completeness of matter; that is, both sides are examined to see if they contain the same atoms, regardless of number

Condition B: checks for the presence of valence changes in the atoms present in the reactants and the products, which would be a redox reaction

Condition C: checks if the reaction in a given state has an analytical solution for the balancing of its matter

Condition D: checks if a matter-balanced reaction is also charge-balanced, that is, if the sum of the charges of the reactants equals that of the products

Condition E: checks if a redox reaction is complete; in other words, if both oxidation and reduction processes are present

Condition F: checks whether in an incomplete redox reaction there is just one half-reaction or more than one oxidation or reduction process

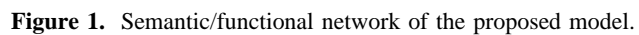
Condition G: checks whether to try and solve an incomplete redox reaction as a half-reaction or as a complete redox reaction

Condition H: checks if in a matter-deficient reaction this deficit is only due to oxygen and/or hydrogen atoms

Condition I: checks whether to proceed with a search for new species whose valences are different from that of the present species; that is, whether to go ahead and change the redox state of the system

3.2. Description of the States. A state characterizes the problem reaction during the solving process. At any given moment in this process, a reaction has a unique state assigned to it and the procedure associated with this state is thus applied.

A state characterizes a reaction and classifies it within the knowledge frame, since for a reaction to reach a state it must satisfy a series of conditions.



The states dealt with in the knowledge frame, as shown in Figure 1, are as follows:

State 0 (*System*): the initial problem reaction

State 0* (*System**): the reaction formed by the half-reactions obtained from the original reaction at a given state, to be dealt with as a new problem reaction

State 1: a matter-complete reaction, where both sides contain the same atomic types

State 2: a matter-deficient reaction

State 3: (1) and a redox process is present

State 4: (1) and no redox process is present

State 5: (2) and the deficit is seemingly due only to oxygen and/or hydrogen atoms

State 6: (2) and the matter deficiency is due to atoms that need not be just oxygen and/or hydrogen

State 7: (1 + 3) and a reduction and oxidation process is present

State 8: (1 + 3) and redox is incomplete, perhaps due to a half-reaction

State 9: (1 + 4) and the reaction has an analytical solution

State 10: (1 + 4) and the reaction has no analytical solution, leading to the deduction that the reaction is incomplete

State 11: (2 + 5) and there is a redox reaction

State 12: (2 + 5) and there is no valence change

State 13: (2 + 6) and a valence change is present

State 14: (2 + 6) and there is no valence change

State 15: (1 + 3 + 7) and the global reaction (without being broken down into half-reactions) has an analytical solution

State 16: (1 + 3 + 7) and the global reaction has no analytical solution

State 17: (1 + 3 + 8) and there is a half-reaction, as there is only one oxidation or reduction process

State 18: (1 + 3 + 8) and more than one oxidation and/or reduction process is present in the reaction

State 19: (1 + 4 + 9) and the reaction is charge-balanced

State 20: (1 + 4 + 9) and the reaction is charge-unbalanced, meaning the reaction is inconsistent

State 21: (2 + 6 + 13) and the redox reaction is complete; that is, there is both an oxidation and a reduction process

State 22: (2 + 6 + 13) and the redox reaction is incomplete

State 23: (2 + 6 + 14) and the reaction can be solved by finding species whose atoms may have a different valence from that found in the reaction

State 24: (2 + 6 + 14) and the reaction can be solved by finding species whose atoms have the same valence as that found in the reaction

State 25: (1 + 3 + 7 + 15) and the adjusted reaction is charge-balanced

State 26: (1 + 3 + 7 + 15) and the adjusted reaction is not charge-balanced

State 27: (1 + 3 + 8 + 17) and the half-reaction may be adjusted without adding new species

State 28: (1 + 3 + 8 + 17) and new species must be considered in order to solve the half-reaction

State 29: (1 + 3 + 8 + 18) and the reaction is dealt with as a complex half-reaction to obtain a solution

State 30: (1 + 3 + 8 + 18) and the complex half-reaction is treated as a complete redox reaction, and new species must be added in order to ensure the redox is complete

State 31: (2 + 6 + 13 + 21) and new species will be found that will change the redox state of the reaction

State 32: (2 + 6 + 13 + 21) and only new species that do not alter the reaction's redox state will be admitted

State 33: (2 + 6 + 13 + 22) and new species will be found that will change the redox state of the reaction

State 34: (2 + 6 + 13 + 22) and only new species that do not alter the reaction's redox state will be admitted

State 35: (1 + 3 + 8 + 17 + 27) and a solution for the half-reaction is found directly, since the electrons are present

State 36: (1 + 3 + 8 + 17 + 27) and electrons must be added to find a charge-balanced solution for the half-reaction

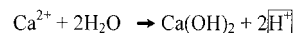
State 37: (1 + 3 + 8 + 17 + 28) and the reaction is treated as a half-reaction in order to obtain a solution, perhaps having to consider species such as H₂O, H⁺, OH⁻, and e⁻, to achieve the solution.

State 38: (1+3+8+17+28) and the reaction is treated as a complete redox reaction; new species must be added in order to ensure redox completeness

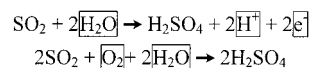
3.2.1. Examples of Reactions for Terminal States, Associated to PIII Procedures. There now follows a description of several terminal states; examples are given of reactions that satisfy these, together with one of the possible solutions the knowledge frame might propose. Inorganic species added during the solving process are shown within boxes.

State 10: an incomplete, nonredox reaction in which analytical adjustment cannot be performed; these are reactions which

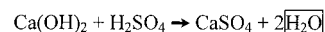
- are simply solved through the consideration of an aqueous solution.



State 11: matter-deficient redox reactions, whose deficit is due only to oxygen and/or hydrogen atoms, and which may be solved as either a half-reaction or a complete redox reaction



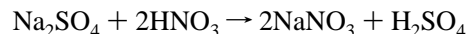
State 12: matter-deficient, nonredox reactions, whose deficit is only due to oxygen and/or hydrogen atoms



State 16: matter-complete, redox reactions with no direct analytical solution; these are dealt with in the same way as those in S26



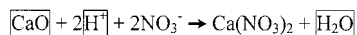
State 19: matter-complete, nonredox reactions that may be solved directly



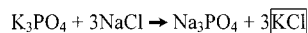
State 20: nonredox reactions, with an analytical solution producing charge imbalance; this is an inconsistent state that should not be presented, so an exception is produced since no solution can be found

State 23: matter-deficient, nonredox reactions, whose deficit may be due to oxygen and/or hydrogen or other

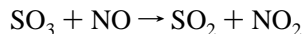
different atoms and which can be solved by adding new species that may alter the redox state of the reaction



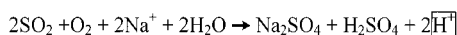
State 24: matter-deficient, nonredox reactions, whose deficit may be due to oxygen and/or hydrogen or other different atoms and which can be solved without altering the redox state of the reaction



State 25: matter-complete, redox reactions having a correct (charge-balanced) analytical solution

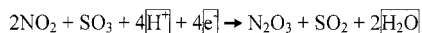


State 26: matter-complete, redox reactions not producing a balanced charge after being globally adjusted

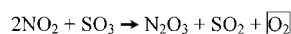


Since there is no charge balance, it is assumed that there is a deficiency in one or more of the following species in the reaction: H_2O , H^+ , OH^- , thus bringing the reaction to a new state *S7*.

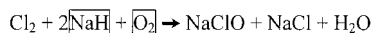
State 29: matter-complete reactions with incomplete redox reactions, where there is more than one oxidation or reduction half-reaction, and which can be solved as a compound half-reaction



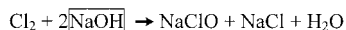
State 30: matter-complete reactions with incomplete redox reactions, where there is more than one oxidation or reduction half-reaction, and which can be solved as a complete redox reaction



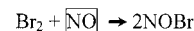
State 31: complete redox reactions where matter deficiency is due to atoms other than oxygen and hydrogen (although possibly to these, too) and which can be solved by finding species of equal and different valences compared to those of the atoms already present in the reaction



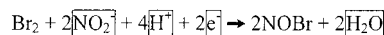
State 32: complete redox reactions where matter deficiency is due to atoms other than oxygen and hydrogen (although possibly to these, too) and which can be solved by finding species of valence equal to that of the atoms already present in the reaction



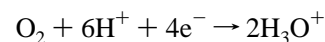
State 33: complete redox reactions where matter deficiency is due to atoms other than oxygen and hydrogen (although possibly to these, too) and which can be solved by finding species whose valence states are equal to and different from those of the atoms already present in the reaction



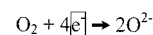
State 34: complete redox reactions where matter deficiency is due to atoms other than oxygen and hydrogen (although possibly to these, too) and which can be solved by finding species whose valence states are equal to those of the atoms already present in the reaction



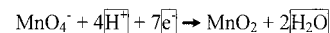
State 35: matter-complete reactions containing a redox half-reaction and the electrons needed to balance the charge when the reaction is solved



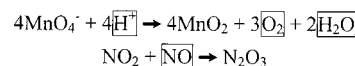
State 36: matter-complete reactions containing a redox half-reaction but requiring the consideration of electrons in order to obtain a charge-balanced solution



State 37: matter-complete reactions containing a redox half-reaction but having no analytical solution; these will be solved as half-reactions using the *P10* procedure, bringing the reaction to the state *S17*



State 38: matter-complete reactions containing a redox half-reaction but having no analytical solution; here, these will be solved as complete redox reactions, using the *P38* procedure to bring the reaction to the state *S7*



3.3. Description of the Procedures. Procedures are functional methods associated with states. Each state is associated with just one procedure, although a procedure may be called by more than one state. Procedures consist of functional methods and/or other procedures. Thus, a procedure associated with one state may call procedures associated with other states.

Depending on the type of procedure associated with a state, states may be classed as stable or volatile. A stable state is one whose associated procedure is of type PI or PII. A volatile state is one whose associated procedure is of type PIII, since this procedure type automatically changes the state of the reaction without submitting it to a new condition.

At the initial stages, in *System* state, a series of checks and calculations is carried out by a set of procedures which determine the features of the reaction. These initial procedures consist of methods that may be called by other procedures corresponding to other states. This initial analysis provides information such as the following:

1. the atoms present on either side of the equation
2. the valence of each atom in all its incidences
3. the occurrences of each atom on either side of the equation
4. the existence of a charge in the reaction, and the charge on each side
5. the existence of valence changes
6. the presence of the species H_2O , H^+ , OH^- , and electrons

7. moreover, these general procedures include all those employed in the manual and analytical solution of the reaction, as described elsewhere⁶

Figure 1 shows the type PIII procedures as rounded boxes.

The *P10* procedure is responsible for adding to the reaction one or more of the following species: H_2O , H^+ , and OH^- . This procedure is used for the completion of matter in redox half-reactions, or in nonredox reactions whose species deficit is solely due to these atoms (type 3 reactions, in accordance with the proposed classification⁶).

The *P25* procedure is responsible for breaking down a redox reaction into its half-reactions, independently finding a solution for each of these, and forming the global reaction for subsequent balancing. It is optionally used to provide the user with extended information about the balancing process, as described in a previous article.⁶

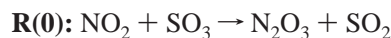
The sole purpose of the *P36* procedure is to add the electrons needed to balance the charge of a half-reaction already balanced in matter.

Conceptually, the *P30*, *P31*, *P32*, and *P38* procedures are highly similar, being responsible for formulating new species and adding them to the reaction (instantiated in the case of type 4 reactions, according to the proposed classification⁶). They all use a common set of methods whose function is to formulate new species within restrictions defined by the current state of the reaction, as well as by the user. Once obtained, these new species are added to one or the other side of the equation, providing one or a set of possible solutions, which will then be dealt with independently.

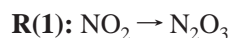
Thus, the *P38* procedure will create solutions by adding species to ensure the reaction is redox complete, while the *P30* procedure, highly similar to *P38*, builds solutions to complete complex redox half-reactions (more than one oxidation or reduction), in both cases using only atoms already present in the reaction. On the other hand, the *P31* and *P32* procedures create solutions by adding species formed by the atoms missing from the reaction; the former procedure considers any valence at all, whereas the latter considers only the valence(s) the present species would have.

For a better understanding of the model, let us examine how the *P30* procedure works. Here and in other forthcoming examples, the set of possible resulting solutions has been simplified for clarity.

Suppose we have the following problem reaction, at state *S30*, where we have an incomplete redox reaction composed of more than one half-reaction (in this case, reduction):



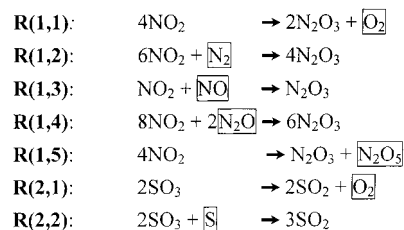
First, employing methods also used in procedure *P25*, the reaction is broken down into its corresponding half-reactions, giving:



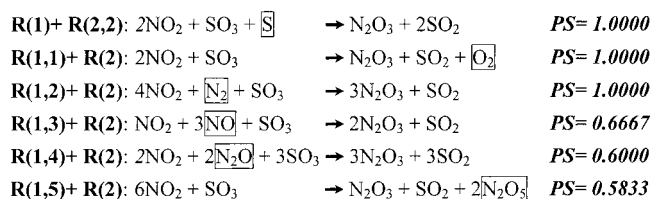
Next, the *P30* procedure, using the same methods as the *P38* procedure, finds the new species required to complete each of the redox half-reactions. These methods are responsible for finding new species formed by the atoms present in the reaction. In the current example, these are S, N_2O ,

NO , N_2O_5 , N_2 , and O_2 (since no ionic species have been considered in this example).

Once these species have been obtained, partial (complete redox) solutions are created for each of the half-reactions (the *P38* procedure would act much in the same way, but on only one half-reaction)



Finally, the new reactions are built by simply combining each of the partial solutions $R(i,j)$, for $j = 1, n$, with each of the half-reactions $R(k)$, for $k \neq i$.



since $R(1,1) + R(2) = R(2,1) + R(1)$, with each of these global solutions being considered as a new reaction at state *S7* (see Figure 1), thus obtaining the above solutions, via a transition of each of these through the states *S7–S15–S25*. Next to each of these solutions the measure of possibility (PS) has been shown; this term will be fully explained later in the article.

3.4. Finding New Species. The knowledge framework contains several procedures responsible for finding new species to add to the reaction and produce a solution. These procedures (*P30*, *P31*, *P32*, and *P38*) include a series of methods designed to formulate these new species, using either all of the atoms present in the reaction or only those deficient on one or other side.

These methods perform their tasks using the following information:

1. The value of a series of flags defining the environment or domain of the task. Examples of such flags would be *Redoxcomplete* (used to evaluate the condition G), *Allvalences* (used to evaluate the condition I), *OHcheck* (used to consider oxygen and hydrogen atoms in the new species to be formulated, even though these are not present in the reaction), and *Ioncheck* (used to consider ionic species). The default value of these flags is set by the user, although it may be automatically updated at run time, depending on the nature of the reaction. For instance, if a ionic species is present, the *Ioncheck* flag is automatically set.

2. The atoms with which new species may be formulated. These will include those already present in the reaction and/or those showing a deficit, in addition to those signaled by the *OHcheck* flag.

3. The valences revealed by these atoms in the reaction and, depending on the values of the *Redoxcomplete* and *Allvalences* flags, whether in the new species to be considered these atoms may appear with the same and/or different valences.

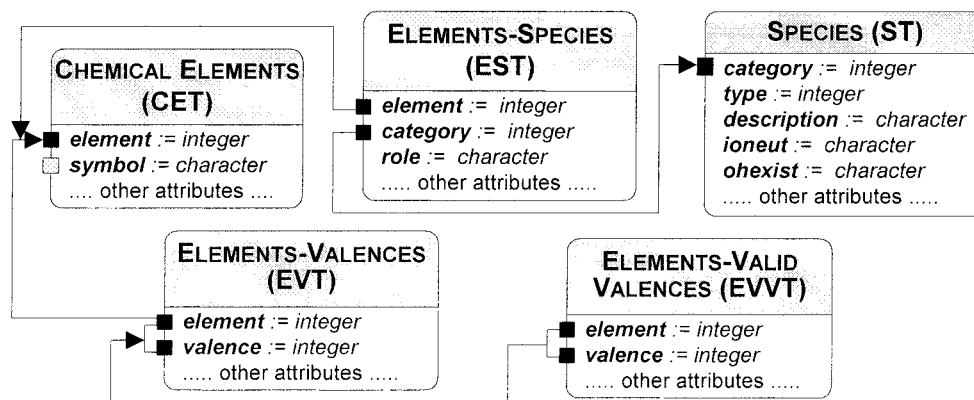


Figure 2. Relational diagram of the tables described in Chart 1.

New species that may be added to the reaction are determined via a complex process that uses not only information managed by the *Formulator*²⁸ system for the translation of inorganic compound names, but also data structures representing the reaction⁶ and a series of working tables. All this information is used to build the EVVT (*element and valid valence table*), which holds information about the elements and their permitted valences which may be used to formulate new species to be added to the reaction.

Chart 1 shows a simplified description of the structure of the most significant tables employed in the process, while Chart 2 shows the pseudocode for the algorithm used to build the working table EVVT. Additionally, Figure 2 contains a relational diagram of the tables described in Chart 1.

It can be seen from its description that the EVVT table stores

- information about the chemical elements that may be considered for the new species to be added to the reaction
- for each chemical element, the value of permitted valences with which they might appear in the new species under consideration; the valence values considered are determined by the values of the *Redoxcomplete* and *Allvalences* flags, which determine the cardinality of the possible set of new species that may be added to the reaction

Chart 2 shows the pseudocode for the procedure responsible for the construction of the EVVT table. This table is used to formulate the set of valid species that might be added in order to complete the reaction, via a process that makes use of:

■ the entries (chemical elements and their valences) in the EVVT table

■ the values of the *OHcheck* and *Ioncheck* flags, which determine the type of species that may be formulated; to achieve the latter, reference is made to the ST and EST tables (see Chart 1)

■ the UMLS and UVLS⁶ lists, which contain information on matter deficiency and valence changes on both sides of the reaction

■ the *Formulator* system that builds the corresponding CTS²⁸ for the new species

Finally, the list of species created is sent back to the corresponding procedure, and other methods are then assigned to building new solutions to the reaction, inserting one or more species from the list into the appropriate side of the equation and modifying the RAS data structure that represents it.⁶

In the example described above for the *P30* procedure, the service performed by these methods can be observed. In this case the procedure has worked with the flag settings *Redoxcomplete* = "Yes", *OHcheck* = "No", and *Ioncheck* = "No". As nitrogen is present with valences of 4 and 2, these incidences are disallowed, and likewise the valences of 6 and 4 for sulfur.

Thus, compounds will only be sought where nitrogen has a valence of -3, 0, 1, 2, and 5, and where sulfur has a valence of 0. Furthermore, since oxygen is present in the reaction, the search will be carried out for species composed of oxygen and these two elements.

Clearly the only feasible solutions are the oxides of nitrogen, N₂O, NO, and N₂O₅, and molecular nitrogen N₂, with valences of 1, 2, 5, and 0, respectively; molecular sulfur S⁰, and molecular oxygen O₂, as depicted in the example of this procedure.

3.5. Possibility of the Solution. As may be observed in the description of the knowledge frame used for solving inorganic reactions, on many occasions the procedures associated with certain states will produce, as indeed would be desired, a set of solutions instead of just one.

The final step in the global process of solving reactions consists of ordering this set of solutions. In the proposed model, the ordering parameter or criterion employed is called the *possibility of the solution* (PS).

The possibility of a solution to a reaction is taken as being inversely proportional to the number of new species considered when achieving the solution and is a direct function of their feasibility. The possibility of a solution may then be expressed thus:

$$PS_i = \frac{\prod_{j=1}^{j=N_i} MF(i,j)}{N_i}$$

That is, the possibility of a given solution PS_i is equal to the product of the molecular feasibilities of the new species *j* considered for the solution MF(*i,j*), divided by the total number of new species considered for this solution N_i.

Thus, the most likely solutions are reckoned to be those that have considered the smallest number of species and/or that have the highest value for molecular feasibility, in other words, species formed by a smaller number of atoms and where these atoms are of lower diversity, which in turns leads

Chart 1. Simplified Structure of Some of the Tables Referenced in the Knowledge Frame

```

Define Table ST (Species Table)
  /* Table of information about species, compounds/sub-
  compounds/elements considered by the Formulator28 system */
  /* The key and order of elements in the table are
  determined according to the attribute category described in
  previous studies24 */
  Category := integer,
    /* value or code assigned to the category or type of
    inorganic species (see references 24-25, 28) */
  type := integer,
    /* flag to indicate category class: compound, subcompound,
    atom, structure, etc. (Ref. 24) */
  description := character,
    /* description of the type of compound, subcompound,
    element or structure */
  ioneut := character,
    /* description of the ionic nature of the species.
    Permissible values are: P: positive, N: negative,
    C: neutral, O: not considered */
  OHexist := boolean,
    /* flag that denotes whether oxygen and/or hydrogen atoms
    may be present in the species */
  ...Other attributes...
End Table ST

Define Table CET (Chemical Element Table)
  /* Table containing information on chemical elements */
  element := integer,
    /* atomic number of chemical element */
  symbol := character,
    /* symbol of chemical element */
  ...Other attributes...
End Table CET

Define Table EVT (Element-Valence Table)
  /* Table containing the set of permissible valences for
  each of the chemical elements. Key and order of the table
  items is determined by the aggregate (element+valence) */
  element := integer,
    /* atomic number of the chemical element */
  valence := integer,
    /* known valence of the chemical element */
  ...Other attributes...
End Table EVT

Define Table EST (Element-Species Table)
  /* Table containing the set of species (compounds,
  subcompounds and atomic states) that may be formulated for
  each of the chemical elements. Key and order of the table
  elements are determined via the aggregate
  (element+category) */
  element := integer,
    /* atomic number of the chemical element */
  category := integer,
    /* value or code assigned to the category or type of
    inorganic species (see references 24-25, 28) */
  role := character,
    /* role the element may play in the chemical species;
    whether it may act as an anion, cation, both or neither */
  ...Other attributes...
End Table EST

Define Table EVVT (Element and Valid Valence Table)
  /* Table containing information on chemical elements and
  the valences that may be considered on either side of the
  reaction in order to find new species */
  element := integer,
    /* atomic number of the chemical element */
  valence := integer,
    /* permissible valence for the chemical element */
  ...Other attributes...
End Table EVVT

```

Chart 2. Pseudocode for the EVVT Table Generating Algorithm

```

Define Procedure BEVVT (Build Element and Valid Valence Table)
  /* Description of pseudo-code for the procedure for
  building the EVVT table */
  ELS *els, *wels
  /* defines two pointers to the list of elements in the
  reaction (see reference 6) */
  *els = Ras.els,
  /* initializes the value of the *els pointer to the first
  element on the list of elements in the reaction */
  int ele, val,
  /* defines two working variables for elements and valences,
  to be used to access the EVT table */
  int ok,
  /* defines a working variable to be used when deciding
  whether to consider the element/valence pair in the EVVT
  table */
  Do Until Ras.els <> *els.nextels
    /* run through the complete list of elements */
    ele = *els.element,
    val = -99,
    /* initializes variables for generic access to the EVT
    table */
    While ele = *els.element
      /* reads the element's known valences from the EVT table */
      readnext EVT(ele, val),
      /* reads the element/valence pair */
      ok = 0,
      /* initializes value to zero, indicating that by default,
      the element/valence pair will not be placed in the EVVT
      table */
      Switch: Allvalences
        /* examines the value of the Allvalences flag */
        Case Yes:
          Ok = 1,
          /* if Allvalences='yes', all the valences of the elements
          from the element list (ELS) will be considered */
          Case No:
            /* if Allvalences='no', it will be the value of the
            Redoxcomplete flag that determines the valences to be
            considered */
            If Redoxcomplete = 'No' and EVT.valence = *els.valence
              ok = 1,
              /* if the Redoxcomplete flag = 'no', only the valences
              present in the reaction will be considered */
              ElseIf Redoxcomplete = 'Yes'
                *wels = *els,
                /* initializes the *wels pointer to scan the ELS list using
                the pointer which addresses equal elements */
                While *els.element = *wels.element
                  /* runs through the element list until the element is
                  changed */
                  If val != *wels.valence
                    ok = 1,
                    /* temporarily considers that the element/valence pair
                    should be placed in the EVVT table */
                    *wels = *wels.nextelement,
                    /* assigns to the pointer the next equal element from the
                    element list (ELS) */
                  Else
                    ok = 0,
                    /* on finding an equal valence, the process considers that
                    the element/valence pair should not be placed in the EVVT
                    table */
                    *els.element = null,
                    /* assigns a null value to the element in order to exit the
                    loop */
                  End If
                /* if the Redoxcomplete flag = 'yes', only valences not
                present in the reaction will be considered */
              End While
            End If
          End Case
        End Case
      End While
    End Until
  End Procedure

```



```

If ok = 1
  Place in EVVT(ele, val),
  /* if the ok flag =1, the element/valence pair will be
  placed in the EVVT table to be considered in the
  formulation of new species */
End If
*els = *els.nextels,
/* the next element from the element list (ELS) is accessed
*/
While existEVVT(*els.element, *els.valence) = 'True'
  *els = *els.nextels,
  /* a check is made to see if the new element/valence pair
  has previously been placed in the EVVT, and if that is the
  case, the process moves on to the next element from the ELS
  */
End While
End While
End Do
End Procedure BEVVT.

```

to the conclusion that the best solution will be that requiring the fewest atoms forming the fewest species.

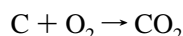
In the example shown for the *P30* procedure, the possibility values are shown for each of the solutions depicted. A higher possibility value means that the solution will be prioritized.

4. APPLICATION OF THE MODEL TO EXAMPLE REACTIONS

The knowledge frame will now be applied to an example reaction, in which several states will be supposed in order to follow the semantic/functional network of the solution; the possible solutions produced have been simplified to improve legibility. In all examples, unless expressly stated otherwise, the working flags (*Allvalences*, *Redoxcomplete*, *Ioncheck*, and *OHcheck*) have been taken to be initially clear.

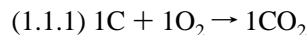
Example 1: a simple redox reaction. (Table 3 shows conveniently formatted computer output from the solving process, step-by-step, for cases 1–3 of this example).

Case 1.1: Take the following initial reaction, a matter-complete redox reaction:

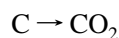


Tracking states: The solution of this reaction via the knowledge framework leads to transitions through the following states: *S1–S3–S7–S15–S25*. Since the reaction is complete, this process is independent of the initial values of the flags.

Solution obtained:



Case 1.2: Suppose the reaction shows an oxygen deficit. Eliminating the O_2 from the left side, we have

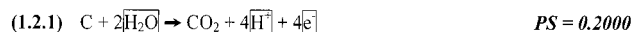


Tracking states: *S2–S5–S11–S3–S8–S17–S28–continues*.

Solutions obtained (only a selection is shown):

■ At state *S28*, as the *Redoxcomplete* flag is clear, the reaction will be solved as a half-reaction, following the transition: *S28–S37–S17–S27–S36–S27–S35*. The *P10* procedure instantiated at state *S37* automatically sets the

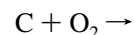
Ioncheck flag, irrespective of its initial value, in order to proceed to an adjustment process for redox half-reactions. The *OHcheck* flag is automatically set as there is oxygen in the reaction.



■ If, on the other hand, the *Redoxcomplete* flag were set, the reaction would be solved as a complete redox reaction, following the transition sequence *S28–S38–S7–S15–S25*.



Case 1.3: Suppose the reaction is deficient in both oxygen and carbon. Eliminating the CO_2 from the right side, we have



Tracking states: *S2–S6–S13–S22–S33–continues*, since there are species in molecular state and it is thus necessary to consider new species that will change the valence state of the atoms present. Condition B is thus satisfied and the *Allvalences* flag is set, since all species present are found to be in molecular state.

Solutions obtained (only a selection is shown): the species considered by *P31* will be CO and CO_2 .

■ In both cases the reaction will progress through the following states: *S33–S3–S7–S15–S25*.



Case 1.4: Suppose that the reaction again reveals a deficit of oxygen and carbon, but this time eliminating all these species from the left side we have



Tracking states: *S2–S6–S14–S24–continues*, and since there are no species on one side and the known species are not in molecular state, there is no information to let us know if it is a redox process, and the *Allvalences* flag is cleared.

Solutions obtained (only a selection is shown):

■ If the *Ioncheck* flag were set, at state *S24*, procedure *P32* would find ionic species such as CO_3^{2-} , producing solutions with the following transition sequence: *S24–S4–*

Table 3

Case	State	Value of the Reaction in the State	Conditions Evaluated	New Flag Values	Action
Initial Flag Values: Allvalences = No OHcheck = No Ioncheck = No Redoxcomplete = No					
(1.1)	0	$C + O_2 \rightarrow CO_2$	A = Yes	OHcheck = Yes	Set State to 1
	1	$C + O_2 \rightarrow CO_2$	B = Yes	None	Set State to 3
	3	$C + O_2 \rightarrow CO_2$	E = Yes	None	Set State to 7
	7	$1C + 1O_2 \rightarrow 1CO_2$	C = Yes	None	Balancing matter Set State to 15
	15	$1C + 1O_2 \rightarrow 1CO_2$	D = Yes	None	Set State to 25
	25	$1C + 1O_2 \rightarrow 1CO_2$	None	None	End. PS= 1.0000
Initial Flag Values: Allvalences = No OHcheck = No Ioncheck = No Redoxcomplete = No					
(1.2.1)	0	$C \rightarrow CO_2$	A = No	OHcheck = Yes	Set State to 2
	2	$C \rightarrow CO_2$	H = Yes	None	Set State to 5
	5	$C \rightarrow CO_2$	B = Yes	None	Set State to 11
	11	$C \rightarrow CO_2$	None	None	Set State to 3
	3	$C \rightarrow CO_2$	E = No	None	Set State to 8
	8	$C \rightarrow CO_2$	F = Yes	None	Set State to 17
	17	$C \rightarrow CO_2$	C = No	None	Set State to 28
	28	$C \rightarrow CO_2$	G = Yes	None	Set State to 37
	37	$C + H_2O \rightarrow CO_2 + H^+$	None	Ioncheck = Yes	Add H ₂ O, H ⁺ Set State to 17
	17	$C + 2H_2O \rightarrow CO_2 + 4H^+$	C = Yes	None	Balancing matter Set State to 27
	27	$C + 2H_2O \rightarrow CO_2 + 4H^+$	D = No	None	Set State to 36
	36	$C + 2H_2O \rightarrow CO_2 + 4H^+ + 4e^-$	None	None	Add electrons Set State to 27
	27	$C + 2H_2O \rightarrow CO_2 + 4H^+ + 4e^-$	D = Yes	None	Set State to 35
	35	$C + 2H_2O \rightarrow CO_2 + 4H^+ + 4e^-$	None	None	End. PS= 0.2000
Initial Flag Values: Allvalences = No OHcheck = No Ioncheck = No Redoxcomplete = Yes					
(1.2.2)	0	$C \rightarrow CO_2$	A = No	OHcheck = Yes	Set State to 2
	2	$C \rightarrow CO_2$	H = Yes	None	Set State to 5
	5	$C \rightarrow CO_2$	B = Yes	None	Set State to 11
	11	$C \rightarrow CO_2$	None	None	Set State to 3
	3	$C \rightarrow CO_2$	E = No	None	Set State to 8
	8	$C \rightarrow CO_2$	F = Yes	None	Set State to 17
	17	$C \rightarrow CO_2$	C = No	None	Set State to 28
	28	$C \rightarrow CO_2$	G = Yes	None	Set State to 38
	38	$C + O_2 \rightarrow CO_2$	None	None	Find O ₂ Set State to 7
	7	$1C + 1O_2 \rightarrow 1CO_2$	C = Yes	None	Balancing matter Set State to 15
	15	$1C + 1O_2 \rightarrow 1CO_2$	D = Yes	None	Set State to 25
	25	$1C + 1O_2 \rightarrow 1CO_2$	None	None	End. PS= 1.0000
Initial Flag Values: Allvalences = No OHcheck = No Ioncheck = No Redoxcomplete = No					
(1.3)	0	$C + O_2 \rightarrow$	A = No	OHcheck = Yes	Set State to 2
	2	$C + O_2 \rightarrow$	H = No	None	Set State to 6
	6	$C + O_2 \rightarrow$	B = Yes	None	Set State to 13
	13	$C + O_2 \rightarrow$	E = No	None	Set State to 22
	22	$C + O_2 \rightarrow$	I = Yes	Allvalences = Yes	Set State to 33
(1.3.1)	33	$C + O_2 \rightarrow CO$	None	None	Find CO Set State to 3
(1.3.1)	3	$C + O_2 \rightarrow CO$	E = Yes	None	Set State to 7
(1.3.1)	7	$2C + O_2 \rightarrow 2CO$	C = Yes	None	Balancing matter Set State to 15
(1.3.1)	15	$2C + O_2 \rightarrow 2CO$	D = Yes	None	Set State to 25
(1.3.1)	25	$2C + O_2 \rightarrow 2CO$	None	None	End. PS= 0.6667
(1.3.2)	33	$C + O_2 \rightarrow CO_2$	None	None	Find CO ₂
(1.3.2)	3	$C + O_2 \rightarrow CO_2$	E = Yes	None	Set State to 7
(1.3.2)	7	$1C + 1O_2 \rightarrow 1CO_2$	C = Yes	None	Balancing matter Set State to 15
(1.3.2)	15	$1C + 1O_2 \rightarrow 1CO_2$	D = Yes	None	Set State to 25
(1.3.2)	25	$1C + 1O_2 \rightarrow 1CO_2$	None	None	End. PS= 0.6000

S10–S4–S9–S19 (the *OHcheck* flag is automatically set since oxygen is present in the reaction).



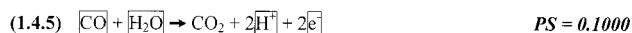
■ If the *Ioncheck* flag is cleared (the default value), an exception will be produced, leading to a transition to state *S23*, where procedure *P31* (which automatically sets the *Allvalences* flag) will find species such as C, CO and O₂, and the following sequence of states will be produced: *S24–S23–S3–S7–S15–S25*, for the next solution:



The sequence *S24–S23–S3–S8–S17–S28–S38–S7–S15–S25*, if it is solved as a complete redox reaction (if the *Redoxcomplete* is set), where first CO is considered by the *P31* procedure and then O₂ or C by the *P38* procedure:

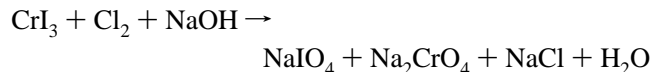


and the sequence *S24–S23–S3–S8–S17–S28–S37–S17–S27–S36–S27–S35*, if it is solved as a half-reaction (if the *Redoxcomplete* flag is clear, i.e., the default value):



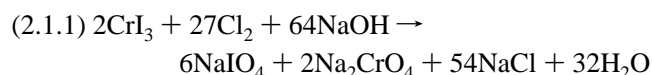
Example 2: a more complex redox reaction

Case 2.1: Take the following initial reaction, a matter-complete redox reaction:

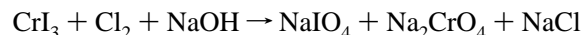


Tracking states: The solution to this reaction via the knowledge framework leads to transitions through the following states: *S1–S3–S7–S15–S25*, irrespective of flag values, since we are dealing with a complete reaction.

Solution obtained:



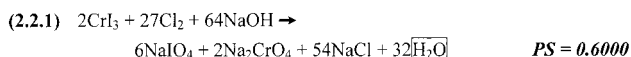
Case 2.2: Suppose the reaction shows a hydrogen deficit. Eliminating H₂O from the right side, we have



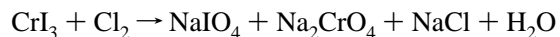
Tracking states: *S2–S5–S11–S3–S7–S16–S26–S7–S15–S25* (the *OHcheck* flag is automatically set since there is oxygen and hydrogen in the reaction).

Solutions obtained (only a selection shown):

■ At state *S26* the *P10* procedure adds H₂O to the right side.



Case 2.3: Suppose the reaction reveals a deficiency of oxygen, hydrogen and another element. Eliminating the *NaOH* from the left side, we have

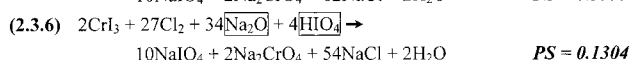
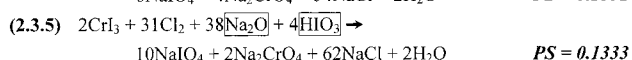
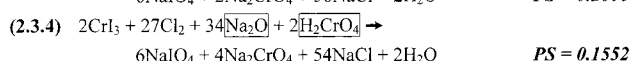
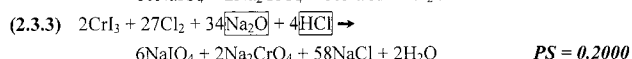
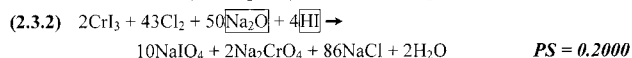
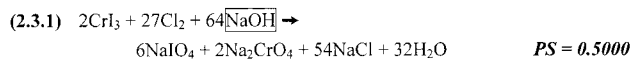


Tracking states: *S2–S6–S13–S21–S32–continues*.

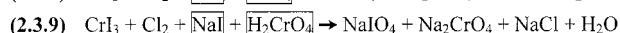
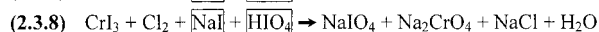
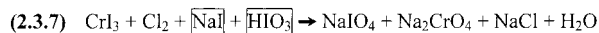
Solution obtained:

■ At state *S32* the *P32* procedure considers the following species: NaI, Na₂O, HI, NaOH, HCl, H₂CrO₄, HIO₄, and HIO₃, discarding other species that would alter the redox state of the reaction (all flags are taken as clear by default, and the *OHcheck* flag is set since there is oxygen and hydrogen present in the reaction).

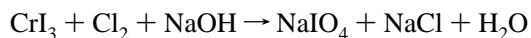
■ Of the solutions being built, the following have the sequence of states: *S32–S7–S15–S25*, a solution being reached in each instance.



■ Among other possible solutions being built, there are those which follow the states: *S32–S7–S16–S26–S7–S26–S1*, and accordingly do not reach a solution due to the fact that no new species can be found to solve the system, since ionic species are not allowed in a complete redox reaction and the *Redoxcomplete*, *Allvalences*, and *Ioncheck* flags are all cleared.



Case 2.4: Suppose the reaction is lacking elements other than oxygen and hydrogen and which will not alter the redox state. Eliminating the Na₂CrO₄ (similar situations would occur if we removed the other species) from the right side, we have

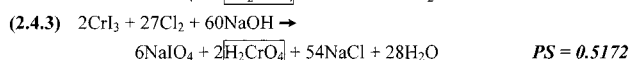
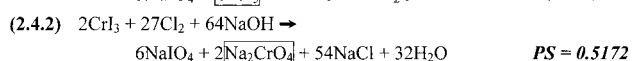
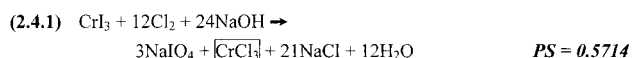


Tracking states: *S2–S6–S13–S21–S32–continues* (all flags are clear by default, except for the *OHcheck* flag, which is set automatically since there is oxygen and hydrogen present in the reaction).

Solutions obtained (only a selection shown):

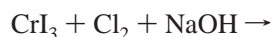
■ At state *S32* the *corresponding* procedure considers the following species, among others, CrCl₃, Na₂CrO₄, and H₂CrO₄, discarding other species that would alter the redox state of the reaction.

■ Of the solutions being built, the following have the sequence of states *S32–S7–S15–S25*, and the following solutions are produced:



In this case it can be seen that of those obtained, the solution showing the highest possibility is not the one we would have expected, since the species CrCl_3 has a greater feasibility than the species Na_2CrO_4 , and both balance the reaction.

Case 2.5: Suppose that in the proposed reaction there are only species on the reactant side, such as

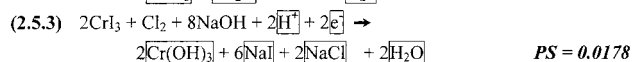
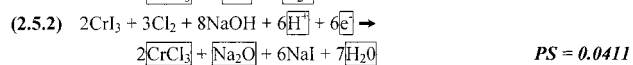
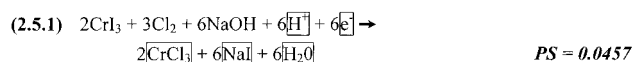


Tracking states: The presence of a species in molecular state means we will always deal with this as a redox process and condition B will thus be satisfied. Hence, the states through which the reaction passes will be $S2-S6-S13-S22-S34$. At this state, an exception is produced due to the fact that no species involving chlorine can be found, since the *Allvalences* flag is clear by default. This exception gives rise to a transition to state $S33$, setting the *Allvalences* flag.

Solutions obtained (only a selection is shown):

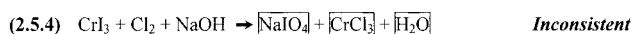
■ At state $S33$ the corresponding procedure considers the following species, among others: CrCl_3 , H_2O , NaCl , HCl , HI , Na_2O , NaI , and $\text{Cr}(\text{OH})_3$.

■ Of the solutions being built, the following have the sequence of states $S33-S3-S8-S17-S28-S37-S17-S27-S36-S27-S35$, since the reaction is treated as a half-reaction. A few of the possible solutions are shown below:

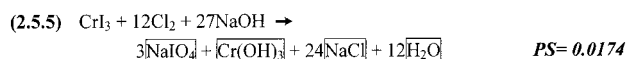


If the *Redoxcomplete* flag were set, then instead of instantiating state $S37$, state $S38$ would be presented, calling procedure $P38$ to find species for the completion of the redox reaction in the previously obtained solutions. The effect of procedure $P38$ would produce solutions with a complete redox reaction and require the addition of many species, so its possibility would be extremely low.

Here, the introduction of other species that would be considered at state $S33$, such as NaIO_4 , NaIO_3 , Na_2CrO_4 , $\text{Na}_2\text{Cr}_2\text{O}_7$, NaClO_2 , NaClO_3 , CrCl_2 , CrI_2 , NaH , CrH_3 , and ICl , will lead to the proposal of solutions that produce inconsistency, such as



Solutions are also produced that may be solved through the states $S33-S3-S7-S16-S26-S7-S15-S25$, if H_2O is not considered initially and is added to state $S26$, or just by the sequence $S33-S3-S7-S15-S25$, such as



5. DISCUSSION

Balancing of reactions may be carried out relatively simply using analytical methods as long as the reactions are consistent; that is, there is no requirement for the consideration of any new species in the reaction.

When dealing with incomplete redox reactions, the consideration of new species is limited to a brief list comprising water, protons, hydroxyls, and electrons, so software to solve such reactions need not be excessively complicated, provided we wish to solve them as redox half-reactions.

In other cases, both when we wish to solve a reaction as a complete redox reaction and when there is a deficit of matter on one or both sides of the equation, the process for completion and solution of such reactions calls for more complex balancing models.

The present article proposes a model that enables the completion and balancing of many inorganic reactions. This model is based on a knowledge frame which takes into account not only the features of the system and the species it contains, but also structural disparity and molecular feasibility parameters, used as fuzzy complexity measures for new species in order to propose the most structurally simple solution to the reaction.

Although the model is capable of solving most incomplete inorganic reactions, it does present a few problems, such as the following:

- If there is a great deficit of matter, for example, if no species are present on either side of the equation, the solutions proposed are of low possibility, perhaps logically so when bearing in mind the paucity of initial information.
- Some solutions that are given priority due to their structural simplicity will not necessarily be the most chemically correct.

- Certain proposed solutions will be chemically unstable. The model does not consider the stability of species in the presence of other species or with respect to the physico-chemical nature of the medium.

- The model proposes neither complex nor hydrated species as possible solutions in the process of completing incomplete systems.

- The set of species that the model may propose as the solution to an incomplete reaction is determined not only by the features of the reaction but also by the limitations of the inorganic formulator²⁸ used in this process.

The authors believe that these disadvantages are largely offset by the advantages provided by the proposed model, namely:

- The model is capable of solving any type of complete reaction.

- The model solves reactions with a hidden matter deficit, such as in redox half-reactions.

- The system solves reactions where there is a deficit of matter on either or both sides of the equation, and even when species are completely absent on one side.

- The new species proposed are formulated automatically through a formulation system.²⁸

- The model obtains a set of possible solutions to a problem reaction. These solutions are ordered according to a set of parameters which take into account the complexity of the species considered. This measure of complexity is performed through measures of disparity and feasibility, as proposed in this model. These parameters allow the proposal of new species without the need for huge databases storing known reactions to be used, at high computational cost, for the proposal of deficient species.

- Although some of the solutions proposed for a reaction are chemically unstable, a great deal of information is

provided to the user, thus enabling the model to be deployed in computer-aided teaching in areas such as virtual chemistry systems and laboratories.

Currently, the authors are working on expanding the knowledge frame, adding information to allow the filtering out of chemically inconsistent solutions. This process involves widening the scope of information in the formulation table and system features, in addition to creating a classification of inorganic reactions that should enable proposed species to produce reactions of one of the defined types.

REFERENCES AND NOTES

- (1) Chunshi, G. A New Inspection Method for Balancing Redox Equations. *J. Chem. Educ.* **1997**, 74 (11), 1365–1366.
- (2) Herndon, W. C. On Balancing Chemical Equations: Past and Present. A Critical Review and Annotated Bibliography. *J. Chem. Educ.* **1997**, 74 (11), 1359–1362.
- (3) Olson, J. A. An Analysis of the Algebraic Method for Balancing Chemical Reactions. *J. Chem. Educ.* **1997**, 74 (5), 538–542.
- (4) Subramaniam, R.; Goh, N. K.; Chia, L. S. The Relationship between the Number of Elements and the Number of Independent Equations of Elemental Balance in Inorganic Chemical Equations. *J. Chem. Educ.* **1995**, 72, 894–895.
- (5) Tóth, Z. Balancing Chemical Equations by Inspection. *J. Chem. Educ.* **1997**, 74 (11), 1363–1364.
- (6) Luque Ruiz, I.; Martínez Pedrajas, C.; Gómez-Nieto, M. A. Design and Development of Computer-Aided Chemical Systems: Representing and Balance of Inorganic Chemical Reactions. *J. Chem. Inf. Comput. Sci.* **2000**, 40 (3), 744–752.
- (7) Cramer, R. D.; Patterson, D. E.; Clark, R. D.; Soltanshahi, F.; Lawless, M. Virtual Compound Libraries: A New Approach to Decision Making in Molecular Discovery Research. *J. Chem. Inf. Comput. Sci.* **1998**, 38 (6), 1010–1023.
- (8) Ellis, L.; Speedie, S. M.; McLeish, R. Representing Metabolic Pathway Information; and Object-Oriented Approach. *Bioinformatics* **1998**, 14 (9), 803–806.
- (9) Hippe, Z. *Artificial Intelligence in Chemistry*; Elsevier: Amsterdam, 1991.
- (10) Hippe, Z. S. Design and Application of an Intelligent Information System SCANKEE for Solving Selected Chemical Problems. *Comput. Chem.* **1998**, 22 (1), 133–140.
- (11) Immirzi, A. Using Symbolic Problems in Teaching Chemistry. *J. Chem. Inf. Comput. Sci.* **1998**, 38 (5), 780–784.
- (12) Karizo, R. W. Mathematica-Assisted Learning in Physical Chemistry. *J. Chem. Inf. Comput. Sci.* **1999**, 39 (1), 96–103.
- (13) Rzepa, H. S.; Tonge, A. P. VchemLab: A Virtual Chemistry Laboratory. The Storage, Retrieval, and Display of Chemical Information Using Standard Internet Tool. *J. Chem. Inf. Comput. Sci.* **1998**, 38 (6), 1048–1053.
- (14) Tratch, S. S.; Zefirov, N. S. Systematic Search for New Types of Chemical Interconversions: Mathematical Models and Some Applications. *J. Chem. Inf. Comput. Sci.* **1998**, 38 (3), 331–348.
- (15) Satoh, H.; Funatsu, K. Further Development of a Reaction Generator in the SOPHIA System for Organic Reaction Production. Knowledge-Guided Addition of Suitable Atoms and/or Atomic Groups to Product Skeleton. *J. Chem. Inf. Comput. Sci.* **1996**, 36 (2), 173–184.
- (16) Flower, D. R. On the Properties of Bit String-Based Measures of Chemical Similarity. *J. Chem. Inf. Comput. Sci.* **1998**, 38 (3), 379–386.
- (17) Moll, R. Reaction Databases and Synthetic Planning-Combined Application and Synergetic Effect. *J. Chem. Inf. Comput. Sci.* **1997**, 37 (1), 131–133.
- (18) Satoh, H.; Sacher, O.; Nakata, T.; Chen, L.; Gasteiger, J.; Funatsu, K. Classification of Organic Reactions: Similarity of Reactions Based on Changes in the Electronic Features of Oxygen Atoms at the Reaction Sites. *J. Chem. Inf. Comput. Sci.* **1998**, 38 (2), 210–219.
- (19) Satoh, K.; Funatsu, K. A Novel Approach to Retrosynthetic Analysis Using Knowledge Bases Derived from Reaction Databases. *J. Chem. Inf. Comput. Sci.* **1999**, 39 (2), 316–325.
- (20) Tractch, S. S.; Zefirov, N. S. A Hierarchical Classification Scheme for Chemical Reactions. *J. Chem. Inf. Comput. Sci.* **1998**, 38 (3), 349–366.
- (21) (a) The IrYdium Project: <http://ir.chem.cmu.edu/irProject/>, (b) *Enzyme Lab.*, http://jchemed.chem.wisc.edu/JCEsoft/Issues/Series_D/5D1/9901W/abs-9901.html. (c) *Equilibrium Calculator*, http://jchemed.chem.wisc.edu/chem.wisc.edu/JCEsoft/Issues/Series_C/4C1/prog2-4C1.html. (d) *Instrument Simulator*, http://jchemed.chem.wisc.edu/JCEsoft/Issues/Series_99/JCEsoft/Issues/Series_D/1D1/prog3-1D1.html. (e) *Frost Diagram*, http://jchemed.chem.wisc.edu/JCEsoft/Issues/Series_SP/SP18/prog1-SP18.html. (f) *ChemDemos*, http://jchemed.chem.wisc.edu/JCEsoft/Issues/Series_SP/SP8/abs-sp8.html. (g) *GCprog1-5d1.html*. (h) *Chemistry Comes Alive!*, <http://jchemed.chem.wisc.edu/>.
- (22) Luque Ruiz, I.; Cruz Soto, J. L.; Gómez-Nieto, M. A. A Tutorial System for Inorganic Chemical Formulation and Reactions. *Proceedings of the 1994 ACM Symposium on Applied Computing*; ACM Press: 1994; pp 583–587.
- (23) Ritter, D.; Johnson, M. Virtual Titrator: a Student-Oriented Instrument. *J. Chem. Educ.* **1997**, 74 (1), 120–123.
- (24) Luque Ruiz, I.; Cruz Soto, J. L.; Gómez-Nieto, M. A. Inorganic Chemical Knowledge Representation Using Dynamic Data Structures. *J. Chem. Inf. Comput. Sci.* **1993**, 33 (3), 378–384.
- (25) Luque Ruiz, I.; Cruz Soto, J. L.; Gómez-Nieto, M. A. Knowledge Representation through Objects in the Development of Expert Systems for Chemical Synthesis and Reactions. In *Advanced Methods in Artificial Intelligence*; Lecture Notes in Computer Science 682; Springer-Verlag: New York, 1993.
- (26) Sowa, J. F. *Knowledge Representation: Logical, Philosophical and Computational Foundations*; Brooks/Cole: 1999.
- (27) Luger, G. F.; Stubblefield, W. A. *Artificial Intelligence: Solutions and Strategies for Complex Problem Solving*; Addison-Wesley: Reading, MA, 1992.
- (28) Luque Ruiz, I.; Cruz Soto, J. L.; Gómez-Nieto, M. A. Computer Translation of Inorganic Chemical Nomenclature to a Dynamic Abstract Data Structure. *J. Chem. Inf. Comput. Sci.* **1994**, 34 (3), 526–533.

CI0000124