(3) American Chemical Society, Committee on Professional Training. *Minimum Standards Used as Criteria in Evaluating Undergraduate Professional Training in Chemistry*; American Chemical Society: Washington, DC, 1960; p 6. *Ibid.*; American Chemical Society: Washington, DC, 1962; p 4.
(4) American Chemical Society, Committee on Professional Training. *Minimum Standards Used as Criteria in Evaluating Undergraduate Professional Education in Chemistry*; American Chemical Society: Washington, DC, 1965; p 4.
(5) American Chemical Society, Committee on Professional Training. *Objectives and Guidelines for Undergraduate Programs in Chemistry*; American Chemical Society: Washington, DC, 1972; p 9.
(6) American Chemical Society, Committee on Professional Training. *Undergraduate Professional Education in Chemistry: Criteria and Evaluation Procedures*; American Chemical Society: Washington, DC, 1977; p 16.
(7) American Chemical Society, Committee on Professional Training. *Undergraduate Professional Education in Chemistry: Guidelines and Evaluation Procedures*; American Chemical Society: Washington, DC, 1983; pp 10 and 19 with four-page Chemical Information Retrieval Appendix.
(8) *Chemical Information Retrieval* (prepared with the help of the ACS Division of Chemical Information); American Chemical Society: Washington, DC, 1983.
(9) Brasted, Robert C., et al., Eds. *Guidelines and Suggested Title List for Undergraduate Chemistry Libraries*; American Chemical Society: Washington, DC, 1982 (under revision).
(10) American Chemical Society. *Tomorrow: the Report of the Task Force for the Study of Chemistry Education in the United States*; American Chemical Society: Washington, DC, 1984; pp 45, 48.
(11) American Chemical Society, Division of Chemical Information, Education Committee. *An Introduction to Computer Searching, STN International Version, 1989*; *An Introduction to Computer Searching, DIALOG Version, 1989*; *An Introduction to Patent Searching, STN International Version, 1989*; *An Introduction to Patent Searching, DIALOG Version, 1989*; all available only from Arleen Somerville, Carlson Library, University of Rochester, Rochester, NY 14627.
(12) An example is STN Mentor, a family of computer-based training tutorials designed to explain how to search for information on databases available on STN International. For example, two STN Mentor disks provide introductory training for CAS Online databases and for author searching on all STN databases.
(13) Clearinghouse for Chemical Information Instructional Materials, c/o Carol Carr, Chemistry Library, University of Pennsylvania, Philadelphia, PA 19104.
(14) LOEX (Library Orientation Instruction Exchange) Clearinghouse, Eastern Michigan University, Ypsilanti, MI 48197.
(15) Heideman, Linda K.; Wiggins, Gary D. *Chemical Literature Guides and Aids: Bibliography and Index*; Indiana University Chemistry Library: Bloomington, IN, 1984.

An Oxidation Number Assignment Expert for CHEMPROF

ARTHUR A. EGGERT* and CATHERINE MIDDLECAMP

Chemistry Tutorial Program, University of Wisconsin, Madison, Wisconsin 53706

ELIZABETH KEAN

Center for Curriculum and Instruction, University of Nebraska, Lincoln, Nebraska 68588

Received January 2, 1990

To construct a computer program module for CHEMPROF capable of tutoring students in the assignment of oxidation number to inorganic compounds, we have studied the constraints on oxidation number assignment and developed a model that is both reliable for computer implementation and comprehensible by students. The model strives to precisely define and circumscribe this domain and to give it a complete declarative description. This rule-based approach will allow computer-aided instruction to expand beyond the limitations of predefined drill and prescripted solutions to problems.

INTRODUCTION

The study of chemistry is the study of the ownership of electrons. In every chemical reaction, that ownership changes. This "ownership" is traditionally represented in terms of the oxidation number of the atom. Chemistry students are then taught the principles of electron ownership and transfer in terms of oxidation–reduction reactions.

The assignment of oxidation numbers, however, has always been somewhat arbitrary. In fact, only for elements and ionic compounds does this ownership number closely approximate an integer. In other cases, electrons are assigned to atoms to which they do not completely belong. Although not chemically exact, such assignments permit the novice to grasp why reactions are occurring and allow the expert to manipulate reaction equations without excessive syntactical baggage.

Several years ago we began work that would establish the basis for CHEMPROF,[1,2] an intelligent tutoring system for general chemistry students. To develop such a tutoring system, domain experts that could solve problems through the application of chemical principles were first needed. Such experts also had to produce solution traces that the computer could then use to teach students how to solve particular problems. The assignment of oxidation numbers was selected as a useful and manageable topic to tackle first. Not unexpectedly, though, in attempting to teach a computer to assign oxidation numbers we found the topic to be more complex than it at first appeared, as the rules commonly used to teach assignments[3-5] were in part incomplete or inconsistent.

The primary problem we faced was that of declarative incompleteness of the domain. This problem in the development of expert systems had already been identified in earlier work such as DENDRAL (which interpreted low-resolution mass spectra)[6] and MYCIN (which identified microbiological organisms).[7] Part of the difficulty arises from the definition of terms. Human experts in a particular field use terminology that is clear to them, but which may be difficult to precisely define for the benefit of nonexperts. Were several experts to supply definitions, these most likely would not be fully compatible. To quibble over any differences may seem like hairsplitting to those unfamiliar with formal logic, but if the inconsistencies cannot be resolved, then a computer program cannot be written to work in the affected area of knowledge. It is, in fact, this lack of clear definition that has caused some chemists to regard oxidation numbers as a questionable parameter. This situation is unfortunate because it is hard to define and work with the concepts of oxidation and reduction without having a means of bookkeeping electrons. Consequently, formalizing the definition of an (inorganic) oxidation

0095-2338/90/1630-0181$02.50/0

number is essential to properly defining more complex ideas. Moreover, the best formalization is in terms of the procedural rules that are used to establish it.

Another part of this incompleteness problem therefore involves the rules and their precedences in application. Declarative rule statements in the predicate calculus or its representational equivalents require the establishment of clauses that match the current problem state and actions that allow the manipulation of the problem state to achieve a new state which is closer to the desired goal state. While one can trivialize some domains by establishing rules that constitute an exhaustive case analysis, the size of the domain of legal inorganic compounds makes such an approach impossible. As a result, it is necessary to formulate more general rules, each of which applies to a subset of the domain. Clearly, for efficiency purposes, one also would want to minimize the number of rules through constraint relaxation and clause disjunctions (OR conditions). Second, since rule application is essentially sequential, one must establish an order of precedence among the rules, so that the more specialized rules are applied before the more general rules. It is in attempting to establish such rules and precedences that the complexity of the computer expert becomes apparent. For example, assigning oxidation numbers in an "obvious" case like MgF_2 cannot be generalized into a rule that would assign each atom its most common value in chemicals of the form of MA_2, because such a rule would fail for CaO_2. In the same way, what is true for PCl_3 does not hold for NBr_3, even though these compounds have identical parsing characteristics, i.e., one trivalent nonmetal bonded to three univalent nonmetals.

It is precisely this type of difficulty that has caused problems such as the stacking of a handful of blocks[8] and the playing of board games designed for six-year-old's[9] to turn into major research projects to develop methods for handling situations where there is a significant interaction among general and specific constraints. The oxidation expert attempts to extend the research done in such "toy worlds" into a real domain which is circumscribable but also not completely axiomizable. Moreover, since the oxidation expert will serve as the basis of a tutoring system, its development not only had to formalize a set of rules and establish an order of precedence but also had to employ rules that could be understood and emulated by relatively naive students.

The detail of the problem can best be understood by examining oxidation number assignment rules as they exist in most textbooks. Atoms in an element usually are first assigned as zero. Next, metals that are invariant in oxidation number within compounds (i.e., group 1A and 2A elements and fluorine) are assigned. Hydrogen and oxygen are then assigned as +1 and −2 respectively, noting that there can be exceptions. Sometimes a rule to assign the most electronegative element by subtracting 8 from its group number (rule of 8) is introduced. Finally, whatever atom remains is assigned a value such that all oxidation numbers sum to zero for compounds or to the charge for ions (Sum rule).

To an experienced chemist these "rules" might be clear enough to be workable, but to a student (and to a computer) they are inadequate and confusing. For example, how are metals with multiple oxidation states handled? When is H assigned as −1? And how are peroxides and superoxides recognized? Furthermore, there is no natural way to teach the progression from simple compounds to more difficult compounds such as $CuSO_4$, KSCN, and $SOCl_2$.

After considerable effort to implement the "usual" approach, even including a library of special cases, we were unable to produce a reliable expert which could solve the oxidation assignment problems that a general chemistry student was likely to encounter. It was at this point that we began developing the set of rules and teaching strategies hereafter described.

METHODS

Development of Assignment Rules. To assign oxidation numbers, we started by establishing two working rules. **Rule 1**: Oxidation numbers are small integers. **Rule 2**: All atoms of one element in a compound have the same oxidation number. While both of these rules had to be amended later to cover all cases, they express general enough principles to provide the framework to develop the initial parts of the assignment procedure. We also decided to classify the chemical to be assigned by the *number* of elements present, rather than by *which* elements are present. Thus, the presence of one, two, or more than two elements is used for the initial sorting. This decision provides the starting point to reach a solution, since it is a straightforward matter for either the computer or student to count the number of elements. With these points in mind, we set about establishing a procedure that contains the steps that we eventually reordered in a way that can teach students to assign oxidation numbers.

The one-element case, although simple, cannot be neglected. It involves three subcases: monatomic elements, polyatomic elements, and simple ions. By the fundamental definition of oxidation numbers,[10] monatomic elements, such as Na, Ca, and Mn, are, of course, assigned an oxidation state of zero. Each atom in a polyatomic element, such as Cl_2 and P_4, is likewise assigned an oxidation number of zero. Finally, the atoms in ions such as Na^+, O^{-2} (Note: CHEMPROF represents O^{2-} as O^{-2} because it automatically superscripts and subscripts as the student types, and it uses the presence of the sign as a flag to adjust the line position), and Hg_2^{+2} are assigned as the charge on the ion divided by the number of atoms.

It may not be immediately apparent why the subcases of monatomic and polyatomic elements must be differentiated. The reasons are twofold. First, students need to know which elements in nature (eight of them) occur as polyatomics, and they need to be corrected if they write the elemental representations incorrectly. As a consequence, the CHEMPROF expert also has such a table of elements and their multiplicity. Second, students sometimes misassign one of the subgroups but not the other. Thus, CHEMPROF needs to keep these groups separate to later remediate student difficulties.

The method of assigning one-element ions (i.e., take the charge on the ion divided by the number of atoms) stems from an important underlying principle: when only one element is left, the oxidation state of each atom of that element can be found by dividing the total charge by the number of atoms present. This is an essential component of the Sum rule, i.e., that the sum of the oxidation numbers for all atoms in a compound or a polyatomic ion must sum to zero or the charge on the ion, respectively. When all but one of the elements have been assigned oxidation numbers, the element remaining can then be treated as if it were a one-element species possessing the amount of charge which is necessitated by the other assignments already made. This approach allows one to handle such ions as Hg_2^{2+} and N_3^- as well.

Species such as N_3^-, however, forced us to modify our initial rules in a manner such that certain elements are permitted to have fractional oxidation states. The number of these species is small and can easily be handled by creating a legal list of exceptions: N^{3-}, I_3^-, I_5^-, I_7^-, alkali superoxides (O_2^-), and transition-metal compounds of the form M_3O_4 and M_3C. In the teaching portion of CHEMPROF, these are introduced as a subtopic under binary species, as will be mentioned subsequently. Nevertheless, the concept of reducing a compound or polyatomic ion to a one-element problem that can

be solved by simple division is important for implementing the assignment of oxidation numbers into a computer expert.

After examining the one-element case, we have the following rule additions and modifications. **Rule 1***: Oxidation numbers are small integers, except for special listed cases. **Rule 2***: When only one element remains to be assigned, the oxidation number of its atoms is the total charge that must be assigned divided by the number of atoms present. (Oxidation numbers determined by this rule must be "legal", that is, be known to occur in nature.) **Rule 3**: Except for N_2, O_2, O_3, F_2, P_4, S_8, Cl_2, Br_2, and I_2, all elements occur in nature as single atoms. (Note: CHEMPROF also accepts sulfur as S because of common usage.)

The two-element case can be divided into six subcases, depending upon whether the species is a compound or an ion and upon which type of elements the species contains: a metal (either one that is invariant in oxidation state or one that can take multiple oxidation states), a metalloid, a nonmetal, or specifically the nonmetal hydrogen. Assigning oxidation numbers to binary compounds depends on applying rules that will allow rapid classification of the compound.

In the first subcase, where the compound consists of an invariant metal cation with either a nonmetal or metalloid (e.g., $AlCl_3$ or Na_3Sb), assigning oxidation numbers is dependent upon *recognizing* which metals have only one oxidation state in their compounds. These metals can be most easily remembered by their position on the periodic table, yielding **Rule 4**: The group 1A (alkali) metals are always +1; the group 2A (alkali earth) metals and Zn and Cd are always +2; and the group 3B metals and Al and Ga are always +3. (A few other elements, such as Lu and Th, are also invariant, but these are not usually encountered in general chemistry. Fluorine need not be included as a special case here as it will be treated later.) Once the invariant metal is assigned, the oxidation number of the other element can be found by **Rule 5**: The sum of the oxidation number of all the atoms in a compound must equal zero (the Sum rule).

In the second subcase in which the compound consists of a metal that can have multiple oxidation states and a nonmetal or metalloid (e.g., CuS or MnAs), the anionic species must be assigned first. Since all anions except fluorine take multiple oxidation states, a new rule must be introduced. **Rule 6**: The oxidation number of a nonmetal (or metalloid) in a species with several elements remaining to be assigned is its group number in the periodic table minus 8 (the rule of 8). For MnO_2, this allows oxygen to be assigned as −2 (i.e., 6 − 8, by the rule of 8) and Mn to be assigned as +4 (by the Sum rule).

In the third subcase where two nonmetals are present, students tend to misassign oxidation states when one of the nonmetals is not oxygen. The rule of 8 will fail if it is applied on the incorrect element, as may happen with NI_3. A simple modification of the rule, however, will prevent such errors and greatly extend its usability. Including the concept of electronegativity yields **Rule 6***: The oxidation number of the most electronegative element in a species with several elements remaining to be assigned equals its group number in the periodic table minus 8 (the rule of 8). The ordering for electronegativity is F > O > Cl > N > Br > I > S > C > Se > P.[11]

In the fourth subcase, where the species is an ion rather than a compound (e.g., VO_2^+ or PO_4^{3-}), the only difference is that the sum of the oxidation states does not equal zero, but rather the charge on the ion. This can be handled by a minor change to Rule 5, yielding **Rule 5***: The sum of the oxidation states of all the atoms in a compound or ion must sum up to zero or the charge on the ion, respectively (the Sum rule).

In the fifth subcase, the compound contains hydrogen. Although hydrogen is a nonmetal, it cannot simply be placed into the nonmetal sequence because the rule of 8 cannot be applied to it. However, hydrogen can be handled in a straightforward manner by creating a new rule that will allow it to be assigned at the outset of the problem. Hydrogen is always assigned as +1 when a nonmetal is present and otherwise as −1.[12] (The presence of metalloids is irrelevant to the rule.) This allows the systematic handling of even such complex compounds as $AuHCO_3$, as well as the correct assignment for H_2Te, and yields **Rule 7**: If hydrogen is present in a compound, assign it as +1 whenever a nonmetal is present in the chemical; otherwise, assign it as −1 (the Hydrogen rule).

Finally, in the sixth subcase binary compounds that can have elements with fractional oxidation states are handled (e.g., NaN_3 and Fe_3O_4). While it is possible to assign integer oxidation states to each atom in these compounds, this would require the abandonment of Rule 2. Without this rule, both computers and students alike would face a much more difficult task in making assignments. Since the primary purpose of defining oxidation numbers, at least at the level of general chemistry, is to observe oxidation-state changes in redox reactions, the concession of fractional oxidation states in a limited number of cases appeared preferable to contorting the rest of the rules.

When compounds contain more than two elements, the rules in most general chemistry texts are incomplete or break down. (This conclusion is based on a survey of 10 general chemistry textbooks.) KSCN well illustrates the difficulties encountered. Most students would have no trouble assigning K as +1, but then would then be left with N, C, and S. If their text included a rule of 8 (which most do not) *and* if their text mentioned that this could be applied more than once (which to our knowledge none does), then the assignments could be made in a straightforward manner. CHEMPROF's oxidation expert, on the other hand, has no trouble with this compound because it assigns the nonmetals N (−3) and S (−2) in that order and then uses the Sum rule to obtain +4 for C. In fact, most compounds, even ones like PH_4VO_3 and $PSCl_3$, can be handled without having to resort to any further rules or special cases.

Some compounds do require another rule, however, for example, US_2WO_4, $MnPO_4$, and NH_4NO_3. In the first case, there are two metals with multiple possible valences; in the second, a metal and a nonmetal with multiple possible valences; and in the third, one element with two valences. The solution to these compounds is to "split" out of them known polyatomic ions (i.e., WO_4^{2-}, PO_4^{3-}, and NH_4^+) and then to solve the parts of the compound separately. This yields **Rule 8**: When a recognizable group of elements of known charge is part of a compound, "split" out this group as if it were an invariant element and then assign the elements within it separately (Splitting rule).

When the CHEMPROF oxidation expert assigns a compound, it examines the chemical formula and always assigns any groups that are in parentheses, not initially looking for other groups. However, after obtaining a solution, it then tries to find other groups that were not bracketed. If any are found, the expert resolves the problem and compares the assignments made both ways. If they are the same, a trace of both is furnished to the teaching logic; if not, or if one element has two different oxidation numbers, the initial trace is marked as inaccurate, and only the last one is used for teaching.

Even with the Splitting rule, it is possible that a compound cannot be completely assigned. Although this occurs only rarely, it can happen when several elements with multiple oxidation states cannot be isolated into different sections of the molecule. In such cases, the CHEMPROF oxidation expert uses trial and error, that is, the expert assigns one of the elements its most common oxidation state and then computes that of the other by the Sum rule. (The oxidation expert

Table I. Oxidation Number Assignment Rules

Step 1:	**Reject** organic compounds. Any compound that contains carbon which is not part of CO_3, CS_3, CO, CS, or CN and that is not a metal carbide is organic.
Step 2:	**Reject** some inorganic compounds. Any compound with B_n ($n > 1$) or with He, Ne, or Ar or with Kr, Xe, or Rn which is not a fluoride is rejected.
Step 3:	Whenever only one element remains to be assigned, **divide** the charge that results from subtracting the oxidation numbers of the previously assigned atoms from the original charge on the chemical **by** the number of atoms present to get the oxidation state. **Check** its legality against the periodic table. Molecules in the elemental form and some elemental ions must have a proper multiplicity of atoms.
Step 4:	**Assign** hydrogen as +1 if any nonmetals are present and −1 otherwise.
Step 5:	**Split** any bracketed or recognizable unbracketed group from the rest of the chemical. If the group charge is known and invariant, **assign** it, and **assign** the charge required by the Sum rule to the rest of the chemical. **Solve** each part separately. If the group charge is unknown or could have several values, **assign** oxidation numbers to the rest of the chemical and apply the Sum to find the charge on the group before solving it.
Step 6:	**Assign** invariant metals, including group 1A (Li, Na, K, Rb, Cs, Fr) as +1, group 2A (Be, Mg, Ca, Sr, Ba, Ra), zinc, and cadmium as +2, and group 3B (Sc, Y, La, Ac) metals, aluminum, and gallium as +3.
Step 7:	**Assign** the most electronegative nonmetal (F > O > Cl > N > Br > I > S > C > Se > P) or subsequently the most electronegative metalloid (At > Te > Po > As > B > Sb > Si > Ge) remaining by the rule of 8. **Repeat** this rule as necessary.
Step 8:	**Assign** trial values to the multivalent metallic elements that remain based on the periodic table until legal values are assigned to all or no legal combination can be found.

actually knows something about which element to try to assign first, but this is not essential for the student to know.) If the combination is legal, it is kept. If not, another value is tried until all reasonable possibilities are exhausted. Groups such as MnO_4 that can have multiple charges (−1, −2, or −3) are treated the same way, being assigned last rather than first as is done with groups whose charge is invariant.

Excluding Inappropriate Compounds. The discussion thus far has addressed how to assign oxidation numbers to a variety of different types of compounds. It is also of great importance, however, for the expert to reject certain types of compounds. For example, students need to be informed when they enter chemicals for which oxidation numbers cannot properly be assigned. The CHEMPROF expert must also deal with chemicals that are feasible, yet beyond its scope.

Entries that are judged inappropriate to assign generally fall into five classes. The first consists of chemicals that cannot exist (e.g., Na_2Cl) because one or more of the elements has an oxidation state that is not feasible. The expert detects these by knowing allowable oxidation numbers (i.e., Cl cannot be −2) and then asks the student to recheck the formula typed. The second class consists of chemicals that are structurally too complex for the expert (and for most general chemistry students), such as $B_3CH_6^-$. Compounds containing more than a small number of boron atoms are therefore excluded, and other thresholds could be implemented if desired. The third class contains the noble gas salts. The CHEMPROF expert currently recognizes only fluoride salts of krypton, xenon, and radon as legal, and these gases clearly require special-case treatment. Fourth, there are compounds in which elements would have legal oxidation numbers, but just do not exist [i.e., $Fe(NO_3)_2$]. CHEMPROF currently accepts and solves such compounds. Practically speaking, students (and even instructors) may not realize that the compounds do not exist; moreover, just because some compounds have not yet been reported, does not mean that they do not exist. The final class is organic compounds. The rules developed for the oxidation expert are not able to adequately handle organic compounds because the oxidation numbers of atoms in organic compounds are dependent on their structure and lack meaning in most contexts.

To reject organic compounds, the method had to be established to determine whether or not a given chemical is organic. For example, H_2CO_2 (formic acid) is "organic", but H_2CO_3 (carbonic acid) is not; $AlC_3H_5O_3$ (aluminum glycol) with three carbon atoms is organic, but $Al_2(CO_3)_3$ (aluminum carbonate) with three carbon atoms is not. The determining factor seems to be that if a carbon–carbon bond or a carbon–hydrogen bond exists, the compound is organic. However, exceptions, such as HCN (which is generally regarded as inorganic), exist. Since it is not clear from the numbers of atoms of an element present or the atomic ratios whether or not a compound is organic, a series of screening rules is needed to make this determination. Identifying these rules was the second major part of the oxidation expert development. Numerous rule schemes interrelating the number and types of elements were tried, with the following proving to be the most reliable.

Organic compounds can be identified by two mechanisms. The first is by simply scanning the formula, left to right, in an effort to identify organic element combinations. The following rules were found to be significant. **Ammonium rule**: NH_4 combinations are considered as H in formula scanning. **CH Rule**: All C_nH_m combinations (e.g., CH_3OH) imply an organic compound. **Chelate/Carbonate rule**: If three consecutive elements (including their numbers present) are collectively on a list of coordinate-bonding species or carbonate-like species (e.g., HCN, CN, CO, CS, CO_3, CS_3, HCO_3, HCS_3) and the compound has not yet been designated as organic, a che/carb group is marked as present and the element scan is restarted. **Oxalate rule**: If more than one carbon is followed by oxygen, sulfur, or nitrogen (e.g., $Na_2C_2O_4$), the compound is organic. **Formate rule**: If two oxygen or sulfur atoms follow one carbon and any number of hydrogens in a formula (e.g., H_2CS), the compound is organic. **HC rule**: If any element follows a hydrogen–carbon sequence in formula (e.g., HCO_2^-), the compound is organic. In scanning, special attention has to be paid to bracketed groups to prevent misassociations from occurring. Moreover, the rules are applied in order so that coordinate bonding, carbonate or ammonium groups are skipped before the later rules are applied which might classify them as part of an organic compound.

The second mechanism used is counting the number of carbon and metal atoms present and applying the following rule. **Excess Carbon rule**: If more than one carbon is present and no metal is present (e.g., C_2Cl_6), the compound is organic. Together these rules seem to properly identify the organic compounds without misclassifying the inorganic carbon compounds.

Ordering the Assignment Rules. Once the rules were developed for assigning oxidation numbers and for rejecting unwanted compounds, it was necessary to order them and place them into a structure so that they could be applied (iteratively or recursively as necessary) to the chemical of interest. This organization is shown in Table I. The first two steps apply to rejecting compounds for which oxidation assignment is impractical. Step 3 is unusual in that it is a "WHEN" step rather than an "IF" step. It is only applied "when" one element remains to be assigned, but then it is *always* applied, no matter where in the sequence one is. It is therefore a "triggered" step rather than a "sequential" step. This step also carries the

requirement that its result must be chemically feasible. If errors exist at this point, the student may need to be referred to electron configurations and the periodic table to assign the chemical properly.

Step 4, the Hydrogen rule, is placed early in the assignment process, since it is always true and will frequently start the student off in the correct direction. Following this, polyatomic ions should be split out if possible (step 5). There is a strong memory component here. The more polyatomic groups that a student can recognize (along with their charges), the more quickly and reliably the student can assign oxidation numbers. The CHEMPROF oxidation expert likewise utilizes "memorized" information, having a library of common groups and their charges that it consults when splitting compounds or verifying the validity of groups. Note that ions exist (such as MnO_4^{n-}) that can take more than one charge. Such ions should be split out early as well, but cannot be used to assign definite charges to the chemical fragments. Such splitting, however, can lead to the rapid assignment of oxidation numbers to the rest of the chemical which can then be carried back to solve the ion as well.

At step 6 the invariant metals are assigned. Here, too, there is information that must be memorized. If the student knows the first 36 elements on the periodic table plus the elements in groups 1A, 2A, 6A, and 7A, then this step is easy to master. Likewise, step 7 requires a knowledge of relative electronegativity values. By the definition of oxidation number, the electrons, and therefore the negative charge, must be assigned to the most electronegative element. Knowing the order of electronegativity allows the student to calculate oxidation numbers by the rule of 8 iteratively, until step 3 becomes applicable or there are no nonmetals left. Step 7 can also be applied to the metalloids once the nonmetals are exhausted.

Step 8 is only used when all else fails, a rare occurrence when the chemicals encountered in general chemistry are assigned. For example, if the oxidation number expert is faced with several multivalent metals, oxidation numbers are assigned recursively by trial and error until the last element can be legally assigned by step 3. The metal in the lowest numbered group on the periodic table is the starting point, and the expert tries assigning the oxidation numbers as the positive value of the group number. Since this step is rarely used by students, they need to understand only its rudimentary application.

DISCUSSION

Our goal in development has been to make the assignment of oxidation numbers implementable with a high degree of accuracy on a computer. This was accomplished by using a rule-based approach (see Table I). Although similar accuracy could have been obtained by using a scheme with tables of values, this would not have permitted the expert to create a trace of how it assigned the oxidation numbers to a chemical so that the teaching logic of CHEMPROF could in turn explain this to the student.

Similarly, CHEMPROF could have been supplied with a large set of worked-out examples for teaching purposes, which would have eliminated the need for an expert. This, however, would have limited CHEMPROF to those problems in its own repertoire and not permitted it to help students with other chemicals that they might want to assign. Aside from the one-element cases, the CHEMPROF oxidation expert actually solves every chemical that it gives to the students as part of instruction, assessment, or practice. An example of the oxidation expert solution trace is given in Figure 1.

Examining the actual application of the expert can best demonstrate the way in which the assignment steps are employed with complex chemicals. Six examples are shown in

```
1,A0,Na(2SO(4,Formula.
1,A1,3,Multiple elements are present.
1,A2,26,Alternate path for polyatomic group.
1,A7,Na,Single occurrence.
1,A7,S,Single occurrence.
1,A7,O,Single occurrence.
1,A3,+,Processing atoms and bracketed polyatomic groups.
1,E1,Na,Element assigned priority.
1,E0,2,Priority grouping from table.
1,E1,S,Element assigned priority.
1,E0,5,Priority grouping from table.
1,E1,O,Element assigned priority.
1,E0,5,Priority grouping from table.
1,E2,Na,Element selected by priority to assign.
1,E5,+1,Oxidation state assigned from table.
1,E4,2,Number of atoms present.
1,E6,+2,Sum of charges assigned so far.
1,E2,O,Element selected by priority to assign.
1,E5,-2,Oxidation state assigned from table.
1,E4,4,Number of atoms present.
1,E6,-6,Sum of charges assigned so far.
1,E2,S,Last element remaining to be assign.
1,E3,+6,Total charge to distribute.
1,E4,1,Only 1 atom present and gets whole charge.
1,E5,+6,Oxidation state assigned.
1,A4,++,Processing of unbracketed polyatomic groups.
1,P3,SO(4,Unbracketed polyatomic group is noted.
1,E1,Na,Element assigned priority.
1,E0,2,Priority grouping from table.
1,G1,SO(4,Polyatomic group is recognized - do first.
1,E1,S,Element assigned priority.
1,E0,5,Priority grouping from table.
1,E1,O,Element assigned priority.
1,E0,5,Priority grouping from table.
1,G2,SO(4)-)2,Analyzing polyatomic group.
1,G0,SO(4,Group for matching purposes.
1,G5,-2,Charge given to each polyatomic group from table.
1,E2,O,Element selected by priority to assign.
1,E5,-2,Oxidation state assigned from table.
1,E4,4,Number of atoms present.
1,E6,-6,Sum of charges assigned so far.
1,E2,S,Last element remaining to be assign.
1,E3,+6,Total charge to distribute.
1,E4,1,Only 1 atom present and gets whole charge.
1,E5,+6,Oxidation state assigned.
1,G7,SO(4)-)2,Analysis of polyatomic group is completed.
1,G6,-2,Sum of charges assigned so far.
1,E2,Na,Last element remaining to be assign.
1,E3,+2,Total charge to distribute.
1,E4,2,Multiple atoms are present; charge is divided.
1,E5,+1,Oxidation state assigned.
1,ZZ,0,End of trace.
```

Figure 1. Trace from the expert. The trace is the series of the steps the expert passes through to assign a compound. Each step contains four parts: the identification of the expert (integer), what was done in the step (one letter, one digit), the operand involved (string of alphanumerics), and the description of the action taken (present only for debugging).

Figure 2 that include the expert's "bookkeeping" of elements and groups remaining to be assigned, the accumulated charge, and which rule is currently applicable. Steps 3 and 7 are responsible for most of the assignments, with steps 4 and 6 reducing the variables and step 8 as a last resort. While CHEMPROF makes as little use of step 8 as possible, it does so effectively by recursively assigning possibilities until the answer is found or the list is exhausted. However, if students do not learn to recognize ions and their charges, they will find themselves using step 8 far too often. Moreover, if they try to assign a compound such as $CuSO_4$ without splitting out the sulfate ion, they may try to assign +10 to copper by two applications of step 7. Because this is clearly unreasonable, they will be forced to backtrack to their assignment of sulfur in step 7 and work between steps 7 and 8 to find the correct solution. Treating sulfur as a multivalent metal will, in fact, work, and the problem can be solved by trial and error in a straightforward fashion. However, a student will employ far better chemical reasoning by knowing common ions and their charges. CHEMPROF teaching logic knows how to deal with this kind of student difficulty by utilizing the trace from the

Chemical: $HSCN$

Problem: Four nonmetals

Step 4: $\Omega_H = +1$ (SCN^-)

Step 7: $\Omega_N = -3$ (SC^{+2})

Step 7: $\Omega_S = -2$ (C^{+4})

Step 3: $\Omega_C = +4$

Chemical: $MnPO_4$

Problem: Two multivalent elements

Step 5: $(Mn^{+3}\ PO_4^{-3})$

Step 3: $\Omega_{Mn} = +3$ (PO_4^{-3})

Step 7: $\Omega_O = -2$ (P^{+5})

Step 3: $\Omega_P = +5$

Chemical: VO_2ClO_4

Problem: Two groups

Step 5: $(ClO_4^-\ VO_2^+)$

Step 7: $\Omega_O = -2$ $(Cl^{+7}\ VO_2^+)$

Step 3: $\Omega_{Cl} = +7$ (VO_2^+)

Step 7: $\Omega_O = -2$ (V^{+5})

Step 3: $\Omega_V = +5$

Chemical: $NaCu(CN)_2$

Problem: Two metals, one multivalent

Step 5: $(NaCu^{+2}\ CN^{-1})$

Step 6: $\Omega_{Na} = +1$ $(Cu^{+1}\ CN^{-1})$

Step 3: $\Omega_{Cu} = +1$ (CN^{-1})

Step 7: $\Omega_N = -3$ (C^{+2})

Step 3: $\Omega_C = +2$

Chemical: $MnPtF_6$

Problem: Two multivalent metals

Step 7: $\Omega_F = -1$ $(MnPt^{+6})$

Step 8: $\Omega_{Pt} = +4$ (Mn^{+2})

Step 3: $\Omega_{Mn} = +2$

Chemical: Na_3AsS_4

Problem: Metal, metalloid, nonmetal

Step 6: $\Omega_{Na} = +1$ (AsS_4^{-3})

Step 7: $\Omega_S = -2$ (As^{+5})

Step 3: $\Omega_{As} = +5$

Figure 2. Six examples of the application of the oxidation rules to chemicals. These demonstrate, respectively, the assignment of hydrogen and multiple nonmetals, splitting to isolate a multivalent metal, splitting to handle a multiply occurring element, splitting to separate several multivalent elements, assignment by trial and error, and assignment involving a metalloid. (Note: Ω_N means the oxidation number of N, and the elements and groups in parentheses remain to be assigned.)

expert, even though the expert rarely is forced into this situation.

In designing the oxidation number expert that would be consistently correct, it proved essential to build in a knowledge of common ions and their charges and the order of nonmetal electronegativity. It is reasonable to extrapolate the need for such knowledge to students. Without it, each new problem can appear as a special case to the student instead of fitting into an established framework.

While not a formally stated goal, it was essential to create an expert that could run rapidly enough to allow students to work interactively with the tutor. This goal has been achieved on the 286 class of PC machines running off hard disk, with chemicals assigned in 2–3 s. Since part of this time is consumed by loading the executable expert module, better performance can be attained by running off of a RAM disk.

The first and second oxidation number experts have both been used by more than 50 students. The first expert, which tried to work in the manner of most textbook rules, consistently had errors that could be traced to rule conflicts or complex pattern keys which could not be easily discerned by the computer. Repair of one problem frequently only introduced more. The expert was generally unstable to any programming changes. The second expert has fared much better. While occasional errors are still found, these can be traced to programming implementation errors or to missing pieces of information that the expert should have in its dictionaries. For example, if the expert does not know that an element can have a certain valence, then it will not permit that valence to be assigned when it is appropriate.

While a chemist might consider the assignment of oxidation numbers for a single compound to usually be easy for even a novice and some of our more complex examples to be somewhat contrived, one must consider the actual conditions under which both the computer expert and the student are asked to make assignments. The computer expert is useless if it cannot handle the difficult cases since it will rarely be asked to solve the easy problems except those generated by CHEMPROF itself. Instructors frequently give students unusual chemicals to see if the students understand the principles well enough to handle them. If the oxidation expert cannot handle such unusual chemicals, then it is of limited value to the student. Second, students are most commonly called upon to assign oxidation numbers as part of balancing redox equations.

Under such conditions, students may be tempted to search for elements that are likely to change oxidation states between compounds before establishing what the oxidation numbers are within each compound. This frequently causes them to miss changes that they do not expect or to assign values without adequate thought. The discipline of following the rules developed for the oxidation expert should reduce these occurrences.

FUTURE WORK

One question that remains in step 3 and step 8 is how to accept or assign an oxidation number for a species which can only occur under specific conditions. For example, in a compound such as Mn_3C, Mn has an oxidation state not seen in other of its compounds. Similarly, although metals may have an oxidation state of zero in metal carbonyls, it may not be useful to include zero in the expert as a legal oxidation state for metals in compounds. A feasible approach might be to build each possible oxidation state for each element as a program object containing named slots which restrict its applicability. This is being explored as part of a general structure for chemical information which will be built to support CHEMPROF experts.

Since names are frequently dependent on oxidation states, the general framework for constructing the oxidation expert also appears to be applicable to building naming experts. Given a chemical input, these experts would either convert inorganic names to formulas or formulas to names. Such experts, which would need to allow for a certain amount of spelling abberration by the students, would also be effective tools for student communication with CHEMPROF. It is also our intent to build an expert to solve redox equations.

ACKNOWLEDGMENT

We acknowledge an equipment grant from IBM under Project Trochos which has given us tools to develop both this expert and CHEMPROF as a whole.

REFERENCES

(1) Eggert, A. A.; Kean, E.; Middlecamp, C. CHEMPROF—An Intelligent Tutoring System. *Proceedings of the Tenth Biennial Chemical Education Conference*; ACS Division of Chemical Education: West Lafayette, IN, 1988.
(2) Eggert, A. A.; Middlecamp, C.; Kean, E. CHEMPROF—An Intelligent Tutor for General Chemistry. Submitted to *J. Chem. Educ.*
(3) Koln, D. The Chemical Equation Part II: Oxidation-Reduction Reactions. *J. Chem. Educ.* **1978**, *55*, 326-333.
(4) Woof, A. A. Oxidation Numbers and Their Limitations. *J. Chem. Educ.* **1988**, *65*, 45-46.
(5) Holleran, E. M.; Jespersen, N. D. Elementary Oxidation-Number Rules. *J. Chem. Educ.* **1980**, *57*, 670.
(6) Lederberg, J. DENDRAL-64: A system for computer construction, enumeration and notation of organic molecules as tree structures and cyclic graphs. Report No. CR-57029, 1964; NASA, Washington, DC.
(7) Shortliffe, E. H. *Computer-based Medical Consultations: MYCIN*; American Elsevier: New York, 1976.
(8) Doyle, J. A Model for Deliberation, Action, and Introspection. Technical Report AI-TR-581, 1980; AI Lab, Massachusetts Institute of Technology, Cambridge, MA.
(9) Burton, R. R.; Brown, J. S. An Investigation of Computer Coaching for Informal Learning Activities. *Int. J. Man-Machine Stud.* **1979**, *11*, 5-24.
(10) Kauffman, J. M. Simple Method for Determination of Oxidation Numbers of Atoms in Compounds. *J. Chem. Educ.* **1986**, *63*, 474-475.
(11) Dickerson, R. E.; Gray, H. B.; Darenbourg, M. Y.; Darenbourg, D. J. *Chemical Principles*, 4th ed.; Benjamin/Cummings: Reading, MA, 1984; p 227.
(12) Kroschwitz, J. I.; Winokur, M. *Chemistry, A First Course*; McGraw-Hill: New York, 1980; p 430.

Computer Perception of Constitutional (Topological) Symmetry: TOPSYM, a Fast Algorithm for Partitioning Atoms and Pairwise Relations among Atoms into Equivalence Classes

GERTA RÜCKER and CHRISTOPH RÜCKER*

Institut für Organische Chemie und Biochemie, Universität Freiburg, Albertstrasse 21, D-7800 Freiburg, FRG

Received February 12, 1990

An algorithm for the perception of constitutional symmetry in molecules (graphs) is presented, which partitions not only atoms (vertices) but also all pairwise relations among skeleton atoms into equivalence classes. The method works without canonical numbering, essentially by raising the connectivity matrix of the arbitrarily numbered molecule (graph) to its second, third, etc. power and evaluating the entries in these higher order matrices.

When developing a computer program for the machine generation of systematic (IUPAC) names for polycyclic organic compounds,[1] we encountered the more fundamental problem of computer perception of symmetry. For example, how many different pairs of potential bridgeheads are present in the bis- and trissecododecahedranes **1-4** (Figure 1): in particular, which pair is equivalent to which other pair? This question is obviously much harder to answer than the question of how many different kinds of atoms are present in these same compounds. Whereas several computer methods exist for the purpose of partitioning atoms in a molecule into equivalence classes[2-8] and some work has also been done on the detection of the identity of bonds,[4,9] no general method appears to be available to treat pairwise (or even higher) relations among the atoms.[10]

The symmetry properties of a molecule (as well as all other properties) are obviously encoded in its structure; the difficulty lies in the decoding process. Since the structure (more precisely, the constitution) can be represented by a connectivity (adjacency) matrix[11] and since the constitutional symmetry is a very fundamental and simple property, we expected that it could be derived by some simple mathematical manipulation of the connectivity matrix. This turned out to be the case, and we report herein on an algorithm (and the computer program TOPSYM based on it) that finds equivalence classes of atoms and pairs of atoms by a purely mathematical approach. That is, we do not need a canonical numbering,[3,5,6] nor do we need to assign numbers to atomic properties (as was done in a somewhat arbitrary manner in some recent solutions of the atom equivalence problem[8]).

0095-2338/90/1630-0187$02.50/0