Empirical Formulas and the Solid State: A Proposal

William B. Jensen

Department of Chemistry, University of Cincinnati, Cincinnati, OH 45221-0172; *jensenwb@email.uc.edu*

The distinction between an empirical formula and a molecular formula is one of the foundation stones of introductory chemical stoichiometry. A quick survey of ten recently published (1999–2001) introductory textbooks, including one-semester terminal texts, full-sized texts for science majors, and tripartite texts for health science majors, showed that all of them not only covered this subject, but also generally illustrated the distinction between the two types of formulas using the example of a typical organic molecule, such as benzene (i.e., CH versus C_6H_6).[1]

A Missed Opportunity

However, the survey also revealed that, with one exception, all of the texts failed to connect this discussion with their later coverage of nonmolecular solids. Since these solids contain infinitely extended framework, layer, or chain polymers rather than discrete molecules *(1),* the concept of a molecular formula has no meaning for them and an empirical formula must be used instead. Nor is this situation exceptional, since even a brief glance at the most recent edition of A. F. Wells' classic monograph, *Structural Inorganic Chemistry*, quickly reveals that the vast majority (over 98% by some estimates) of inorganic compounds, intermetallic alloys, and simple substances are nonmolecular solids under ambient conditions *(2).*

Though coverage of solid-state structures in all of the books was minimal at best—amounting in most cases to less than 1% of the total text—it is still surprising that only one author felt compelled to point out explicitly the necessity of using empirical formulas for nonmolecular solids, especially since many of the texts were careful to use such terms as "formula unit" and "unit formula weight" when working stoichiometric problems involving these solids in order to avoid the incorrect implication that they could be assigned a discrete "molecular weight". Regrettably, however, even in the case of this single exception, the issue of molecular versus empirical formulas for solids was incorrectly presented as one of covalent versus ionic bonding[1]:

> For molecular compounds, the empirical and molecular formulas are usually different. For ionic compounds, the actual formulas are almost always identical to the empirical formulas. For example, the actual formula for sodium chloride is NaCl, not Na_2Cl_2, or some other multiple. There is one exception, however, mercury(I) compounds. Notice that the actual formula of mercury(I) chloride, Hg_2Cl_2, is twice the empirical formula.

The true reason for this difference, of course, lies not in the so-called ionicity of NaCl and HgCl, but rather in the fact that NaCl forms an infinitely extended framework structure in the solid state and therefore can only be assigned an empirical formula, whereas HgCl forms discrete molecules in the solid state and can therefore by assigned a meaningful molecular formula, Hg_2Cl_2, as well *(3).* It is simply not true,

as implied by this quote, that covalent bonding always leads to discrete molecules with molecular formulas, and ionic bonding to nonmolecular solids having only empirical formulas. Thus, such covalently bonded substances as diamond and boron nitride are nonmolecular and can only be assigned empirical formulas, whereas the discrete molecule SF_6 is highly ionic and can be assigned a molecular formula *(4).* This confusion of the issue of covalent-versus-ionic bonding with the issue of molecular-versus-nonmolecular structures is pervasive in introductory textbooks and has been commented on by several authors in the past, though apparently with little effect on the textbook literature *(5–10).*

An Expedient Solution

Given this limited and often incorrect coverage, how is an introductory student, in the absence of a sophisticated feel for periodic trends in bonding and structure, to decide—short of brute memorization—whether a given compositional formula is empirical or molecular in nature? In coping with this problem in my own classes, I have adopted the convention of writing all empirical formulas in quotation marks and all molecular formulas in normal type.[2] This simple act of courtesy relieves the students of a great deal of memorization and guess work. Since virtually the only reason for using an empirical formula for substances of already known composition is the absence of a discrete molecular structure, it also automatically sensitizes them to the pervasive existence of nonmolecular solids without having to greatly expand the already minimal coverage of this topic found in the typical introductory text. All that is required is an explicit statement within the section on solid-state structures that nonmolecular solids have only empirical compositional formulas.

A final added advantage is that this simple convention immediately allows a student to differentiate between the compositional formula of an isolated atom and that of a nonmolecular simple substance—a common source of misunderstanding in introductory chemistry. Thus C stands for the molecular formula of an isolated carbon atom whereas "C" stands for the empirical formula of either the infinitely extended framework structure found in diamond or the infinitely extended layer structure found in graphite. But, a purist may well ask, how can an isolated atom have a molecular formula? This objection brings me to a final point concerning the appropriateness of using the terms "empirical" and "molecular" in describing these two types of compositional formulas.

Some Anachronisms

The adjective "empirical" was originally intended to describe a compositional formula derived directly from experimental data and free of all theoretical or structural interpretations. But, to the best of my knowledge, there is no legitimate use of the adjective empirical that would make

a formula based on experimental gravimetric data and atomic weight values more empirical than one that also makes use of an experimentally determined molecular weight value. In other words, the formula C_6H_6 for benzene is no more theory laden and less empirical than the formula "CH" and the adjective empirical really does not properly differentiate between the two types of formulas *(11)*. Indeed, now that structure can be directly determined using X-ray diffraction data, rather than indirectly inferred from chemical reactivity, as was the case in the 19th century when these distinctions were first made, it is no longer apparent why a structural formula should be regarded as somehow inherently less empirical than a compositional formula.

A More Descriptive Terminology

In reality, all formulas used in chemistry may be classified according to two criteria: they are either compositional or structural in content, and they are either relative or absolute in nature *(12)*. What is currently called an empirical formula is more accurately described as a "relative compositional formula," since it describes only the kinds and relative numbers of atoms present, and what is currently called a molecular formula is more accurately described as an "absolute compositional formula," since it describes both the kinds and absolute numbers of atoms present. Obviously, within the context of this proposal there is no semantic problem with talking about the absolute composition of either a single atom or a polyatomic molecule.

By definition, purely compositional formulas list only the relative or absolute numbers of atoms present, but make no attempt to indicate structure by grouping these atoms into radicals, complex ions,[3] functional groups, solvent of crystallization, and so forth. Imposition of this information on a purely compositional formula results in either a relative or an absolute structural formula, depending on which type of compositional formula is being structurally articulated. Thus the formulas $H(C_2H_3O_2)$ and CH_3COOH for acetic acid are both absolute structural formulas (as are the two- and three-dimensional formulas drawn with bond lines and wedges) since they are structural elaborations of the absolute compositional formula, $C_2H_4O_2$, of acetic acid rather than of its relative compositional formula, "CH_2O". Likewise, the formula, "$(NH_4)(NO_3)$", for the nonmolecular solid, ammonium nitrate, is a good example of a "relative structural formula," since it represents a partial structural elaboration of the corresponding relative compositional formula, "$N_2H_4O_3$". It is, however, important to note that, though relative compositional formulas are, by definition, based on the *smallest* whole number ratios between the atoms, relative structural formulas are often based on a higher multiple of that formula in order to correctly represent the absolute compositions of discrete radicals or complex ions found within an otherwise overall nonmolecular structure. Thus, though the relative compositional formula of sodium oxalate is written as "$NaCO_2$", its relative structural formula is written as "$Na_2(C_2O_4)$", in order to correctly indicate the absolute composition of the discrete oxalate anion. Nevertheless, the stoichiometric coefficient for Na is a relative, rather than an absolute number, representing the 2:1 combining ratio between the cation and the anion for the overall nonmolecular structure of the resulting neutral salt.[4]

Summary and Conclusion

A survey of current introductory textbooks reveals that many authors fail to connect explicitly their discussions of empirical versus molecular formulas with their later discussions of the composition and structure of nonmolecular solids, and that, even when this connection is made, they often imply incorrectly that the issue of empirical versus molecular formulas is really an issue of ionic versus covalent bonding. In addition to calling attention to these shortcomings, the present article proposes that some method, such as the use of quotation marks, be adopted to differentiate unambiguously empirical formulas from molecular formulas, and that the terms "empirical" and "molecular" be replaced by the more descriptive terms "relative" and "absolute." In short, it is argued that our terminology should reflect the information content of the various formulas and not some outdated and largely imaginary distinction concerning their supposed grounding in experimental data.

Notes

1. A list of the textbooks surveyed can be obtained from the author.

2. I originally used italics for relative formulas. However, this method is difficult to employ when writing formulas on a blackboard or overhead. One reviewer suggested underlining the empirical formulas instead, but this is unattractive when it appears on a printed page, and the use of quotation marks appears to be the best compromise between these two extremes. What is important is that some method be used to consistently distinguish empirical compositional formulas, on the one hand, from molecular compositional formulas, on the other, and not the precise nature of that method, which, as far as I am concerned, is still open to discussion and debate.

3. At least one reviewer was insistent that these be called "molecular ions," to distinguish them from simple or atomic ions. Though I have no objection to this in principle, I am bothered by the fact that, in contrast to neutral monoatomic atoms, the neutral molecules corresponding to such ions as NH_4^+, SO_4^{2-}, et cetera, are generally unknown, even as transient species. In addition, the term may be potentially confusing to students, as most discrete molecular ions are usually found in nonmolecular solids rather than in discrete neutral molecular solids. Consequently I have preferred to use the more traditional terms "complex ion" or "polyatomic ion" when referring to these species.

4. This point is important in distinguishing, for example, between the structural formula for mercury(I) chloride, $(Hg_2)Cl_2$, and that for mercury(I) nitrate "$(Hg_2)(NO_3)_2$". The first formula represents a doubling of the relative compositional formula "HgCl" and is written without quotation marks because it is based on the absolute compositional formula of an actual discrete molecule. The second formula also represents a doubling of the relative compositional formula "$HgNO_3$" in order to correctly indicate the structural presence of the diatomic Hg_2^{2+} ion but is written with quotation marks to indicate that it is still a relative formula, since there are no neutral discrete molecules present in the overall structure, which is composed of an infinite array of Hg_2^{2+} and NO_3^- ions combined in a 1:2 ratio. As with the case of ammonium nitrate given in the text, I have found it useful in these formulas to always enclose discrete radicals and polyatomic ions in parentheses, even when they are weighted with an implicit relative stoichiometric coefficient of 1.

Literature Cited

1. Strictly speaking, the term "polymer" is compositional in nature and should be used only to describe the relationship between two species having compositional formulas that are whole number multiples of one another. The question of whether the bonding leading to that polymerization is covalent, ionic, or metallic is irrelevant. While it is true that most nonmolecular inorganic solids cannot be assigned a specific absolute compositional formula, there is no doubt that the structural units present (infinite frameworks, layers, and chains) are multiples of those present in the corresponding high-temperature, albeit transient, gaseous species (which often have molecular formulas identical to the corresponding empirical formulas) and can be described legitimately as polymers of these. Regrettably, for historical reasons, many textbooks continue to apply the term polymer only to covalently bonded, high-molecular weight organic species, even though it has also long been used for predominantly ionic inorganic species in the literature dealing with glass chemistry, cement chemistry, geochemistry, and inorganic polymer chemistry. See, for example, reference 2; Vogel, W. *Chemistry of Glass;* American Ceramic Society: Columbus, OH, 1985; Ray, N. H. *Inorganic Polymers;* Academic Press: New York, 1978; and Wilson, A. D.; Prosser, H. J. *Developments in Ionic Polymers;* Applied Science: New York, 1983.

2. Wells, A. F. *Structural Inorganic Chemistry,* 5th ed.; Clarendon: Oxford, 1983.

3. Mak, T.; Zhou, G. *Crystallography in Modern Chemistry: A Resource Book of Crystal Structures;* Wiley: New York, 1992; pp 473–475. Given that the coordination of Cl is one in the linear Cl−Hg−Hg−Cl molecules, it is unlikely that the Hg−Cl bond is particularly ionic—a conclusion in keeping with the ionic polarization model.

4. Reed. A. E.; Weinhold, F. *J. Am. Chem. Soc.* **1986,** *108,* 3586–3593.

5. Pauling, L. *The Nature of the Chemical Bond,* 3rd ed.; Cornell University Press: Ithaca, NY, 1960; pp 71–73.

6. Laing, M. *Educ. Chem.* **1993,** *30,* 160–163.

7. Lingafelter, E. C. *J. Chem. Educ.* **1993,** *70,* 98–99.

8. Jensen, W. B. *Educ. Chem.* **1994,** *31,* 94.

9. Jensen, W. B. *J. Chem. Educ.* **1998,** *75,* 817–828.

10. Gillespie, R. J. *J. Chem. Educ.* **1998,** *75,* 923–925.

11. Consultation of older chemical dictionaries shows that for most of the 20th century the term empirical was correctly used to describe both types of compositional formula, and that these were, in turn, differentiated from one another by the terms stoichiometric versus molecular. See for example, *The Van Nostrand Chemical Dictionary;* Honig, J. M., Ed.; Van Nostrand: Princeton, NJ, 1953; p 301 and *Hackh's Chemical Dictionary,* 4th ed.; Grant, J., Ed.; McGraw-Hill: New York, 1962; p 278. The term stoichiometric (from the Greek *stoicheion,* meaning "element" and *metron,* meaning "to measure") is, of course, no better than the term empirical in distinguishing between the two types of compositional formula, since both measure the quantity of the elements present, albeit one in a relative manner and the other in an absolute manner.

12. Use of the terms relative versus absolute, rather than empirical versus molecular, was first suggested by Berzelius nearly 170 years ago. See Berzelius, J. J. *Jahres-Bericht* **1832,** *11,* 42.