Reports

An Interpolated Molecular Formula

Abstract. The necessity of counting hydrogen atoms in molecular formulas may be obviated by substituting a "hydrogen reciprocal," which can be obtained more easily.

When molecular formulas are used for searching or filing chemical compounds, handling of the hydrogen atoms proves troublesome. Counting the H's is laborious and is responsible for the majority of errors in these formulas; yet, the sum of the H's does not provide much information about a compound. As a result, this sum has been relegated to the end of the formulas in some collections (*1*) and in others it has been omitted altogether (*2*).

I propose, here, to omit the sum of hydrogen atoms from molecular formulas of covalent compounds and to replace it with the number of rings and the degree of unsaturation of the molecules. It has already been shown that, mathematically, these two expressions are equivalent (*3, 4*).

A few examples of the proposed interpolated formulas are given below, along with their conversion into conventional molecular formulas. In the proposed formulas, the number before the comma represents the number of rings, *R*, and the number after the comma represents the degree of unsaturation, Δ. This choice of two numbers is arbitrary. One alternative is to use a set of three numbers: *f* for the number of double bonds connected to hetero atoms (as in $>C{=}O$); *s* for the number of double bonds linking carbon atoms (as in $>C{=}C<$); and *R* for the number of rings. Such division is a matter of convenience. If the three numbers *f*, *s*, and *R* are used, the expression will be different for keto and enol tautomers. On the other hand, if *R* and Δ (which is $f + s + R$) are added to form a single number [Soffer's ρ (*4*), which here may properly be called a hydrogen reciprocal], the expression will remain unaltered even if the represented compound undergoes glycoside formation.

From an equation derived by Soffer (*4*), the conversion formula is obtained. In its general form, it is Eq. 1:

$$n_H = 2 + \Sigma\, n_v(v - 2) - 2(R + \Delta) \quad (1)$$

where n_v is the number of atoms (except hydrogen atoms) of covalence *v*, the sum of which is taken over all the *v*'s; n_H is the number of hydrogen atoms; *R* is the number of rings; and Δ is the number of double bonds (one triple bond counts for two double bonds; unsaturated linkages in functions are counted too) (*5*).

In most cases, Eq. 1 will reduce to the following:

$$n_H = 2 + 2n_C + n_{N,P} - n_{Hal} - 2(R + \Delta) \quad (2)$$

where n_C is the number of carbon atoms, $n_{N,P}$ is the number of (trivalent) N and P atoms, and n_{Hal} is the number of halogens. For example:

1) For ergosterol, the interpolated formula is $C_{27}O$-4,3. To find n_H, substitute into Eq. 2 the values $n_C = 27$, $R = 4$, and $\Delta = 3$: $n_H = 2 + 2 \times 27 - 2(4 + 3) = 42$. The molecular formula is $C_{27}H_{42}O$.

2) For riboflavin phosphate, the interpolated formula is $C_{17}N_4O_9P^V$-3,8 (see *5*). To find n_H, substitute into Eq. 2 the values $n_C = 17$, $n_N = 4$, $R = 3$, $\Delta = 8$, and $n_{P^V} = 1$: $n_H = 2 + 2 \times 17 + 4 - 2(3 + 8) + 3 \times 1$ ($= 3n_{P^V}$, Eq. 1) $= 21$. The molecular formula is $C_{17}H_{21}N_4O_9P$.

3) Attention must be paid to abnormal valencies; in carbon monoxide, for instance, $2n_C$ does not apply, since $v = 2$ here and the $n_C(v - 2)$ from Eq. 1 becomes 0.

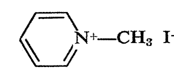
It is seen that the advantages of interpolated formulas are that no information is lost, and that these formulas are more meaningful, more easily obtained, and less apt to contain errors than conventional formulas. Furthermore, interpolated formulas can be obtained for many compounds with unknown structures, as enough data for their calculation are often available at an early stage. For other compounds, the conventional molecular formulas could be filed, without inconvenience, among interpolated formulas. The presence of hydrogen simply would indicate the lack of other information.

ALFRED FELDMAN

Department of Radiobiology, Walter Reed Army Institute of Research, Washington, D.C.

References and Notes

1. J. H. Fletcher and D. S. Dubbs, *Chem. Eng. News* **34**, 5888 (1956); G. M. Dyson, *Chem. & Ind.* (*London*) **1952**, 676 (1952).
2. H. Skolnik and J. K. Hopkins, *J. Chem. Educ.* **35**, 150 (1958).
3. A. A. Pontet, *Chimia* (*Switz.*) **5**, 39 (1951).
4. M. D. Soffer, *Science* **127**, 880 (1958).
5. In the file now in use in our department at Walter Reed, some simplifications have been introduced. For elements that can exhibit several valencies, the "normal" one is recorded (that is, 3 for N and P, 2 for S), provided that the additional valencies consist of double bonds. These double bonds may then be disregarded. Thus, the formula for $CH_3{-}NO_2$ is CNO_2-0,1; for $CH_3{-}SO_3H$ it is CO_3S-0,0, and for riboflavin phosphate it would be $C_{17}N_4O_9P$-3,7. However, in 1-methylpyridinium iodide,



the additional valencies do not consist of a double bond; therefore, the formula is C_6IN^V-1,3.

2 August 1960

Instructions for preparing reports. Begin the report with an abstract of from 45 to 55 words. The abstract should *not* repeat phrases employed in the title. It should work with the title to give the reader a summary of the results presented in the report proper.

Type manuscripts double-spaced and submit one ribbon copy and one carbon copy.

Limit the report proper to the equivalent of 1200 words. This space includes that occupied by illustrative material as well as by the references and notes.

Limit illustrative material to *one* 2-column figure (that is, a figure whose width equals two columns of text) or to *one* 2-column table or to *two* 1-column illustrations, which may consist of two figures or two tables or one of each.

For further details see "Suggestions to Contributors" [*Science* **125**, 16 (1957)].

Syphacia muris, the Rat Pinworm

Abstract. A migration of gravid *Syphacia muris* pinworms down the large intestine of the rat host is shown to occur from the seventh day on in the worm's life cycle. Eggs obtained from migrating worms have proved to be infective to helminth-free rats after incubation in saline for 30 minutes at room temperature and 4 hours at 37°C.

The pinworm, *Syphacia muris*, is a very common cecal parasite of the laboratory albino rat. Few details of the life history of this pinworm are known, however, mainly because previous workers (*1*) have been unable to obtain infective stages of the parasite for use in experimental infections. These investigators reported that eggs liberated from gravid worms taken from the cecum of infected rats would not continue their development in vitro. Chan (*2*), using the closely related mouse pinworm *S. obvelata*, then demonstrated that eggs obtained from gravid female worms migrating down the colon of the host mice at the conclusion of their life cycle would develop to the infective stage. This suggested that