

A Novel Formalism To Characterize the Degree of Unsaturation of Organic Molecules

Martin Badertscher,^{*,†} Kaspar Bischofberger,[†] Morton E. Munk,[‡] and Ernö Pretsch[†]

Department of Organic Chemistry, Swiss Federal Institute of Technology (ETH), Universitätstrasse 16, CH-8092 Zürich, Switzerland, and Arizona State University, Tempe, Arizona 85287-1604

Received September 27, 2000

The existing formalism to calculate the degree of unsaturation from the molecular formula of organic molecules cannot be applied to charged and/or disconnected species. Moreover, the calculated value depends on the assumed formal valence of each of the elements. In this work, we introduce a new formalism that eliminates these problems. The suggested property, degree of unsaturation, can be calculated from the molecular formula as well as from any structural representation of a molecule corresponding to that molecular formula.

1. INTRODUCTION

A molecular formula defines a finite, albeit usually very large number of chemical structures. If the formal valence of all elements involved is known, the sum of the number of rings, double bonds, and twice that of triple bonds is constant for all possible isomers. This sum is usually called “double bond equivalents, *DBE*” or “index of hydrogen deficiency, *IHD*”. The different simplified formalisms found in the literature^{1,2} can be derived from the general equation³

$$IHD = 1 + \frac{1}{2} \left(\sum_i n_i (\nu_i - 2) \right) \quad (1)$$

with n_i the number and ν_i the formal valence of the element i .

The assessment of the molecular formula is usually the first step of the structure elucidation process. The calculation of *IHD* and of the various contributions to it, i.e., the determination of the number of double bonds, triple bonds, and rings is helpful during structure elucidation. Computer programs for automatic structure generation^{4–10} make explicit or implicit use of this information.

One drawback of the index of hydrogen deficiency as defined in eq 1 is that the formal valence of each element must be known. This is not a problem with most of the organic molecules containing only C, H, N, O, and halogens. However, if sulfur is present in a molecule, its formal valence depends on its form and has, for example, different values if sulfur occurs in the molecule as a thioether, a sulfoxide, or a sulfone. Analogous problems occur with phosphorus. Available computer programs for automatic generation of structures, therefore, require the definition of the oxidation state of such elements, an information which may not be known to the user.

Another problem is that some functional groups can be drawn in several ways, with a different number of double bonds. Three drawings of nitromethane illustrate this problem (Figure 1). Today, usually it is drawn as **2** with one double

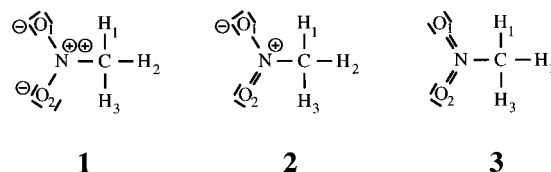


Figure 1. Three resonance structures of the same constitution, leading to the same *DU* of 1, but three different values for *IHD*.

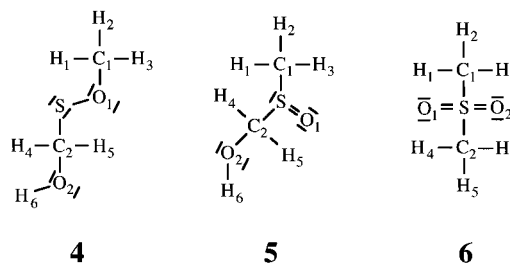


Figure 2. Three resonance structures of three constitutional isomers, leading to the same *DU* of 0, but three different values for *IHD*.

bond. The alternative resonance structures **1** and **3** exhibit zero and two double bonds, respectively. Nowadays, one might reject **3** due to the 5-valent nitrogen atom, but **1** is definitely a valid resonance structure. As a further possibility (not shown), the single O–N bond in the nitro group is sometimes drawn as a coordinative covalent bond, represented by an arrow pointing to the electronegative atom. Another problem is illustrated in Figure 2 where the three structures shown for molecular formula $C_2H_6SO_2$ have 0, 1, and 2 double bonds, respectively. Structure generators usually generate one of these, depending on the predefined formal valence of S. A further constraint of the current formalism is that it cannot handle radicals, ions, and disjoint parts (see below). Advanced structure generators being able to handle these additional structural features require a new formalism to quantify unsaturation because it is no longer clear what it means, when the user of a structure generator requests the presence of a double bond in the structures generated. Structure **2** may be constructed in a particular case. If the presence of a double bond is forbidden, quite the opposite requirement, structure **1** may be generated. **1** and **2** are resonance structures of the same compound. The formalism introduced reflects

* Corresponding author phone: +41 1 632 2904; fax: +41 1 632 1353; e-mail: badertscher@org.chem.ethz.ch.

[†] Swiss Federal Institute of Technology (ETH).

[‡] Arizona State University.

Table 1. Number of Valence Electrons of Elements Formally Partitioned into the Standard Number of Bond Electrons and the Standard Number of Localized Valence Electrons^a

element	no. of valence electrons	std no. of bond electrons b_i	std no. of localized valence electrons l_i
H	1	1	0
Li	1	1	0
C	4	4	0
N	5	3	2
O	6	2	4
halogens	7	1	6
Si	4	4	0
P	5	3	2
S	6	2	4

^a The values in column 2 are the sum of the values in columns 3 and 4.

the new situation. Parameters such as the index of hydrogen deficiency become useless, as there may be candidate structures generated for which they are not defined. In this work a new parameter is defined that allows to quantify the degree of unsaturation for a much larger variety of structural representations. The index of hydrogen deficiency can be derived from the new formalism as a special case for the classical structural variety covered by most of today's structure generators.

2. RESULTS AND DISCUSSION

New Formalism. Here we define as a related new measure, the degree of unsaturation, DU . It can be calculated from the molecular formula and has the same value irrespective of the formal valence of the elements involved. This is made possible by introducing as a further parameter the number of excess localized electrons, ELE . The valence electrons of an element are partitioned into bond electrons and electrons localized on an atom. The number of excess localized electrons of a molecule is the difference between the actual number of electrons localized on all atoms in its representation and the sum of the standard numbers as given in column 4 of Table 1. When calculating the degree of unsaturation, DU , for various forms of a molecule (Figure 1) or various isomers with the same molecular formula (Figure 2), the difference in the number of double bonds will be compensated for by ELE .

A bond has an even number of electrons shared between two atoms. There are "single", "double", and "triple" bonds corresponding to 2, 4, and 6 electrons involved in the bond, respectively. Each atom contributes 1, 2, or 3 bond electrons to the covalent bond. If a structure cannot be represented with the bond types defined, it cannot be handled by this scheme. An aromatic compound must be represented as any resonance structure explicitly showing all bonds. No "aromatic bond" with multiplicity 1.5 is defined.

The electrons localized on an atom are lone pair electrons and, in case of radicals, unpaired electrons. Table 1 lists the values for some of the elements as they occur in the chemical environments most frequently encountered in organic chemistry. The total number of valence electrons is divided into a standard number of bond electrons b_i and a standard number of localized valence electrons l_i . Note that for this scheme, it is not necessary to consider inner shell electrons. The degree of unsaturation, DU , of any particular representa-

Table 2. Calculation of DU from the Molecular Formula CH_3NO_2 ($q = 0$)

element	n_i	b_i	$n_i(b_i-2)$
C	1	4	2
H	3	1	-3
N	1	3	1
O	2	2	0
		$\sum n_i(b_i - 2)$	0
		DU	1

Table 3. Calculation of DU from the Molecular Formula $\text{C}_2\text{H}_6\text{SO}_2$ ($q = 0$)

element	n_i	b_i	$n_i(b_i-2)$
C	2	4	4
H	6	1	-6
S	1	2	0
O	2	2	0
		$\sum n_i(b_i - 2)$	-2
		DU	0

tion of a structure is defined in terms of differences between standard properties of elements as given in Table 1 and the actual properties of the atoms in a particular representation of a molecule. Since the calculation of DU from the molecular formula uses b_i and that for a representation depends on l_i (see below), the total number of valence electrons can be assigned quite arbitrarily, as long as the relation between b_i and l_i is not altered. Along these lines, Table 1 can be easily extended by any other elements.

The degree of unsaturation, DU , can be calculated from the molecular formula as

$$DU = 1 + \frac{1}{2}(-q + \sum_i n_i(b_i - 2)) \quad (2)$$

with q as total charge of the molecule (signed), n_i the number of atoms of element i , and b_i the standard number of bond electrons of element i (cf. Table 1).

As compared to the earlier definition of index of hydrogen deficiency, IHD (eq 1), for DU the actual formal valences of the atoms have been replaced by tabulated standard values, and the charge has been added as an additional contribution. DU is a multiple of $1/2$. It is a whole number for molecules with an even number of electrons. Radicals and ions may lead to odd multiples of $1/2$. For the molecular formulas CH_3NO_2 and $\text{C}_2\text{H}_6\text{SO}_2$ the contributions to DU are given in Tables 2 and 3, respectively.

DU can be calculated from any particular structural representation by itemizing all contributions

$$DU = DB + 2 \cdot TB + RING + (1 - DIS) + \frac{1}{2} ELE \quad (3)$$

with DB as the number of double bonds, TB the number of triple bonds, $RING$ the number of rings, DIS the number of disjoint parts, and ELE the number of excess localized electrons as already defined.

Note that a ring unsaturation unit disappears if a ring is opened at a single bond and two hydrogen atoms are added to the ends. This procedure can be repeated for all ring unsaturation units. If the procedure is applied to a bond not being part of a ring, two disjoint molecules result (also reducing DU by 1). The parameter DIS allows the definition of DU for disjoint parts: Any two atoms A1 and A2 within

Table 4. Number of Excess Localized Electrons, *ELE*, for Each Atom in **1–3**

atom	<i>ELE</i> in 1	<i>ELE</i> in 2	<i>ELE</i> in 3
C	0	0	0
H1, H2, H3	0	0	0
N	−2	−2	−2
O1	2	2	0
O2	2	0	0
total	2	0	−2

Table 5. Number of Unsaturations for **1–3**

<i>DU</i> unit	no. of unsaturations		
	in 1	in 2	in 3
<i>DB</i>	0	1	2
$\frac{1}{2}$ <i>ELE</i>	1	0	−1
total	1	1	1

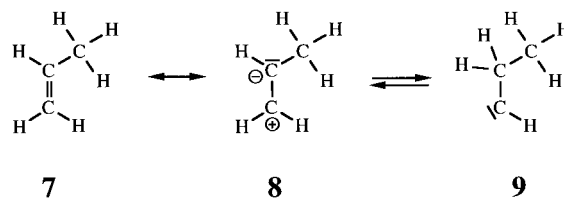
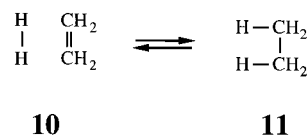
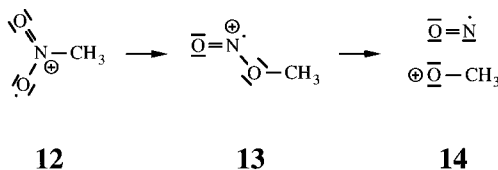
a molecule are connected by at least one path of atoms and bonds A1-...-A2. If no such path can be found between two atoms, they are members of distinct disjoint parts. Only single, double, and triple bonds as defined above are considered “bonds”. A charge separation, as in a salt, is not.

It is usual to assign a formal charge to each atom in a resonance structure, defined as the difference between the number of valence electrons as given in Table 1 and the actual number of valence electrons. The formal charges of all atoms add up to the total charge of the molecule. The formal charge is shown in the examples, though *DU* does not depend on it when the unsaturation units are itemized. The scheme is exemplified in Tables 4 and 5 by assigning the contributions to the resonance structures **1–3** (cf. Figure 1). Equation 3 is reduced to $DU = DB + \frac{1}{2}ELE$ for this example as $0 = TB = RING = (1-DIS)$. In **2** the nitrogen atom has a formal charge of +1. Therefore, only 4 of its usual 5 valence electrons are present. All of them are engaged in bonds. There is no electron localized on this atom. Atom O2 is uncharged. Two of its 6 valence electrons are engaged in the double bond. The 4 remaining localized electrons are shown as two dashes representing lone pairs. Atom O2 has 3 lone pairs. As its formal charge is −1, there is a seventh valence electron, engaged in the single bond. Table 1 shows the standard number of localized valence electrons l_i to be 2 on nitrogen and 4 on oxygen. O1 has 4 localized electrons indeed, while O2 has 6, which is 2 in excess to the standard value. This atom has 2 excess localized electrons. The nitrogen atom has no localized electrons, 2 less than the standard value, a negative excess of −2. The carbon and hydrogen atoms have no localized electrons, in accord with the standard value of 0. The sum of all excess localized electrons, *ELE*, is 0. The number of disjoint parts, *DIS*, is 1, since the atoms specified in the molecular formula are combined to form a single molecule. Using eq 3, the degree of unsaturation, *DU*, is calculated as 1, in agreement with most definitions. The resonance structures **1** and **3** lead to the same value. In **1** a double bond has vanished, when compared with **2**, compensated by 2 more excess localized electrons on O2. The opposite is true for **3**. Note that in contrast to *DU*, the values of the index of hydrogen deficiency, *IHD*, are not the same for the three representations of the compound.

The previous example deals with mesomeric structures belonging to the same constitution. In contrast, Figure 2 shows three different constitutional isomers **4–6** correspond-

Table 6. Number of Unsaturations for **4–6**

<i>DU</i> unit	no. of unsaturations		
	in 4	in 5	in 6
<i>DB</i>	0	1	2
$\frac{1}{2}$ <i>ELE</i>	0	−1	−2
total	0	0	0

**Figure 3.** Formal rearrangement of propene to a carbene. The three resonance structures correspond to the same *DU* of 1.**Figure 4.** Addition of H_2 to ethene. The two representations correspond to the same *DU* of 0.**Figure 5.** Gas phase reaction frequently observed in mass spectrometry. The three representations correspond to the same *DU* of $\frac{1}{2}$.

ing to a molecular formula $C_2H_6SO_2$, leading to $DU = 0$. The contributions are summarized in Table 6. Again eq 3 is reduced to $DU = DB + \frac{1}{2}ELE$ for this example as $0 = TB = RING = (1-DIS)$. Only the sulfur atom shows a contribution from excess localized electrons, being 0, −2, and −4 for the isomers **4**, **5**, and **6**, respectively. As in the first example *IHD* values as calculated from eq 1 are different.

Figure 3 shows an example of a (purely formal) rearrangement involving a carbene **9**. The extra lone pair compensates the double bond present in **7**. *DU* for **7–9** amounts to 1. This is an example of a pure hydrocarbon where the concept of *IHD* fails to yield consistent numbers. **7** has *IHD* = 1 in agreement with *DU*, but the structural representations **8** and **9** display no double bonds and on that basis would be assigned *IHD* values of 0.

DU as Invariant in Chemical Reactions. As *DU* can be calculated from the molecular formula, it does not change in a chemical reaction, as long as all atoms involved are combined in the molecular formula. The addition of hydrogen to ethene is an obvious example (Figure 4). The total molecular formula is C_2H_6 . Both **10** and **11** lead to a *DU* value of 0.

In ordinary electron impact mass spectra the molecule ions of nitro compounds often lose NO corresponding to a mass difference of 30 atomic units. The reaction can be understood as a rearrangement followed by a N–O bond cleavage according to Figure 5. Each representation **12–14** leads to the same *DU* of $\frac{1}{2}$ as can also be calculated from the formula of the radical cation $CH_3NO_2^+$.

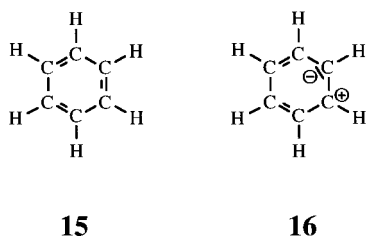


Figure 6. Two resonance structures of benzene leading to the same DU of 4 but different values for IHD .

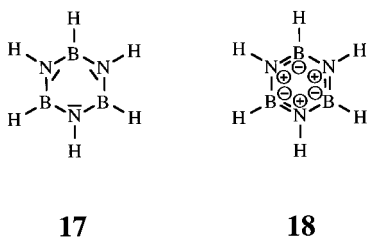
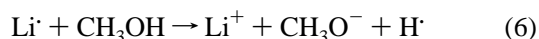
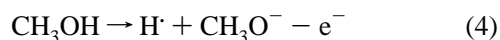


Figure 7. Two resonance structures of borazine, an inorganic analogue of benzene. Both representations correspond to the same DU of 1.

Ambiguous Role of Charge Separation. Charge separations do not explicitly enter the formula itemizing DU . Their appearance may be accompanied by a gain or by a loss of DU units, dependent on the formal mechanism of their formation. Charge separations are mere side effects when electrons are moved within a structure. The following two examples are given for clarification. Figure 6 shows two resonance structures of benzene, one involving a charge separation. Both DU and the index of hydrogen deficiency, IHD , of benzene are 4, as calculated from the molecular formula. They can be attributed to 3 double bonds and 1 ring in the Kekulé form **15**. In **16**, one of the double bonds has vanished in favor of a charge separation. The DU contribution of the former double bond is now found as an additional lone pair on one of the carbon atoms leading to 2 additional excess localized electrons. It is not clear how to calculate IHD of **16**. Equation 1 is not applicable to such a representation. If the number of double bonds and rings is counted, IHD is 3, in contradiction to the Kekulé form.

Borazine (cf. Figure 7) is a compound expressing a certain degree of aromaticity.¹¹ It is possible to draw a resonance structure **18** analogous to the Kekulé form of benzene. However, DU and IHD calculated from the molecular formula are only 1. In **17** the unsaturation unit is devoted to a ring. The 3 additional double bonds in **18** are compensated by the engagement of 3 lone pairs in the bonding, according to a loss of 6 excess localized electrons. Again the formula to calculate the contribution to IHD does not apply to **18**.

Redox Reactions Involving Free Electrons. Occasionally, redox reactions are formulated as a separate reduction and oxidation part, each involving free electrons.



Eq 6 is the sum of eqs 4 and 5, with the free electrons canceled. The respective DU values are also additive.

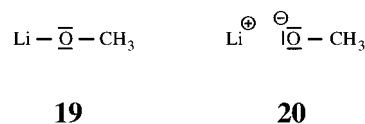


Figure 8. Lithium methanolate may be considered a molecule (**19**) or a salt. In either case DU is calculated to be 0.

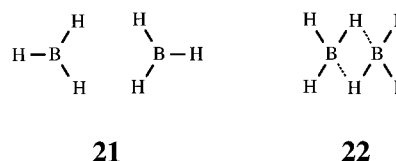


Figure 9. The molecular formula of diborane, B_2H_6 , corresponds to DU of -1 . Two disjoint parts, as shown left, are in contradiction to experiment. A representation involving three-center bonds (**22**) best fits experimental results but cannot be handled by the scheme described in this paper.

Whenever molecules or other atom ensembles are combined, their individual DU values add up to DU of the combination. When calculating DU from the molecular formula, explicit electrons enter with their charge only. When itemizing the contributions, free electrons are not considered disjoint parts. Instead each electron constitutes $1/2$ degree of unsaturation. Therefore, either side of eq 4 corresponds to $DU = 0$, the sides of eq 5 to $DU = -1/2$. Each side of the sum (eq 6), therefore, corresponds to $DU = -1/2$.

The reaction product lithium methanolate may be considered as a molecule **19** or ion pair **20** shown in Figure 8. In either case DU is calculated as 0. In **20** the oxygen has 2 excess localized electrons, but the compound is drawn as 2 disjoint parts.

Limits of the Formalism. Certain compounds cannot be handled employing the scheme of bonds between two atoms. Diborane (Figure 9), with molecular formula B_2H_6 , corresponds to a DU value of -1 . There is no way to draw a constitution comprising a complete molecule. **21** shows the negative unsaturation unit as two disjoint fragments. Three-center bonds are employed in **22** to account for the structural features of the compound. Thus, no contributions to the degree of unsaturation can be itemized for **22**.

3. CONCLUSIONS

The degree of unsaturation, DU , as defined in this work is, in essence, a mathematical construct of electron book-keeping, independent of any graphical representation of a structure. It may serve as a tool in structure generators used in the process of structure elucidation. Advanced structure generators being able to handle an extended structural variety such as charged species cannot employ the index of hydrogen deficiency, IHD , or similar constructs as a tool to quantify unsaturation. IHD may not be the same for all structures generated and, more importantly, it may not be defined for some of the structures. The degree of unsaturation, DU , is the same for all structures compatible with a molecular formula. Hence, DU can be calculated from any structure as well as from the molecular formula. The scheme covers almost all structural features common in organic chemistry. It can handle charged species, single and multiple radicals, disjoint parts, and reactions involving free electrons. It only fails when the concept of covalent bonds between two atoms does not hold, such as in certain boron compounds.

ACKNOWLEDGMENT

The financial support of the National Institutes of Health (Grant GM 37963), Sterling Drug, Inc., and Allied Signal, Inc. is gratefully acknowledged.

REFERENCES AND NOTES

- (1) Hendrickson, J. B.; Cram, D. J.; Hammond, G. S. *Organic Chemistry*, 3rd ed.; McGraw-Hill Kogasusha: Tokyo, 1970.
- (2) Pellegrin, V. Molecular Formulas of Organic Compounds. *J. Chem. Educ.* **1983**, *60*, 626–633.
- (3) Pretsch, E.; Bühlmann, P.; Affolter, Ch. *Structure Determination of Organic Compounds*, 3rd ed.; Springer-Verlag: Berlin, 2000.
- (4) Munk, M. E.; Sodano, C. S.; McLean, R. L.; Haskell, T. H. Actinobolin. I. Structure of Actinobolamine. *J. Am. Chem. Soc.* **1967**, *89*, 4158–4165.
- (5) Sasaki, S.; Kudo, Y.; Ochiai, S.; Abe, H. Automated Chemical Structure Analysis of Organic Compounds: An Attempt to Structure Determination by the Use of NMR. *Mikrochim. Acta* **1971**, 726–742.
- (6) Shelley, C. A.; Munk, M. E. CASE, a computer model of the structure elucidation process. *Anal. Chim. Acta* **1981**, *133*, 507–516.
- (7) Dubois, J. E.; Carabédian, M.; Dagane, I. Computer-aided elucidation of structures by carbon-13 NMR. *Anal. Chim. Acta* **1984**, *158*, 217–233.
- (8) Christie, B. D.; Munk, M. E. Structure generation by reduction: A new strategy for computer-assisted structure elucidation. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 87–93.
- (9) Kerber, A.; Laue, R.; Moser, D. Ein Strukturgenerator für molekulare Graphen. *Anal. Chim. Acta* **1990**, *235*, 221–228.
- (10) Funatsu, K.; Sasaki, S. Recent advances in the automated structure elucidation system CHEMICS. Utilization of two-dimensional NMR spectral information and development of peripheral functions for examination of candidates. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 190–204.
- (11) Chiavarino, B.; Crestoni, M. E.; Di Marzio, A.; Fornarini, S.; Rosi, M. Gas-Phase Ion Chemistry of Borazine, An Inorganic Analogue of Benzene. *J. Am. Chem. Soc.* **1999**, *121*, 11204–11210.

CI000135O