DETERMINATION OF THE NUMBER
OF STRUCTURAL ISOMERS

By

G. E. VERESS

Department of General and Analytical Chemistry, Technical University Budapest

In order to clear up the structure of covalently bounded organic com-
pounds it is necessary to describe all the structural isomers in knowledge of
the empirical formula. To gather, however, all structural isomers belonging to
a given empirical formula is rather difficult, since no general method is known
for the determination of their number. The study of the structure is rather
problematic and even could lead to a deadlock if the exact number of isomers
is not known, since just the really existing structural isomer may be forgotten.
Thus, in structural investigations it is essential to know the number of structur-
al isomers belonging to a given empirical formula. Concerning the structural
isomers the handbooks contain data relating only to the paraffines [1] and
after some experiments [2] HENZE and BLAIR deal with the number of structural
isomers only of the paraffines [3], saturated monoalcohols [4], the members
of the olefine series [5], some significant aliphatic compounds [6]. Their
method is a combinatorial one; they compose the molecules from particles
divided into different groups. This method, however, very original, cannot
be applied for the determination of the number of structural isomers of any
molecule. When determining the structural isomers of organic compounds it
is advisable first of all to form groups of the possible structural isomers.These
groups are to be formed according to the variations of the order of carbon
atoms derived from the number of carbon and hydrogen atoms. So on the basis
of results obtained by physico-chemical and chemical experiments certain
cases can automatically be excluded and then the possible structural isomers
can be described.

In this paper a method suitable to determine the possible variations of
the order of carbon atoms in organic compounds is presented giving a possi-
bility to determine the number of the structural isomers in covalently bound-
ed compounds of any composition.

The variations of orders of carbon atoms

Let the molecule of composition C_nH_m be taken as example, where conventionally n and m are number of the carbon and hydrogen atoms in the molecule, respectively. From among n carbon atoms let x_i be that of i-th order $(i = 1, \ldots, 4)$. According to the rules of stochiometry validity of the following relationship is evident:

$$\left.\begin{aligned} n &= x_1 + x_2 + x_3 + x_4 \\ m &= 3x_1 + 2x_2 + x_3 \end{aligned}\right\} \tag{1}$$

In mathematical sense these equations can be regarded as simultaneous diophantic equations. Their positive integer solutions offer the possible variations of the order of carbon atoms, while the number of their positive integer solutions give the number of the possible variations.

Determination of variations

How can the positive integer solutions of the equation system (1) be determined?

If the general solution is required to be got (e.g. [7]) three cases can be distinguished depending on the residue of the division: m divided by three.

Let us introduce the congruence

$$a \equiv b \, (\text{mod } c)$$

i.e. number a divided by number c, the residue of the division is equal to b.

Further, be the integer part of the division a/b denoted by $\left[\dfrac{a}{b}\right]$ (read entier a/b).

Then the three cases are as follows:

1. $m \equiv 0 \, (\text{mod } 3)$

Introducing v and z new variables:

$$\left.\begin{aligned} x_1 &= \frac{m}{3} - \frac{2x_2 + x_3}{3} = \frac{m}{3} - v \\ x_2 &= v + \frac{v - x_3}{2} = v + z \\ x_3 &= v - 2z \\ x_4 &= n - \frac{m}{3} - v + z \end{aligned}\right\} \tag{2a}$$

2. $m \equiv 1 \pmod 3$

$$x_1 = \left[\frac{m}{3}\right] - v$$

$$x_2 = v + z$$

$$x_3 = v - 2z + 1$$

$$x_4 = n - \left[\frac{m}{3}\right] - v + z - 1$$

$$(2b)$$

3. $m \equiv 2 \pmod 3$

$$x_1 = \left[\frac{m}{3}\right] - v$$

$$x_2 = v + z$$

$$x_3 = v - 2z + 2$$

$$x_4 = n - \left[\frac{m}{3}\right] - v + z - 2$$

$$(2c)$$

Substituting any chosen value for variables v and z, the integer solutions of Eqn. (1) are obtained. However, from among the solutions the positive integer ones are only of interest which are to obtain by the way of analytic geometry.

Since for every solution x_i

$$x_i \geqslant 0, \qquad (3)$$

therefore performing this substitution in the above mentioned equations, the expressions x_i determine a half plane each in the system of coordinates v, z. The pair of value (v, z) corresponding to the positive integer solutions of the equation system (1) are given by the lattice points in the common parts of the four half planes.

Let us see it for example in the case 1. Using Eqn. (3) the following relationships are obtained from Eqn. (2a):

$$v \leqslant \frac{m}{3} \qquad (x_1)$$

$$z \geqslant -v \qquad (x_2)$$

$$z \leqslant \frac{1}{2} v \qquad (x_3)$$

$$z \geqslant v - n + \frac{m}{3} \qquad (x_4)$$

Let these inequalities be plotted in the system of coordinates v, z attributing given values to letters n and m. In Fig. 1, the striped part corresponds to the field determined by the four inequalities.

The positive integer solutions are obtained by substituting the lattice points (v, z) of the pencilled field into the equation system (2a).

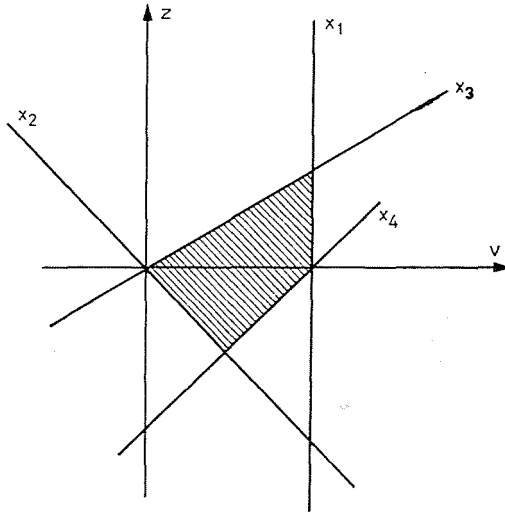Similar system is to be applied in the second and third cases.

Fig. 1. The field of the possible solutions

In many cases, however, it is simpler to determine the positive integer solutions of the equation system (1) by guesswork.

In order to obtain all the solutions it is necessary to know the number of the solutions.

Number of variations

How many solutions belong to the equation system (1)?

Generally it is true [8] that there are as many positive integer solutions as the coefficient of the member $x^m y^n$ of the power series in the function:

$$f(x, y) = \frac{1}{(1 - x^3 y)(1 - x^2 y)(1 - xy)(1 - y)}$$

This can theoretically be calculated by partial differentiation:

$$\operatorname{coeff} x^m y^n \text{ in } f(x, y) = \left[\frac{\partial^n}{\partial y^n} \left(\frac{\partial^m f(x, y)}{\partial x^m} \right) \cdot \frac{1}{nm} \right]_{(0,0)}$$

The calculation of this coefficient means a great problem that is why a simplier way is advisable to choose.

Without deduction SYLVESTER [9] reports several formulae relating to the positive integer solutions of the following simultaneous equations depending on the ratio of m/n:

$$ax + by + cz + \ldots\ldots + lw = m$$

$$x + y + z + \ldots\ldots + w = n$$

These results, however, cannot be generalized.

Geometrically it is clear that the number of the positive integer solutions of the equation system (1) in the range $n - 2 \leqslant m \leqslant 2n + 2$ is obtained by substracting the number of the positive integer solutions of Eqn. (4) from that of Eqn. (5).

$$3x_1 + 2x_2 + x_3 = m \tag{4}$$

$$2x_1 + x_2 + x_4 = m - n - 1 \tag{5}$$

On this basis the wanted numbers of the positive integer solutions of equation systems (4) and (5) is simply got by known methods (e.g. [7], [8]).

The number ($m \geqslant 0$) of the positive integer solutions of the Eqn. (4) is

$$g_1(m) = \operatorname{coeff} t^m \text{ in } G_1(t) \tag{6}$$

where

$$G_1(t) = \frac{1}{(1 - t^3)(1 - t^2)(1 - t)} =$$

$$= \frac{1}{(1 - t)(1 - \varepsilon t)(1 - \varepsilon^2 t)(1 + t)(1 - t)(1 - t)} =$$

$$= \frac{1}{(1 - t)^3(1 + t)(1 - \varepsilon t)(1 - \varepsilon^2 t)}$$

ε: is the third primitive complex unity root

$$\varepsilon^\nu = \cos\frac{2\pi\nu}{3} + i\sin\frac{2\pi\nu}{3}$$

$$\varepsilon^\nu = 1 \qquad\qquad (\nu \equiv 0 \mod 3)$$

$$\varepsilon^\nu = -\frac{1}{2} + i\frac{\sqrt{3}}{2} \quad (\nu \equiv 1 \mod 3)$$

$$\varepsilon^\nu = -\frac{1}{2} - i\frac{\sqrt{3}}{2} \quad (\nu \equiv 2 \mod 3)$$

Decomposed into partial fractions:

$$G_1(t) = \frac{A}{(1-t)^3} + \frac{B}{(1-t)^2} + \frac{C}{1-t} + \frac{D}{1+t} + \frac{E}{1-\varepsilon t} + \frac{F}{1-\varepsilon^2 t}$$

The coefficients are as follows:

$$A = \lim_{t \to 1} G_1(t) \cdot (1-t)^3 = \frac{1}{6}$$

$$D = \lim_{t \to -1} G_1(t) \cdot (1+t) = \frac{1}{8}$$

$$E = \lim_{t \to \varepsilon^2} G_1(t) \cdot (1-t) = \frac{1}{9}$$

$$F = \lim_{t \to \varepsilon} G_1(t) \cdot (1-t) = \frac{1}{9}$$

substituting $t = 0$ and $t = 2$

$$B = \frac{1}{4} \qquad C = \frac{17}{72}$$

Thus

$$G_1(t) = \frac{1}{6} \cdot \frac{1}{(1-t)^3} + \frac{1}{4} \cdot \frac{1}{(1-t)^2} + \frac{17}{72} \cdot \frac{1}{1-t} + \frac{1}{8} \cdot \frac{1}{1+t} + \frac{1}{9} \cdot$$

$$\cdot \frac{1}{1-\varepsilon t} + \frac{1}{9} \cdot \frac{1}{1-\varepsilon^2 t}$$

Expanding in series the members of $G_1(t)$:

$$\frac{1}{6} \cdot \frac{1}{(1-t)^3} = \frac{1}{6}\left(1 + 3t + 6t^2 + \ldots + \frac{(m+1)(m+2)}{2} \cdot t^m + \ldots\right)$$

$$\frac{1}{4} \cdot \frac{1}{(1-t)^2} = \frac{1}{4} \cdot (1 + 2t + 3t^2 + \ldots + (m+1) \cdot t^m + \ldots)$$

$$\frac{17}{72} \cdot \frac{1}{1-t} = \frac{17}{72}(1 + t + t^2 + \ldots + t^m + \ldots)$$

$$\frac{1}{8} \cdot \frac{1}{1+t} = \frac{1}{8}(1 - t + t^2 - \ldots + (-1)^m t^m + \ldots)$$

$$\frac{1}{9} \cdot \frac{1}{1-\varepsilon t} = \frac{1}{9}(1 + \varepsilon t + \varepsilon^2 t^2 + \ldots + \varepsilon^m t^m + \ldots)$$

$$\frac{1}{9} \cdot \frac{1}{1-\varepsilon^2 t} = \frac{1}{9}(1 + \varepsilon^2 t + \varepsilon^4 t^2 + \ldots + \varepsilon^{2m} t^m + \ldots)$$

From Eqn. (6) the wanted number of the solutions:

$$g_1(m) = \frac{(m+1)(m+2)}{12} + \frac{m+1}{4} + \frac{17}{72} + \frac{1}{8}(-1)^m + \frac{1}{9}(\varepsilon^m + \varepsilon^{2m})$$

or

$$g_1(m) = \frac{1}{72}[(m+1)(m+5)6 + 17 + 9(-1)^m + 8(\varepsilon^m + \varepsilon^{2m})] \qquad (7)$$

Similarly to the above mentioned the number $(u \geqslant 0)$ of the positive integer solutions of the Equ. (5) is the following:

$$g_2(u) = \text{coeff } t^u \text{ in } G_2(t) \qquad (8)$$

where

$$u = m - n - 1$$

and

$$G_2(t) = \frac{1}{(1-t^2)(1-t)(1-t)} = \frac{1}{(1+t)(1-t)^3} = \frac{A}{1+t} + \frac{B}{(1-t)^3} +$$

$$+ \frac{C}{(1-t)^2} + \frac{D}{1-t}$$

The coefficients are:

$$A = \lim_{t \to -1} G_2(t) \cdot (1+t) = \frac{1}{8}$$

$$B = \lim_{t \to 1} G_2(t) \cdot (1-t)^3 = \frac{1}{2}$$

and by substituting $t = 0$ and $t = 2$

$$C = \frac{1}{4} \cdot \qquad D = \frac{1}{8}$$

Hence

$$G_2(t) = \frac{1}{8(1+t)} + \frac{1}{2(1-t)^3} + \frac{1}{4(1-t)^2} + \frac{1}{8(1-t)}$$

Expanding in series the members of $G_2(t)$:

$$\frac{1}{8} \cdot \frac{1}{1+t} = \frac{1}{8} \cdot (1 - t + t^2 - \ldots + (-1)^u \cdot t^u + \ldots)$$

$$\frac{1}{2} \cdot \frac{1}{(1-t)^3} = \frac{1}{2}(1 + 3t + 6t^2 + \ldots + \frac{(u+1)(u+2)}{2} \cdot t^u + \ldots)$$

$$\frac{1}{4} \cdot \frac{1}{(1-t)^2} = \frac{1}{4}(1 + 2t + 3t^2 + \ldots + (u+1)t^u + \ldots)$$

$$\frac{1}{8} \cdot \frac{1}{1-t} = \frac{1}{8}(1 + t + t^2 + \ldots + t^u + \ldots)$$

On the basis of equation (8) the wanted number of the solutions:

$$g_2(u) = \frac{1}{8}(-1)^u + \frac{(u+1)(u+2)}{4} + \frac{u+1}{4} + \frac{1}{8}$$

or

$$g_2(u) = \frac{(u+1)(u+3)}{4} + \frac{1}{8}[1+(-1)^u] \tag{9}$$

Consequently the number of the positive integer solutions of equation system (1) is

$$g(n,m) = g_1(m) - g_2(m-n-1) \tag{10}$$

As the value of m should always be an even number, so

$$(-1)^m = 1$$

and

$$\varepsilon^m + \varepsilon^{2m} = \begin{cases} +2, \text{ if } m \equiv 0 \pmod 3 \\ -1, \text{ if } m \not\equiv 0 \pmod 3 \end{cases}$$

therefore instead of Eqn. (7) it can be written

$$g_1(m) = \frac{(m+1)(m+2)}{12} + \frac{m+1}{4} + \frac{17}{72} + \frac{1}{8} + \frac{1}{9}\left(\begin{cases} +2, \text{ if } m \equiv 0 \pmod 3 \\ -1, \text{ if } m \not\equiv 0 \pmod 3 \end{cases}\right) \tag{11}$$

Substituting the value of u into Eqn. (9)

$$g_2(m-n-1) = \frac{(m-n)(m-n+2)}{4} + \frac{1}{8}(-1)^{m-n-1}$$

since

$$(-1)^{m-n-1} = -(-1)^n$$

thus

$$g_2(m-n-1) = \frac{(m-n)(m-n+2)}{4} - \frac{1}{8}(-1)^n + \frac{1}{8} \tag{12}$$

Substituting Eqn. (11) and (12) into (10), the number of the positive integer solutions of equation system (1), i.e. the number of the variations of carbon order:

$$g(n,m) = \frac{(m+1)(m+5)}{12} - \frac{(m-n)(m-n+2)}{4} + \frac{17}{72}$$

$$+ \frac{1}{8}(-1)^n + \frac{1}{9}\left(\begin{cases} +2, \text{ if } m \equiv 0 \pmod 3 \\ -1, \text{ if } m \not\equiv 0 \pmod 3 \end{cases}\right) \tag{13}$$

General description of molecules

In knowledge of the positive integer solutions of equation system (1) the compound of the general formula C_nH_m can be rewritten as a compound of molecule number $g(n, m)$

$$X_{x_1}^I \ X_{x_2}^{II} \ X_{x_3}^{III} \ X_{x_4}^{IV}$$

by general formula, where

X^I is the part of molecule $CH_3 -$
X^{II} is the part of molecule $CH_2 =$
X^{III} is the part of molecule $CH \equiv$
X^{IV} is the part of molecule $C \equiv$

In most cases — on the basis of chemical and physicochemical experiments — certain variations can be omitted at the beginning of research work.

So far compounds containing only carbon and hydrogen have been dealt with. A molecule containing heteroatom — or more generally the given part of molecule — should be substituted by an "atom" of given valency.

On the basis of the above mentioned a molecule consists of "atoms" x^{ij}, where i is the valency number, the second superscript j is the sort of atoms having valency i. μ_{ij} means the number of atoms x^{ij} in the molecule. Thus, each molecule can be described in the following form:

$$\pi_i \ \pi_{j(i)} \ x^{ij}_{\mu_{ij(i)}} \tag{15}$$

Determination in structural isomers

On the basis of the above mentioned method the possible variations of the carbon order in a covalently bounded compound of any composition can be determined. Consequently the number of the structural isomers and the structural isomers themselves, are easier to determine.

Two ways are offered to determine the structural formula of the molecules described by empirical formula (14), where the number of molecules $g(n, m)$ is obtained by the outlined method. One way of the determination is guessing, the other is to describe all the reachability matrices characterizing the relation between the possible bounds. The latter is intended to be reported on in a subsequent paper.

Example for illustration

Be the tested molecule a composition C_5H_{12}, then the given form of the diophantic equation system (1) is as follows:

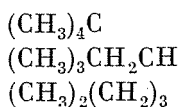$$x_1 + x_2 + x_3 + x_4 = 5$$
$$3x_1 + 2x_2 + x_3 = 12$$

The number of the solutions of the equation system on the basis of relationship (13) is as follows:
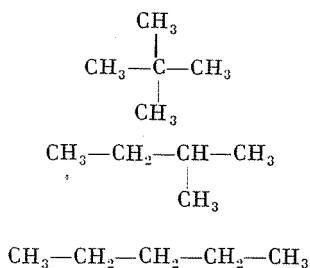
$$g(5, 12) = 3$$

The solutions of the equation system are:

x_1	x_2	x_3	x_4
4	0	0	1
3	1	1	0
2	3	0	0

According to the above, the molecule wanted can be described as the variation of the following, and only of the following three carbon orders:

$(CH_3)_4C$
$(CH_3)_3CH_2CH$
$(CH_3)_2(CH_2)_3$

On the basis of the variations of carbon orders the structural isomers are:

$$CH_3-\overset{\overset{\displaystyle CH_3}{|}}{\underset{\underset{\displaystyle CH_3}{|}}{C}}-CH_3$$

$$CH_3-CH_2-\underset{\underset{\displaystyle CH_3}{|}}{CH}-CH_3$$

$$CH_3-CH_2-CH_2-CH_2-CH_3$$

(In this example to each variation of carbon order a single special structural isomer belongs).

Summary

A method suitable to determine the number of the possible variations of the order of carbon atoms in organic compounds is presented giving a possibility to determine the number of the structural isomers in covalently bounded compounds of any composition.

References

1. BRUCKNER, G.: Szerves Kémia I. vol. 48. p.
2. CAYLEY, Ber. 8, 1056 (1875)
3. HENZE, H. R., BLAIR, C. M.: J. Am. Chem. Soc. 53, 3077 (1931).
4. HENZE, H. R., BLAIR, C. M.: J. Am. Chem. Soc. 53, 3042 (1931)
5. HENZE, H. R., BLAIR, C. M.: J. Am. Chem. Soc. 55, 680 (1933)
6. HENZE, H. R., BLAIR, C. M.: J. Am. Chem. Soc. 56, 157 (1934)
7. TURÁN, P.,—LÁNCZI, I.: Számelmélet (Kézirat) 1962. Bp.
8. DICKSON, L. E.: History of the Theory of Numbers Vol II. Chelsea Publ. Co. N. Y. 1952
9. SYLVESTER, J. J.: Coll. Math. Papers, II. 113—117 (Cambridge) (1908)

Dr. Gábor E. VERESS, H-1521 Budapest