Isonumeric Compounds: Derivative Enumeration

Igor Novak

Department of Chemistry, National University of Singapore, Singapore 0511, Singapore

The basic expressions governing derivative enumeration are given and a sufficient condition proposed which applies to isonumeric compounds.

Introduction

The links between combinatorics, graph theory and group theory (symmetry) have been studied in depth with the aim of systematic and exhaustive enumeration of isomers [1, 2]. The well known Pólya enumeration method can be used for that purpose.

There is, however, an interesting related question which has not been posed before: *"Assuming that one has a parent compound (molecule) with **k** monovalent substitution sites and **n** different possible monovalent ligands, ligand repetition being allowed, what is the number of possible derivatives **d**?"*.

A corollary to that question is: *"When will two different parent compounds have the same number of derivatives, i.e. be isonumeric?"*

In this article we try to derive general conditions which will obviate the need for complex and often tedious counting procedures and give analytical expressions for the count. This work thus extends the chemical enumeration to its most comprehensive form by including not only isomers but all derivatives.

Discussion

We shall assume that d includes only a single member of each conformer population, if ligands and parent compound are non-rigid. Otherwise ligands will be achiral and rigid (without conformational flexibility) and freely rotating around substituent bonds. We shall also assume (without loss of generality) that bonds linking the same atoms are identical. While this is usually true, there are some counter-examples (e.g.

axial and equatorial Te–Cl bonds in $TeCl_4$). The requirements of monovalency and ligand rigidity are introduced in order to simplify the analysis and produce possible generalisations. Furthermore, we shall assume that the sites are permanently substituted (with hydrogen atom also being counted as substituent).

It can be expected that n, k and molecular symmetry $\mathfrak{G}$ (expressed through its point group) will be independent variables in a functional relationship determining d values:

$$d = f(n, k, \mathfrak{G}), \quad d, n, k \geq 1 . \tag{1}$$

The functional relationship f can in fact be written explicitly:

$$d = c_1 \binom{n}{1} + c_2 \binom{n}{2} + c_3 \binom{n}{3} + \dots$$

$$= \sum_{s=1}^{k} c_s(\mathfrak{G}) \binom{n}{s}, \quad d > c_s \geq 1 . \tag{2}$$

The series in (2) can be deduced by assuming that the sites contain a single type of substituent, two different types, three different types, etc. Index s gives the number of different ligands (chosen from the pool n) present in a molecule at any one time. Since $s \leq k$ in (2), ligand repetition is automatically taken into account.

The binomial coefficients represent combinations of n objects taken s at a time, without repetition. Integer c_s coefficients account for various possibilities arising from molecular symmetry $\mathfrak{G}$ (see examples) and can be derived from coefficients of a suitable counting polynomial [2]. We shall designate $[c_1, c_2, \dots]$ a "structure vector". If $n \leq s$ the corresponding binomial coefficient in (2) vanishes. Equation (2) shows that one can separate combinatorial $\binom{n}{s}$, from molecular structure dependent factors $[c_s]$.

Equation (2) can be simplified since $c_1 \equiv 1$ and $\binom{n}{1} = n$,

$$d = n + \sum_{s=2}^{k} c_s(\mathfrak{G}) \binom{n}{s}. \tag{3}$$

The reason why $c_1 \equiv 1$ is that if only a single type of substituent is present in the molecule at a time, then irrespective of symmetry there will be as many derivatives as there are available ligands: $d = n$ if $k = 1$.

The derivation of d in general is tantamount to the derivation of the structure vector $[c_s]$, a process complicated by different possible molecular symmetries. Direct use of Pólya's theorem for evaluation of $[c_s]$ of large molecules is cumbersome, although analytical expressions were given for planar molecules [3]. We shall than try the easier task of defining an interval within which d must be found.

In order to find the boundaries of the interval we must consider two extreme cases:

a) all substituent sites (bonds) are equivalent by symmetry, i.e. transformed by symmetry operations of the point group (excluding the trivial case of C_1 symmetry);

b) they are all different (C_1 point group).

$$\binom{n+k-1}{k} < d < n^k. \tag{4}$$

Equation (4) is based on elementary formulae for the number of combinations of n elements taken k at a time with repetition (case a)), and for the number of permutations of k elements selected from n elements with repetition (case b)).

When all the sites are nonequivalent, the molecule may be chiral, so the upper boundary in (4) becomes $2n^k$, which will then account for the existence of enantiomeric pairs.

For $k = 2$, $c_2 = 1$ or 2, i.e. c_2 can only assume values of 1 and 2 depending on the symmetry (non)equivalence of the two substituent bonds. For $k \geq 2$ the symmetry begins to influence d and a simple, general derivation is no longer possible. The total number of possible derivatives can also be expressed in terms of the point group (molecular symmetry):

$$d = n + \sum_{s=2}^{p} \chi_s(\mathfrak{G}_s, k, n), \quad p, \chi_s \geq 1, \tag{5}$$

where $\mathfrak{G}_1$ is the symmetry (point group) of the derivative with a single type of substituent, while $\mathfrak{G}_2, \mathfrak{G}_3, \ldots, \mathfrak{G}_p$ are its subgroups. $[\chi_s]$ coefficients form p components of another vector, which we shall call "group vector". The first term "n" in (5) arises from the parent group, when all the substituents are identical. The $[\chi_s]$ vector is dependent on $[c_s]$ since its components are linear combinations of $[c_s]$.

The calculation of the $[c_s]$ and $[\chi_s]$ vectors' components may be done manually by drawing and counting possible structures, but this is impractical for large molecules with many sites. In that case more complicated and laborious methods based on permutation group theory must be used.

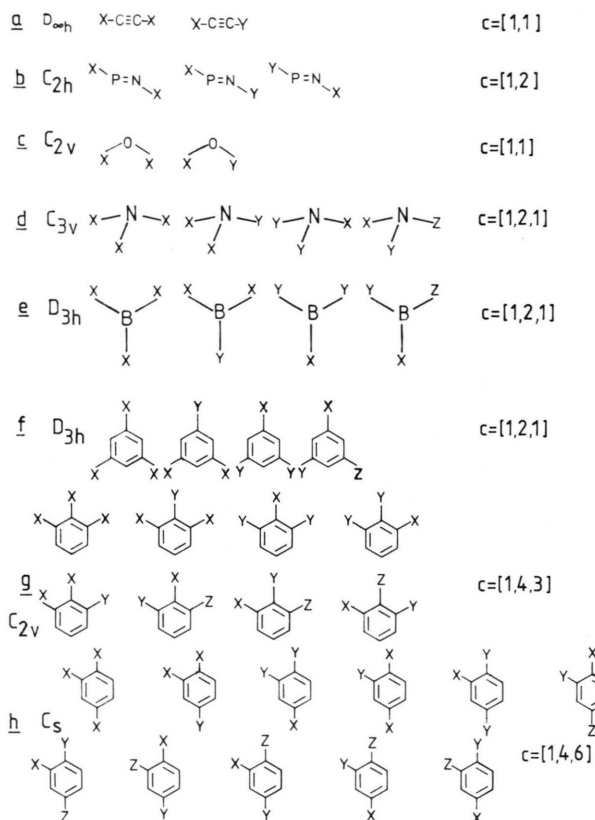We shall now turn our attention to the *isonumericity*.

Equations (2) and (5) are linear Diophantine equations in $k-1$ and $p-1$ unknowns, respectively. They have many possible solutions, and subsequently (for given n, k) many possible isonumeric compounds exist. While a simple, necessary and sufficient condition for *isonumericity* has not been derived, we shall give a weaker, sufficient condition if the substituent bonds in the two isosymmetric parent molecules span the same set of irreducible representations, than the compounds are isonumeric. This requirement ensures that upon substitution the bonds (basis vectors) transform identically under symmetry operations and give identical subgroups ($\mathfrak{G}_p$).

Examples (Fig. 1)

The rather abstract discussion so far can be made more accessible by examining several examples of real molecules. The examples shall be confined to $k = 2, 3, 4$ because for larger k values the number of generic structures which needs to be drawn and counted increases fast.

Examples **a–c** are for molecules with $k = 2$ and different symmetries. The structure vectors c_s are given for each molecule. It is obvious that molecules **a** and **c** are isonumeric but not isosymmetric. The set of examples for $k = 3$ demonstrates once again that the two isonumeric parent molecules may be isosymmetric, (**e**, **f**), but not necessarily so, (**d**, **e**). In both **e** and **f** the substitutent bonds span $a_1' + e'$ representations of D_{3h} point group which is a sufficient condition for isonumericity.

Symmetry of the parent molecule is determined by its available substitution sites, not all of which have to be occupied or available. The example of trisubstituted benzenes is a case in point. The more complicated

<u>a</u> $D_{\infty h}$ X–C≡C–X   X–C≡C–Y   c=[1,1]

<u>b</u> C_{2h} $\underset{X}{\overset{X}{}}P{=}N\underset{X}{}$ $\underset{Y}{\overset{X}{}}P{=}N\underset{X}{}$ $\underset{X}{\overset{Y}{}}P{=}N\underset{X}{}$ c=[1,2]

<u>c</u> C_{2v} $\underset{X}{}\overset{O}{}\underset{X}{}$ $\underset{X}{}\overset{O}{}\underset{Y}{}$ c=[1,1]

<u>d</u> C_{3v} x—N—x   x—N—y   y—N—x   x—N—z   c=[1,2,1]

<u>e</u> D_{3h} (B structures)   c=[1,2,1]

<u>f</u> D_{3h} (benzene-ring structures)   c=[1,2,1]

<u>g</u>
C_{2v} c=[1,4,3]

<u>h</u> C_s c=[1,4,6]

examples of thiophene, oxirane, pyrazine and ethene, with $k = 4$ are not shown in Fig. 1, but instead we give the structure vectors for each molecule:

thiophene: $c_s = [1, 8, 18, 6]$, C_{2v},
oxirane: $c_s = [1, 5, 9, 6]$, C_{2v},
ethene: $c_s = [1, 5, 9, 6]$, D_{2h},
pyrazine: $c_s = [1, 5, 9, 6]$, D_{2h}.

The isonumericity of ethene and pyrazine can be deduced without reference to the structure vectors, because 4 substituent bonds span $a_g + b_{3g} + b_{1u} + b_{2u}$ representations of the D_{2h} point group. This example also demonstrates that two molecules may be isosymmetric and yet not isonumeric (thiophene and oxirane).

Conclusion

We have derived concise, general expressions for the number of derivatives in an arbitrarily selected molecule and an isonumericity condition.

Several interesting and difficult questions remain:

– is it possible to predict, on the basis of molecular symmetry and irreducible representations of substituent bonds alone, which of the two molecules has a larger number of derivatives?

– is there a definite relationship between d's of the molecules whose symmetries are in group-subgroup relationship?

[1] S. Fujita, Symmetry and Combinatorial Enumeration in Chemistry, Springer-Verlag, Berlin 1991.
[2] N. Trinajstić, Chemical Graph Theory, 2nd ed., CRC Press: Boca Raton 1992, p. 275.
[3] A. Yu. Kruglyak and M. E. Dochtmanov, J. Mol. Struct., Theochem. **233**, 13 (1991).