communications in mathematical
match
and in computer chemistry

no. 39, March 1999

ISSN 0340-6253

MATCDY (39) 21-37 (1999)

Sample Applications of an Algorithm for the Calculation of the Number of Isomers With More Than One Type of Achiral Substituent†

Hans Dolhaine[a], Helmut Hönig*[b], Markus van Almsick[c]

[a]Henkel KGaA, Henkelstraße 67, D-40191 Düsseldorf, Germany
[b]Institute for Organic Chemistry, Technical University Graz, Stremayrgasse 16, A-8010 Graz, Austria. email: hoenig@orgc.tu-graz.ac.at
[c]Am Markuskreuz 6, D-45133 Essen, Germany

Abstract

An isomer enumeration program written in a symbolic manipulation language and incorporating an efficient algorithm[1] based on Pólya's theorem is applied to a number of molecules of interest for organic chemists: acenaphthylene, anthracene, azulene, benzene, chrysene, corannulene, coronene, cubane, cyclopropane to cyclohexane, dibenzo[2.2]paracyclophane, C_{20}-[5]-fullerane (dodecahedrane), C_{24}-[4,6]-fullerane, C_{36}-[5,6]-fullerene, C_{60}-[5,6]-fullerene, *as*-indacene, *s*-indacene, naphthalene, pentalene, perylene, phenanthrene, prismane, [5]-prismane, [6]-prismane, pyrene, tetrahedrane, triphenylene, triptycene and twistane. The numbers of enantiomeric pairs and achiral derivatives are listed whenever induced chirality by substitution is possible. A discussion of inositols and their respective oligomers is included.

Introduction

A rather simple sounding question was the starting point for this work. Given an unlimited supply of 25 different substituent types, how many isomers of benzene are there in total, if a combination of *n* substituents is allowed at the 6 benzene substituent sites? This question arose in the course of refining a computer program for the improved prediction of carbon-13 chemical shifts of (i. a.) aromatic systems[2,3]. The improvement mentioned is based on the introduction of

† Dedicated to Prof. Dr. Herfried Griengl on the occasion of his 60th birthday.

additional pairwise correction increments, which in the case of e.g. 25 different substituents, do amount to as many as 975 possible pairs, as will be shown below. We wanted to determine the relationship between necessary input parameters (number of disubstituted benzenes) and all possible isomers (up to hexasubstitution at the ring), the ^{13}C-nmr spectra of which could then be predicted in an improved way. This relationship could serve as an indicator for the predictive power of the method, which other research groups[4] consider for adoption.

Furthermore, there are several areas of interest related to this question. How many isomers should one consider, if infrared spectroscopy of an unknown compound exhibits three substituents like nitro, methoxy and cyano and proton nmr suggests 1,2,4-substitution? How many isomers exist in the case of 1,3,5-substitution? How many members are there in a library of all possible halogenated (F through I) pyrenes? If one takes, for example, 7 different substituents on benzene as descriptors, the physical properties of how many different compounds could be predicted by Katritzky's method[5]? Similar questions were recently posed to one of us (H.H.)[6]. How many dimeric inositols are there theoretically? Could one estimate the number of theoretically possible hexa-O-inositoyl-inositols? How many of those would be achiral?

These questions may sound rather academic, but the applications are quite useful for a number of applications. For present and future automatic structure elucidation packages, the total number of isomers, the amount of diastereoisomers and/or achiral derivatives are essential in order to check whether all possible combinations have been considered. Of course, duplications due to symmetry are to be avoided. There are also many applications for combinatorics in spectroscopy[7]. While the first examples mentioned earlier could be done by hand, the others would require a patient researcher devoting a lot of time, yet the results would still be susceptible to errors[8].

There are early approaches to the theoretical background of this problem[9] - and very elegant solutions to parts of it - all linked to Pólya's theorem[10,11]. However, most of the relevant references enumerate substitutional isomers with only one class of substituent[12,13] (e.g. chloro) to answer the equally interesting question of how many PCBs are possible[14]? The general principles with different structures are outlined in two illustrative papers[15,16], but again no actual figures for problems with many different substituents are given. One paper[17] gives the actual isomer enumerations for pentane, biphenyl, adamantane, and some for dodecahedrane. Another paper[18] presents cyclohexane with up to four different substituents. There is a reference on the cycle indices of polysubstituted peri-condensed aromatic compounds[19], yet only one functional group is considered and no actual figures are given. Another interesting and comprehensive paper on isomer enumeration in the field of fullerenes appeared recently[20], but again, the maximum

number of different types of substituents were limited to three (in case of the smaller fullerenes). For the bigger fullerenes, only one type was considered because of computational constraints as in other papers[21].

In this paper we present isomer enumerations that take large numbers of different substituent types into account. As indicated by the initial question, we do not consider a simple substitutional isomer enumeration for a given skeleton and a given set of substituents. Instead, we assume an unlimited supply of *n* different substituent types. We thus count all substitutional isomers that can be formed with a given skeleton and any combination of substituents possible. Thanks to Prof. Kerber, our attention was drawn to a recent reference[22] which depicts an elegant and simple general formula used in evaluating the sizes of theoretical combinatorial libraries. The general formulas in[22] for 1,3,5,7-tetrachlorocarbonyl cubane (I) [identical symmetry and substituent site distribution as tetrahedrane (**30**)], for a tetrachlorocarbonyl xanthene (II) (identical to furane or our 1,2,3,4-benzene) and for the benzene triacid chloride III (a 1,3,5-benzene) do exactly match our results given below and in the *EXCEL*-chart[25] on the web.

The calculations described here can be applied to many organic molecules, some of which are shown in figure 1: Planar aromatics like acenaphthylene (**1**), anthracene (**2**), azulene (**3**), benzene (**4**), chrysene (**5**), corannulene (**6**), coronene (**7**), pentalene (**8**), *as*-indacene (**9**), *s*-indacene (**10**), naphthalene (**11**), perylene (**12**), phenanthrene (**13**), pyrene (**14**) and triphenylene (**15**) mostly exhibit point groups D_{nh} or C_{nv} and will not show any enantiomers after substitution with achiral substituents. Some more sophisticated bodies are depicted in figure 2: The average conformations of the cycloalkanes cyclopropane to cyclohexane (**16**, **17**, **18**, **19**) feature a D_{nh} symmetry as prismane (**20**), [5]-prismane (**21**) and [6]-prismane (**22**). Dibenzo[2.2]para-cyclophane (**23**) and triptycene (**24**) are included as examples of aromatics with the possibility of enantioisomerism after substitution. Cubane (**25**), C_{20}-[5]-fullerane (dodecahedrane, **26**), C_{24}-[4,6]-fullerane (**27**), C_{36}-[5,6]-fullerene (**28**), C_{60}-[5,6]-fullerene (**29**) and tetrahedrane (**30**) are highly symmetric molecules with I_h, O_h or T_d symmetry, also showing many possibilities of enantioisomerism after substitution. For the discussion of the inositols, the possible stereoisomeric forms are also depicted (figure 3).

Method

The coefficients yielding the numbers of substituted isomers of the molecules shown were obtained by using the generating polynomial constructed according to Pólya's theorem. This has been done for benzene and is published in the early literature[23]. One can extract the results for

some other structures from the literature already cited. However, for the other molecules, as well as for any other structures, we have developed an easy to use computer application written in a modern computer algebra language[24]. This application not only allows an easy input of any structural symmetry, but also checks on the data given and determines the corresponding point group.

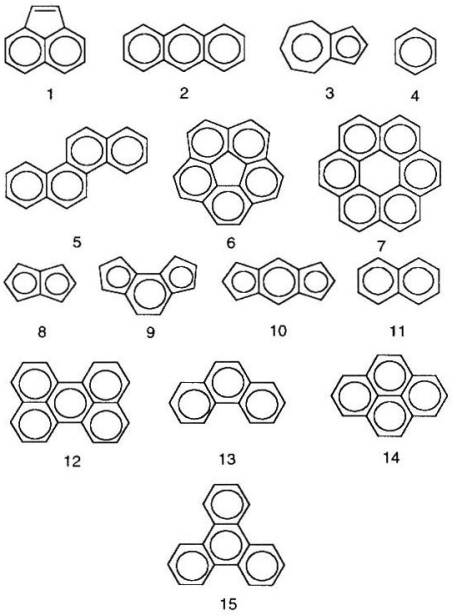

Figure 1: Planar aromatics under consideration.

To handle bigger molecules with more than 8 substitution sites, we were forced to introduce a special, more efficient algorithm calculating polynomial coefficients. An in-depth description of this program will be given elsewhere[1].

The program provides a command to define a molecule via its point group generator(s). With the possible reflection symmetry contained in these point group generators, it is also possible to

calculate the numbers of enantiomeric pairs and achiral isomers for substituted molecules like **16** to **30**.

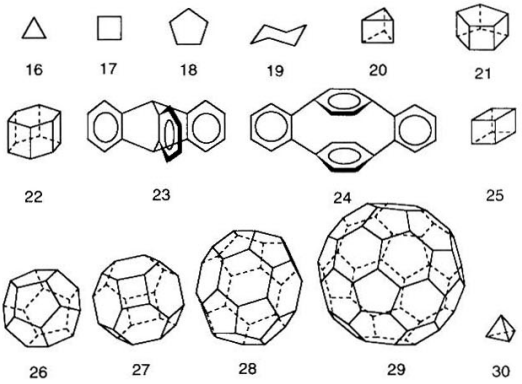

Figure 2: Spherical bodies treated.

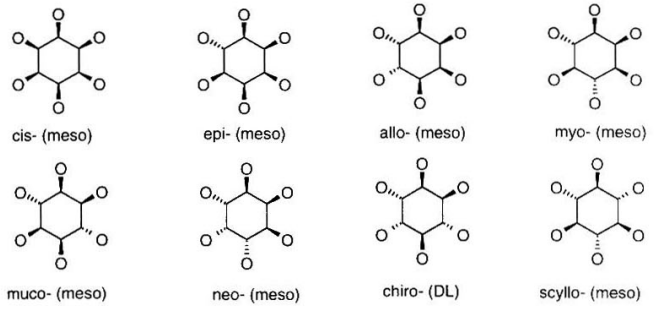

Figure 3: All possible inositols.

Here we will show one example to depict the effortless handling of the program. For tetrahedrane (**30**) we adopt an arbitrarily chosen numbering of substituent sites as shown in figure 4:

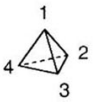

Figure 4: Chosen numbering of carbons in tetrahedrane (**30**)

There are three twofold axes and four threefold axes. Only one rotation around one of the twofold and one rotation around one of the threefold axes are needed to define all rotations of the point group. The rotations are encoded as permutations of substitution site labels. Prior to rotations we assume a 1-2-3-4 label sequence as depicted in figure 4. A rotation around the twofold axis through the middle of 1-2 and 3-4 reads: 2-1-4-3. Similarly a 120°-rotation around the threefold axis through 1 and the middle of the plane 2-3-4 renders permutation 1-3-4-2. For the enumeration of enantiomers, one defines just one mirror symmetry, e.g. the reflection across the plane that includes 1 and 2 and cuts through the middle of 3-4, which is represented by permutation 1-2-4-3. The appropriate command implementing tetrahedrane in the program is:

DefineParentCompound[Tetrahedrane, { R[2, 1, 4, 3], R[1, 3, 4, 2], R_σ[1, 2, 4, 3]}].

One could include the remaining axes or planes of symmetry (rotations R; reflections R_σ) without altering the results. To verify the input provided by the user, the program automatically performs a list of consistency checks and presents the resulting point group(s). With some simple commands the program then prints the respective tables, either in general form or with actual numbers, some of which are shown below.

Results

Three facts have to be kept in mind when examining the following tables: First, we talk about „isomers" assembled with substituents taken from a supply of n different substituent types. Thus, the resulting ensemble of compounds is in fact not an isomeric ensemble but rather a set of diamutamers[1, 18]. We retain the expression „isomers" because it is used in most of the corresponding papers on this issue. Second, the figures given in the tables below include

„isomers" with multiple occurrence of any of the n different substituent types. Think of the tetrahedrane (T) example mentioned earlier in combination with 4 different substituents (a, b, c and d): Thus not only T(abcd), but also T(aaaa), T(aaab) etc. are valid combinations. Within the program there are options to calculate any partition of substituent types explicitly. Third, a note about molecules with initial chirality. For molecules like *chiro*-inositol (see figure 3) the program also considers the mirror image of the skeleton. Thus, even though only one parent enantiomer is the basis of the input, both, the input skeleton and its enantiomeric counterpart, are considered in the enumeration thereby rendering an isomer count that is twice as large than might be expected.

The following tables present the isomer enumeration results according to the number of occupied substituent sites X listed by row. Hence, the first row in each table contains the number of isomers with monosubstitution. Only the substituent type and substituent site are left unspecified so that the resulting isomeric combinations of these two criteria are counted. In the second row, the isomers with disubstitution are enumerated, and so on. For each skeleton the number of occupied substituent sites X has an upper boundary, since each molecule can only exhibit a finite number of substituent sites. The number of substituent types n, however, may be arbitrarily large. We therefore provide only a small selection of n-values for the actual numbers of benzene isomers in table 1, where the columns are used to list the number of substituent types n. All the following tables provide formulas in n, so that the columns can be used to list different skeletons. An *EXCEL*-chart containing all the examples of this paper - and quite some additional ones - can be downloaded from the web[25].

Users wanting to compile their own *EXCEL*-tables should consider the following consistency check: The sum over all possible arrangements („Sum" in table 1) for a given n should be equal to the value for full substitution ($X = 6$ in case of benzene) for the $n + 1$ - case minus 1. See also footnotes [§] and [$] in table 1. This is due to the fact, that in the n - series, formally the $(n + 1)^{st}$ substituent is always hydrogen. If one replaces now this hydrogen by a new substituent ($n + 1$ - series) all former arrangements from $X = 1$ to $X = n$ from the n - series are now encompassed in the full substituted X_{max} - case of the $n + 1$ - series. The additional term „minus 1" results from the fact, that formally the fully substituted molecule with this particular substituent hydrogen in the n - series is the parent (unsubstituted) molecule and is thus not included in the n - series. In the $n + 1$ - series, of course there is a fully substituted molecule with this additional substituent

type. One should perform this check with n equal at least 2, since the results for $n = 1$ are independent of the exponent of n and thus typing errors in the exponents still could be overlooked.

A cautionary note has to be added: For larger molecules with more than 20 occupied substituent sites X, our program may need a long time to calculate the symbolic isomer enumeration results in n. For a C_{40}-fullerene on a 233 MHz Pentium PC one should allow for about two hours of computing time for the complete table as given in the *EXCEL*-chart[25]. However, numeric calculations rendering actual numbers of isomers are considerably faster.

X	n = 1	n = 2	n = 3	n = 6	n = 10	n = 25	n = 50	n = 100
1	1	2	3	6	10	25	50	100
2	3	9	18	63	165	975	3,825	15,150
3	3	18	55	398	1,770	26,675	210,850	1,676,700
4	3	29	126	1,773	13,125	496,875	7,878,125	125,512,500
5	1	20	135	3,996	50,500	4,890,625	156,312,500	5,000,500,000
6	1	13§	92$	4,291	86,185	20,448,025	1,303,687,925	83,358,668,350
Sum	12§	91$	429	10,527	151,755	25,863,200	1,468,093,275	88,486,732,800
Sum / 2	4	10	24	167	920	26,526	383,815	5,840,708

Table 1: Number of isomers of substituted benzenes (**4**) listed as a function of the number of occupied substituent sites X. § and $: Related numbers, see text above. Sum is the sum over all numbers in one column for $X = 1$ to $X = 6$. The fraction in the number of isomers for *Sum* over $X = 2$ is presented in the last row as rounded integer.

In the following tables, molecules with the same enumeration results have been placed in one column:

X	Tetrahedranes (**30**)	Enantiomers of Substituted **30**	Achiral Isomers of Substituted **30**
1	n	0	n
2	$n*(n+1)/2$	0	$n*(n+1)/2$
3	$n*(n^2+2)/3$	$n*(n^2-3n+2)/3$	n^2
4	$n^2*(n^2+11)/12$	$n*(n^3-6n^2+11n-6)/12$	$n*(n^2+1)/2$

Table 2: Formulas for the number of substituted isomers, enantiomers and achiral isomers listed as a function of the number of occupied substituent sites X for tetrahedrane (**30**).

X	Cyclopropanes (**16**) and Prismanes (**20**)	Enantiomers of Substituted **16** and **20**	Achiral Isomers of Substituted **16** and **20**
1	n	0	n
2	$n*(5n + 3) / 2$	$2n^2$	$n*(n + 3) / 2$
3	$2n*(5n^2 + 1) / 3$	$2n*(5n^2 - 3n + 1) / 3$	$2n^2$
4	$n^2*(5n^2 + 3) / 2$	$n^2*(5n^2 - 2n + 1) / 2$	$n^2*(n + 1)$
5	n^5	$n^3*(n^2 - 1)$	n^3
6	$n^2*(n^4 + 3n + 2) / 6$	$n*(n^5 - 3n^3 + 2n^2 + 2n - 2) / 6$	$n*(3n^3 + n^2 + 2) / 6$

Table 3: Formulas for the number of substituted isomers, enantiomers and achiral isomers listed as a function of the number of occupied substituent sites *X* for cyclopropane (**16**) and prismane (**20**).

X	Acenaphthylene (**1**)and as-Indacene (**9**)	Azulene (**3**)	Pentalene (**8**)	s-Indacene (**10**)	Naphthalene (**11**)
1	4n	5n	2n	3n	2n
2	$2n*(7n + 1)$	$n*(29n + 3) / 2$	$2n*(2n + 1)$	$5n*(3n + 1) / 2$	$n*(7n + 3)$
3	$28n^3$	$n^2*(28n + 3)$	$n^2*(5n + 1)$	$n^2*(14n + 3)$	$14n^3$
4	$n^2*(35n^2 + 3)$	$n^2*(70n^2 + 3n + 3) / 2$	$n^2*(15n^2 + 2n + 7) / 4$	$n^2*(35n^2 + 3n + 6) / 2$	$n^2*(35n^2 + 9) / 2$
5	$28n^5$	$n^3*(28n^2 + 3)$	$n^3*(3n^2 + 1) / 2$	$n^3*(14n^2 + 3)$	$14n^5$
6	$2n^3*(7n^3 + 1)$	$n^3*(28n^3 + 3n + 1) / 2$	$n^3*(n^3 + n + 2) / 4$	$n^3*(14n^3 + 3n + 3) / 2$	$n^3*(7n^3 + 3)$
7	$4n^7$	$n^4*(4n^3 + 1)$		$n^4*(2n^3 + 1)$	$2n^7$
8	$n^4*(n^4 + 1) / 2$	$n^5*(n^3 + 1) / 2$		$n^4*(n^4 + 2n + 1) / 4$	$n^4*(n^4 + 3) / 4$

Table 4: Formulas for the number of isomers of substituted molecules with 6 to 8 substitution sites listed as a function of the number of occupied substituent sites *X*.

X	Anthracene (**2**) and Pyrene (**14**)	Corannulene (**6**)	Phenanthrene (**13**)
1	3n	n	5n
2	$n*(23n + 7) / 2$	$n*(9n + 5) / 2$	$5n*(9n + 1) / 2$
3	$2n^2*(15n + 1)$	$12n^3$	$60n^3$
4	$n^2*(105n^2 + 2n + 13) / 2$	$n^2*(21n^2 + 5)$	$5n^2*(21n^2 + 1)$
5	$3n^3*(21n^2 + 1)$	$2n*(63n^4 + 2) / 5$	$126n^5$
6	$3n^3*(35n^3 + n + 4) / 2$	$n^3*(21n^3 + 5)$	$5n^3*(21n^3 + 1)$
7	$2n^4*(15n^3 + 1)$	$12n^7$	$60n^7$
8	$n^4*(45n^4 + 4n + 11) / 4$	$n^4*(9n^4 + 5) / 2$	$5n^4*(9n^4 + 1) / 2$
9	$n^5(5n^4 + 1) / 2$	n^9	$5n^9$
10	$n^5*(n^5 + n + 2) / 4$	$n^2*(n^8 + 5n^3 + 4) / 10$	$n^5*(n^5 + 1) / 2$

Table 5: Formulas for the number of isomers of substituted molecules with 10 substitution sites listed as a function of the number of occupied substituent sites *X*.

X	Chrysene (**5**)	Coronene (**7**)	Perylene (**12**)	Triphenylene (**15**)
1	6n	n	3n	2n
2	$3n*(11n+1)$	$n*(11n+7)/2$	$3n*(11n+3)/2$	$n*(11n+3)$
3	$110n^3$	$n*(55n^2+2)/3$	$55n^3$	$2n^3*(55n^2+2)/3$
4	$15n^2*(33n^2+1)/2$	$5n^2*(33n^2+7)/4$	$45n^2*(11n^2+1)/4$	$15n^2*(11n^2+1)/2$
5	$396n^5$	$66n^5$	$198n^5$	$132n^5$
6	$2n^3*(231n^3+5)$	$n*(231n^5+35n^2+3n+1)/3$	$3n^3*(77n^3+5)$	$2n^2*(77n^3+5n+1)$
7	$396n^7$	$66n^7$	$198n^7$	$132n^7$
8	$15n^4*(33n^4+1)/2$	$5n^4*(33n^4+7)/4$	$45n^4*(11n^4+1)/4$	$15n^4*(11n^2+1)/2$
9	$110n^9$	$n^3*(55n^6+2)/3$	$55n^9$	$2n^3*(55n^6+2)/3$
10	$3n^5*(11n^5+1)$	$n^5*(11n^5+7)/2$	$3n^5*(11n^5+3)/2$	$n^5*(11n^5+3)$
11	$6n^{11}$	n^{11}	$3n^{11}$	$2n^{11}$
12	$n^6*(n^6+1)/2$	$n^2*(n^{10}+7n^4+2n^2+2)/12$	$n^6*(n^6+3)/4$	$n^4*(n^8+3n^4+2)/6$

Table 6: Formulas for the number of isomers of substituted molecules with 12 substitution sites listed as a function of the number of occupied substituent sites *X*.

In the following tables, only the number of isomers and the number of enantiomers (not enantiomeric pairs !) are listed, the according figures for the achiral derivatives can be obtained by substraction or from the *EXCEL* - chart[25]:

X	Cyclopentane (**18**) and [5]-Prismane (**21**)	Enantiomers of Substituted **18** and **21**	Cyclohexane (**19**) and [6]-Prismane (**22**)	Enantiomers of Substituted **19** and **22**
1	n	0	n	0
2	$n*(9n+5)/2$	$2n^2$	$n*(11n+7)/2$	$4n^2$
3	$12n^3$	$2n^2*(3n-1)$	$n*(55n^2+2)/3$	$2n^2*(26n^2-6n+1)/3$
4	$n^2*(21n^2+5)$	$n^2*(21n^2-2n+1)/2$	$5n^2*(33n^2+7)/4$	$n^2*(41n^2-6n+1)$
5	$2n*(63n^4+2)/5$	$n*(63n^4-15n^2+2)/5$	$66n^5$	$2n^3*(33n^2-2n-3)$
6	$n^3*(21n^3+5)$	$n^3*(21n^3-3n+2)/2$	$n*(231n^5+35n^2+3n+1)/3$	$n*(231n^5-3n^4-27n^3+7n^2+3n-1)/3$
7	$12n^7$	$2n^4*(3n^3-1)$	$66n^7$	$2n^4*(33n^3-3n-2)$
8	$n^4*(9n^4+5)/2$	$n^4*(9n^4-4n+3)/4$	$5n^4*(33n^4+7)/4$	$3n^4*(55n^4-2n^2-8n+3)/4$
9	n^9	$n^5*(n^4-1)/2$	$n^3*(55n^6+2)/3$	$n^3*(55n^6-12n^4-3n^2+2)/3$
10	$n^2*(n^8+5n^3+4)/10$	$n*(n^9-5n^5+4n^4+4n-4)/20$	$n^5*(11n^5+7)/2$	$n^5*(11n^5-2n^2-3n+2)/2$
11			n^{11}	$n^2*(n^9-1)$
12			$n^2*(n^{10}+7n^4+2n^2+2)/12$	$n^2*(n^{10}-3n^6+2n^4+2n^2-2)/12$

Table 7: Formulas for the number of substituted isomers and enantiomers for cyclopentane (**18**), [5]-prismane (**21**), cyclohexane (**19**), and [6]-prismane (**22**) listed as a function of the number of occupied substituent sites *X*.

X	Cyclobutane (**17**)	Enantiomers of Substituted **17**	Cubane (**25**)	Enantiomers of Substituted **25**
1	n	0	n	0
2	$n*(7n + 5) / 2$	$2n^2$	$3n*(n + 1) / 2$	0
3	$7n^3$	$2n^2*(3n - 1)$	$n*(7n^2 + 2) / 3$	$2n^2*(2n^2 - 3n + 1) / 3$
4	$n*(35n^3 + 15n + 2) / 4$	$n^2*(17n^2 - 6n + 1) / 2$	$n*(35n^3 + 43n + 6) / 12$	$n^2*(8n^2 - 9n + 7) / 3$
5	$7n^5$	$n^3*(7n^2 - 2n - 1)$	$n^3*(7n^2 + 2) / 3$	$n^3*(7n^2 - 6n - 1) / 3$
6	$n^3*(7n^3 + 5) / 2$	$n^3*(7n^3 - n^2 - 3n + 1) / 2$	$n^2*(7n^4 + 9n + 2) / 6$	$n*(7n^5 - 3n^4 - 9n^3 + 5n^2 + 2n - 2) / 6$
7	n^7	$n^5*(n^2 - 1)$	$n^3*(n^4 + 2) / 3$	$n^3*(n^4 - 3n^2 + 2) / 3$
8	$n^2*(n^6 + 5n^2 + 2) / 8$	$n^4*(n^2 - 1)^2 / 8$	$n^2*(n^6 + 17n^2 + 6) / 24$	$n^2*(n^6 - 6n^4 + 13n^2 - 8) / 24$

Table 8: Formulas for the number of substituted isomers, enantiomers and achiral isomers listed as a function of the number of occupied substituent sites X for cyclobutane (**17**) and cubane (**25**).

X	Dibenzo[2.2]para-cyclophane (**24**)	Enantiomers of Substituted **24**	Achiral Isomers of Substituted **24**
1	4n	n	2n
2	$6n*(5n + 1)$	$n*(23n - 1)$	$7n*(n + 1)$
3	$140n^3$	$2n^2*(63n - 4)$	$2n^2*(7n + 4)$
4	$7n^2*(65n^2 + 3)$	$n^2*(875n^2 - 56n - 3) / 2$	$n^2*(35n^2 + 56n + 45) / 2$
5	$1,092n^5$	$2n^3*(539n^2 - 28n - 6)$	$2n^3*(7n^2 + 28n + 6)$
6	$14n^3*(143n^3 + 3)$	$n^3*(1,995n^3 - 70n^2 - 42n - 1)$	$n^3*(7n^3 + 70n^2 + 42n + 43)$
7	$2,860n^7$	$2n^4*(1,429n^3 - 28n^2 - 42n - 4)$	$2n^4*(n^3 + 28n^2 + 42n + 4)$
8	$15n^4*(429n^4 + 7) / 2$	$n^4*(12,869n^4 - 112n^3 - 420n^2 - 112n - 1) / 4$	$n^4*(n^4 + 112n^3 + 420n^2 + 112n + 211) / 4$
9	$2,860n^9$	$2n^5*(1,430n^4 - 4n^3 - 42n^2 - 28n - 1)$	$2n^5*(4n^3 + 42n^2 + 28n + 1)$
10	$14n^5*(143n^5 + 3)$	$n^6*(2,002n^4 - n^3 - 42n^2 - 70n - 7)$	$n^5*(n^4 + 42n^3 + 70n^2 + 7n + 42)$
11	$1,092n^{11}$	$2n^7*(546n^4 - 6n^2 - 28n - 7)$	$2n^7*(6n^2 + 28n + 7)$
12	$7n^6*(65n^6 + 3)$	$n^8*(910n^4 - 3n^2 - 56n - 35) / 2$	$n^6*(3n^4 + 56n^3 + 35n^2 + 42) / 2$
13	$140n^{13}$	$2n^9*(70n^4 - 4n - 7)$	$2n^9*(4n + 7)$
14	$6n^7*(5n^7 + 1)$	$n^{10}*(30n^4 - n - 7)$	$n^7*(n^4 + 7n^3 + 6)$
15	$4n^{15}$	$2n^{11}*(2n^4 - 1)$	$2n^{11}$
16	$n^8*(n^8 + 3) / 4$	$n^{12}*(n^4 - 1) / 4$	$n^8*(n^4 + 3) / 4$

Table 9: Formulas for the number of substituted isomers, enantiomers and achiral isomers for Dibenzo[2.2]-paracyclophane (**24**) listed as a function of the number of occupied substituent sites X.

X	Triptycene (**23**)	Enantiomers of Substituted **23**	Achiral Isomers of Substituted **23**
1	3n	0	3n
2	n*(31n + 7) / 2	$8n^2$	n*(15n + 7) / 2
3	2n*(91n^2 + 2) / 3	4n*(38n^2 - 9n + 1) / 3	$2n^2$*(5n + 6)
4	n^2*(1,001*n^2 + 79) / 6	2n^2*(239n^2 - 45n + 10) / 3	n^2*(15n^2 + 60n + 13) / 2
5	n^3*(1,001n^2 + 4) / 3	2n^3*(496n^2 - 60n - 25) / 3	n^3*(3n^2 + 40n + 18)
6	n^2*(1,001n^4 + 35n + 4) / 2	n*(1,500n^5 - 90n^4 - 135n^3 + 29n^2 + 6n - 2) / 3	n*(3n^5 + 180n^4 + 270n^3 + 47n^2 + 4) / 6
7	4n^3*(143n^4 + 1)	4n^3*(143n^4 - 3n^3 - 15n^2 - 3n + 1)	12n^4*(n^2 + 5n + 1)
8	13n^4*(77n^4 + 3) / 2	n^2*(3,003n^6 - 12n^5 - 270n^4 - 180n^3 + 79n^2 - 4) / 6	n^2*(6n^5 + 135n^4 + 90n^3 + 19n^2 + 2) / 3
9	n^3*(1,001n^6 + 4) / 3	n^3*(1,001n^6 - 54n^4 -120n^3 - 9n^2 + 4) / 3	n^3*(18n^2 + 40n + 3)
10	n^4*(1,001n^6 + 63n + 16) / 6	n^4*(1,001n^6 - 18n^4 - 180n^3 - 45n^2 + 42n + 16) / 6	n^3*(6n^3 + 60n^2 + 15n + 7) / 2
11	2n^5*(91n^6 + 2) / 3	2n^5*(91n^6 - 18n^3 - 15n^2 + 2) / 3	2n^7*(6n + 5)
12	n^4*(91n^8 + 21n^2 + 2) / 6	n^2*(91n^{10} - 12n^7 - 45n^6 + 14n^4 + 2n^2 - 2) / 6	n^2*(12n^7 + 45n^6 + 7n^4 + 2) / 6
13	n^5*(7n^8 + 2) / 3	n^5*(7n^8 - 9n^4 + 2) / 3	3n^9
14	n^6*(n^8 + 3n + 2) / 6	n^3*(n^{11} - 3n^7 + 2n^4 + 2n^3 - 2) / 6	n^3*(3n^7 + n^4 + 2) / 6

Table 10: Formulas for the number of substituted isomers, enantiomers and achiral isomers for triptycene (**23**) listed as a function of the number of occupied substituent sites X.

For the following tables 11, 12, 13 and 14 of the fulleranes **26** and **27** as well as the fullerenes **28** and **29** only the enumerations for up to ten occupied substitution sites are shown. Higher values for X are available in the *EXCEL* -chart[25].

X	C_{20}-[5]-Fullerane (Dodecahedrane, **26**)	Enantiomers of Substituted **26**
1	n	0
2	n*(7n + 5) / 2	$2n^2$
3	n*(19n^2 + 2)	2n*(9n^2 - 4n + 1)
4	n^2*(323n^2 + 61) / 4	n^2*(161n^2 - 24n + 15) / 2
5	2n*(646n^4 + 5n^2 + 4) / 5	2n*(646n^4 - 20n^3 - 65n^2 + 4) / 5
6	n^2*(646n^4 + 30n + 5)	n^2*(646n^5 - 2n^4 - 42n^3 + 14n^2 + 5n - 1)
7	2n^3*(646n^4 + 5)	2n^3*(646n^4 - 14n^2 - 28n + 5)
8	n^4*(4,199n^4 + 115) / 2	n^2*(4,199n^6 - 14n^4 - 168n^3 + 73n^2 - 2) / 2
9	2n^3*(4,199n^6 + 10) / 3	2n^3*(4,199n^6 - 84n^3 - 105n^2 + 10) / 3
10	n^2*(46,189n^8 + 945n^3 + 200n^2 + 36) / 15	n*(46,189n^9 - 210n^6 - 1,575n^5 + 672n^4 + 200n^3 + 36n - 12) / 15

Table 11: Formulas for the number of isomers of C_{20}-[5]-fullerane (**26**, dodecahedrane) including the number of enantiomers listed as a function of the number of occupied substituent sites X.

X	C_{24}-[4,6]-Fullerane (**27**)	Enantiomers of Substituted **27**
1	n	0
2	$n*(23n + 9) / 2$	$8n^2$
3	$n*(253n^2 + 8) / 3$	$8n*(29n^2 - 3n + 1) / 3$
4	$n*(1,771n^3 + 99n + 6) / 4$	$2n^2*(217n^2 - 14n + 1)$
5	$1,771n^5$	$28n^3*(63n^2 - 2n - 1)$
6	$n^2*(33,649n^4 + 495n + 56) / 6$	$2n*(8,407n^5 - 105n^4 - 147n^3 + 17n^2 + 14n - 2) / 3$
7	$14,421n^7$	$28n^4*(515n^3 - 2n^2 - 7n - 2)$
8	$3n^2*(81,719n^6 + 495n^2 + 10) / 8$	$n^4*(61,289n^4 - 56n^3 - 490n^2 - 392n + 65) / 2$
9	$2n^3*(81,719n^6 + 28) / 3$	$2n^3*(81,719n^6 - 12n^5 - 294n^4 - 588n^3 - 105n^2 + 28) / 3$
10	$11n^5*(7,429n^5 + 27)$	$n^5*(81,719n^5 - n^4 - 98n^3 - 490n^2 - 245n + 59)$

Table 12: Formulas for the number of isomers of C_{24}-[4,6]-fullerane (**27**) including the number of enantiomers listed as a function of the number of occupied substituent sites X.

X	C_{36}-[5.6]-Fullerene (**28**)	Enantiomers of Hetero-**28**
1	3n	0
2	$21n*(5n + 1) / 2$	$44n^2$
3	$n*(595n^2 + 2)$	$2n*(290n^2 - 22n + 1)$
4	$357n^2*(55n^2 + 1) / 4$	$n^2*(4,891n^2 - 122n + 11)$
5	$31,416n^5$	$2n^3*(15,701n^2 - 106n - 151)$
6	$n*(162,316n^5 + 476n^2 + 11n + 1)$	$n*(162,309n^5 - 249n^4 - 817n^3 + 109n^2 + 11n - 1)$
7	$695,640n^7$	$2n^4*(347,819n^3 - 98n^2 - 697n - 644)$
8	$255n^3*(9,889n^4 + 7)$	$n^4*(10,086,779n^4 - 392n^3 - 6,490n^2 - 13,552n + 2,279) / 4$
9	$110n^3*(213,962n^6 + 1) / 3$	$2n^3*(11,767,910n^6 - 42n^5 - 1,911n^4 - 8,484n^3 - 5,733n^2 + 55) / 3$
10	$714n^5*(29,667n^5 + 7)$	$7n^5*(6,052,068n^5 - n^4 - 182n^3 - 1,860n^2 - 2,782n + 565) / 2$

Table 13: Formulas for the number of isomers and enantiomers of heterofullerenes of **28** (carbons replaced by e.g. silicon) listed as a function of the number of replaced carbons X.

X	C_{60}-Fullerene (**29**)	Enantiomers of Hetero-**29**
1	n	0
2	$n*(59n + 15) / 2$	$14*n^2$
3	$n*(1,711n^2 + 20) / 3$	$2n*(427n^2 - 21n + 5) / 3$
4	$29n^2*(1,121n^2 + 15) / 4$	$7n^2*(1,161n^2 - 6n + 1) / 2$
5	$2n*(227,563n^4 + 12) / 5$	$n*(227,563n^4 - 70n^3 - 945n^2 + 12) / 5$
6	$n^2*(2,503,193n^4 + 3,045n + 190) / 3$	$n*(2,503,193n^5 - 21n^4 - 1,701n^3 + 385n^2 + 190n - 10) / 6$
7	$6,436,782n^7$	$9n^4*(357,599n^3 - 21n - 182)$
8	$261n^4*(653,543n^4 + 105) / 4$	$9n^4*(18,952,747n^4 - 42n^2 - 2,184n + 567) / 8$
9	$19n^3*(12,967,669n^6 + 20)$	$n^4*(246,385,711n^6 - 3,276n^5 - 20,475n^2 + 380) / 2$
10	$3n^2*(4,188,557,087n^8 + 118,755n^3 + 88) / 10$	$3n*(4,188,557,087n^9 - 2,730n^6 - 102,375n^5 + 28,938n^4 + 88n - 8) / 20$

Table 14: Formulas for the number of isomers and enantiomers of heterofullerenes of **29** (carbons replaced by e.g. silicon) listed as a function of the number of replaced carbons X.

For the discussion of the number of inositoyl-dimers and hexa-O-inositoyl-inositols, the respective numbers of interest are given in table 15. Again, parent compounds with equal enumeration results are combined (*epi-* and *myo-* as well as *muco-* and *neo*-inositol) and the full tables (including $X = 2$ to 5) can be obtained from the *EXCEL* -chart[25]:

Inositol	*X*	Total Isomers	Enantiomers	Achiral Isomers
cis	*1*	n	0	n
	6	$n*(n^5 + n^2 + 2n + 2) / 6$	$n*(n^5 - 3n^3 - 2n^2 + 2n + 2) / 6$	$n^{3}(n + 1) / 2$
epi or myo	*1*	6n	4n	2n
	6	n^6	$n^4*(n^2 - 1)$	n^4
allo	*1*	6n	6n	0
	6	n^6	$n^3*(n^3 - 1)$	n^3
muco or neo	*1*	3n	2n	n
	6	$n^3*(n^3 + 1) / 2$	$n^4*(n^2 - 1) / 2$	$n^{3}(n + 1) / 2$
chiro	*1*	6n	6n	0
	6	$(n^6 + n^3) / 2$	$(n^6 + n^3) / 2$	0
scyllo	*1*	n	0	n
	6	$n^2*(n^4 + 3n + 2) / 6$	$n*(n^5 - 3n^3 + 2n^2 + 2n - 2) / 6$	$n^{2}(3n^2 + n^2 + 2) / 6$

Table 15: Formulas for the number of isomers, enantiomers and achiral isomers of mono- and hexa-substituted inositols as a function of occupied positions *X*

Conclusions

From these tables, the following conclusions can be drawn: We begin with the answers to the questions posed in the introduction.

- The relation between the number of disubstituted derivatives (pairs) and the total number of benzene-isomers becomes bigger the more different substituents one considers. This is trivial. The exact growth of this relation however can only be calculated by our program. Thus the predictive power of any carbon-13 NMR chemical shift estimation program using pairwise correction terms as parameters for benzene shifts[2-4] theoretically would increase with the number of substituents, if all disubstituted pairs would be known experimentally. Unfortunately, already with 50 different substituents, ^{13}C-NMR-data on 3,825 pairs would be needed, which clearly is not available from the literature. Therefore, other approaches for larger sets of functional groups in such prediction programs will be needed.
- With three different substituents, there are 27 benzene isomers in case of 1,2,4- and 10 isomers in case of 1,3,5-trisubstitution[26].

All halogenated pyrenes (four different halogens plus hydrogen as substituents) amount to as many as 2,446,874 isomers.

- Seven different substituents at benzene yield a family of 10,528 isomers.
- With a benzene skeleton and just 25 different substituent types one can assemble as many substitutional isomers as there are chemical compounds presently known in the scientific literature[27].
- 100 substituent types - a number, which many shift prediction programs easily provide - give rise to over $8 * 10^{10}$ isomers, an incredible number in terms of experimental verification.
- A heteroisomer enumeration of C_{60}-fullerene with 8 silicon atoms replacing carbon ($C_{52}Si_8$) already yields over 42 million isomers, most of them being enantiomers.

There are 32 monosubstituted inositols in total. One obtains this number by setting $X = 1$ and $n = 1$, and summing all applicable numbers of isomers in the „Isomer" column of table 15. The numbers for *epi*- and *myo*- as well as *muco*- and *neo*-inositol have to be counted twice, since those numbers each represent two symmetrically identical, but nevertheless individual inositols. Of those monosubstituted isomers, only 8 are achiral ($X = 1$, $n = 1$, sum over all achiral isomers counts). Taking the one monosubstitueted *cis*-inositol for example, we have 32 monosubstitutable inositol isomers to attach, thereby generating 32 dimers. For the complete series of inositols we count 528 dimers. This number simply represents the $32 * (32 + 1) / 2$ pairings of all 32 monosubstitutable inositols. Note, that a dimer assembled from an A-B inositol pairing is the same as a B-A dimer. Consequently one has to subtract $32 * (32 - 1) / 2$ redundant dimers from the set of $32 * 32$ ordered pairings, thus deriving $32 * (32 + 1) / 2$. Of the 528 isomeric dimers, only $8 * (8 + 1) / 2$ are combinations of achiral inositols. Adding to this another 12 dimers for achiral *meso*-forms resulting from a combination of two enantiomeric inositols, (for $X = 1$, count the enantiomeric pairs of all inositols for $n = 1$ in table 15), one obtains 48 achiral dimers of inositols in total. The number of 990 possible dimeric inositols given in[6] thus was overestimated. In the series of hexa-O-inositoyl-inositols you could expect e.g. 178,962,784 isomers for *cis*-inositol ($X = 6$, $n = 32$), from which at least 2,304 would be achiral ($X = 6$, $n = 8$). The exact calculation of the additional achiral *meso*-forms due to symmetry in chirally substituted derivatives is non-trivial and will be one of our future projects.

- Calculating these general formulas in „n" by hand is virtually impossible, we have tried it at the beginning. So this program with its simple input should help all interested readers to calculate any isomer number of polysubstituted molecules. The presented results should enable researchers as well as teachers to explore the wealth of possible isomers in this field, including enantiomers and heteroisomers of fullerenes and questions arising in the field of inositols.

References and Notes

(1) Almsick, M.van; Dolhaine, H.; Hönig, H. An Efficient Algorithm to Enumerate Diamutamers with more than one Type of Substituent. *J. Chem. Inf. Comput. Sci.* **1999**, 39 (submitted).

(2) Hönig, H. An Improved ^{13}C-NMR-Shift Prediction Program for Polysubstituted Benzenes and Sterically Defined Cyclohexane Derivatives. *Magn. Reson. Chem.* **1996**, 34, 395 - 406.

(3) Hönig, H. HIPPO-CNMRS: **H**ighly **I**mproved **P**rediction **P**rogram **O**f **C**arbon **N**uclear **M**agnetic **R**esonance **S**hifts. *J. Chem. Educ.* **1996**, 73, 1078 and *J. Chem. Educ.: Software* **1997,** 4D, 30-40.

(4) Thomas, St.; Brühl, I.; Heilmann, D.; Kleinpeter, E. ^{13}C NMR Chemical Shift Calculation for Some Substituted Pyridines: A Comparative Consideration. *J. Chem. Inf. Comput. Sci.* **1997**, 37, 726-730.

(5) see e.g. Katritzky, A. R.; Maran, U.; Karelson, M.; Lobanov, V. S. Prediction of Melting Points for the Substituted Benzenes: A QSPR Approach. *J. Chem. Inf. Comput. Sci.* **1997**, 37, 913-919.

(6) T. Hudlicky, Univ. of Florida, Gainesville at a seminar given during his sabbatical stay at the Technical University Graz, passed on by L. Brecker. See also the introduction in: Hudlicky, T.; Abboud, K.A.; Entwistle, D.A.; Fan, R.; Maurya, R.; Thorpe, A.J.; Bolonick, J.; Myers, B. Toluene-Dioxygenase-Mediated cis-Dihydroxylation in Enantioselective Synthesis. Iterative Glycoconjugate Coupling Strategy and Combinatorial Design for the Synthesis of Oligomers of nor-Saccharides, Inositols and Pseudosugars with Interesting Molecular Properties. *Synthesis* **1996**, 897-911.

(7) Balasubmaranian, K. Applications of Combinatorics and Graph Theory to Spectroscopy and Quantum Chemistry. *Chem. Rev.* **1985**, 85, 599-618.

(8) See e.g. corrections by reference 20 of values calculated in reference 21 or our comments to reference 18 in reference 1 or the comments in this paper to values from reference 6.

(9) An excellent overview of developments up to 1974 can be found in: Rouvray, D.H. Isomer Enumeration Methods. *Chem. Soc. Rev.* **1974**, 355-372.

(10) Harary, F; Palmer, E. M.; Robinson, R. W.; Read, R.C. Pólya's contribution to chemical theory. In *Chemical Applications of Graph Theory;* Balaban, A. T., Ed.; Academic Press, London, 1976, p. 11-24.

(11) Hässelbarth, W. A Note on Pólya enumeration theory. *Theoret. Chim. Acta. (Berl.)* **1984**, 66, 91-110.

(12) Dolhaine, H.; Zerfaß, H.-R. Substitutionsisomere und Symmetrie - Ein Computer-Programm zur Erzeugung von Formeln der Substitutionsisomeren eines Moleküls. *Chem. Ztg.* **1981**, 105, 365-369.

(13) Dolhaine, H. A Computer Program for the Enumeration of Substitutional Isomers. *Computers & Chemistry* **1981**, 5, 41-48. See also the note by Dolhaine, H.; Garavelli, J. S.; Leonard, J. E. Comments on Papers Concerning Computer Enumeration of Permutation Isomers. *Computers & Chemistry* **1986**, 10, 239-240.

(14) Dolhaine, H. Spins und Substitutionsisomerie. *Chem. Ztg.* **1980**, 104, 287-290.

(15) Hansen P. J.; Jurs, P.C. Chemical Applications of Graph Theory. Part II. Isomer Enumeration. *J. Chem. Ed.* **1988**, 65, 661-664.

(16) Novak, I. Inverse Problem of Isomer Enumeration. *J. Chem. Ed.* **1994**, 71, 512-513.

(17) Garavelli, J. S.; Leonhard, J. E. Improvements in the Computer Enumeration of Permutation Isomers. *Computers & Chemistry* **1985**, 9, 133-147.
(18) Leonhard, J. E. Isomer Numbers of Nonrigid Molecules. The Cyclohexane Case. *J. Phys. Chem.* **1977**, 81, 2212-2214.
(19) Balasubramanian, K. Combinatorial Enumeration of Chemical Isomers. *Ind. J. Chem.* **1978**, 16B, 1094-1096.
(20) Balasubramanian, K. Enumeration of Chiral and Positional Isomers of Substituted Fullerene Cages (C_{20} - C_{70}). *J. Phys. Chem.* **1993**, 97, 6990-6998.
(21) Fujita, S. Unit Subduced Cycle Indices with and without Chirality Fittingness for I_h Group. An Application to Systematic Enumeration of Dodecahedrane Derivatives. *Bull. Chem. Soc. Jpn.* **1990**, 63, 2759-2769.
(22) Kerber, A.; Laue, R.; Wieland, T. Discrete Mathematics for Combinatorial Chemistry. Manuscript in preparation. See also: Wieland, T. Combinatorics of combinatorial chemistry. *J. Math. Chem.* **1997**, 21, 141-157.
(23) Lunn, C.; Senior, J.K. Isomerism and Configuration. *J. Phys. Chem.* **1929**, 33, 1027-1079. For benzene see pages 1054 f.
(24) e.g. *Mathematica*: A System for Doing Mathematics by Computer, Wolfram Res. Inc., Champain, Ill. USA. This program was i.a. chosen because it is in widespread use within the scientific community. See e.g.: Davies, N. D. Computing in Spectroscopy: Discussion Meeting at EUCMOS XXIV in Prague - 27 August 1998. *Spectroscopy Europe* **1998**, 10/4, 22-24.
(25) http://www-orgc.tu-graz.ac.at/research/softnew.htm
(26) See also a very recent article on benzene and coronene: Fujita, S. Benzene Derivatives with Achiral and Chiral Substituents and relevant Derivatives Derived from D_{6h} Skeletons. Symmetry-Itemized Enumeration and Symmetry Characterization by the Unit-Subduced Cycle Index Approach. *J. Chem. Inf. Comput. Sci.* **1999**, 39, 151-163.
(27) As of 11 / 1998, the CAS-Registry file contained approximately 18.6 million compounds.