Pergamon

Tetrahedron Vol. 51, No. 13, pp. 3831-3840, 1995
Copyright © 1995 Elsevier Science Ltd
Printed in Great Britain. All rights reserved
0040-4020/95 $9.50+0.00

0040-4020(95)00106-9

On the Application of Sieve Formula to the Enumeration of Stable Stereo and Position Isomers of Deoxyaldytols.

Robert M. NEMBA*, Maurice FAH

Laboratory of Physical Chemistry, Faculty of Science, University of Yaounde I,
PO. Box 812 Yaounde, Cameroon.

Abstract : *The method of sieve formula based on the principle of inclusion and exclusion of non valid configurations is developed for the enumeration of stable stereoisomers of polyalcohols. An application is shown in the series of deoxyalditols* $C_nH_{n+3}(OH)_{n-1}$ *where* $3 \leq n \leq 10$.

INTRODUCTION

Deoxyalditols are acyclic polyalcohols presenting both constitutional isomerism due to the possible displacement of the position of the unsubstituted carbon atom along the linear carbon chain and the existence of optical stereoisomerism relevant to the alternant positions of the various hydroxy groups located on asymmetric carbon atoms. It must be recalled that despite the wide occurrence of deoxyaldoses in nature(e.g. deoxyribose), only one deoxyalditol has been found, namely 1 deoxy-D-glycero-D-talo-heptitol (siphulitol) as a constituent of the lichen Siphula ceratites [1]. Terminal deoxyalditols are conveniently synthesized via 1-alkylthio-1-deoxy derivatives by desulfuration with raney nickel of the corresponding aldose diakyl dithioacetals[2]. Alternatively they can be prepared by reduction of alditol substituted at a primary position by an alkane or arene-sulfonoxy group using lithium aluminium hydride[3], or of a deoxyaldose in which the deoxy group is at terminal position or of toluene-β-sulphonylhydrazones of aldoses with sodium borohydride[4].

Deoxyalditols with the deoxy group at a secondary carbon atom are in general prepared by reduction of the corresponding deoxyaldose[5]. The enumeration of deoxyalditols is not convenient by the classical methods and the unwieldiness of this problem is in part due to the fact that polyalcohols that have more than one hydroxy group on a carbon atom are unstable and represent non valid structures. The purpose of this paper is to apply Balasubramanian' s method[6] for counting stable stereo and position isomers of deoxyalditols with the formula $C_nH_{n+3}(OH)_{n-1}$. The results are presented for $3 \leq n \leq 10$.

MATHEMATICAL APPROACH

Let us represent the tridimensional skeleton of an acyclic polyalcohol by a stereograph shown in figure 1 where the labeled vertices of degree 4 are carbon atoms while those of degree 1 are unspecified and considered as boxes into which (n+3) hydrogen atoms and (n-1) OH groups are to be put. The problem is to find the number of ways we can map the unspecified vertices into the set consisting of (n+3)H and (n-1)OH. Thus we have an enumeration problem of Polya-type[7, 8]. Our task is to discover the appropriate permutation group and its cycle index. In so doing we can partition the molecular graph shown in fig. 1 into 4 subgraphs : Q, T_0, T_1 and T_2 represented in fig. 2.

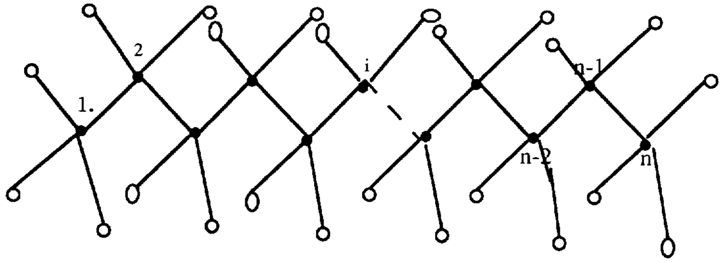

Figure 1 : Molecular stereograph of a linear polyalcohol $C_nH_{n+3}(OH)_{n-1}$

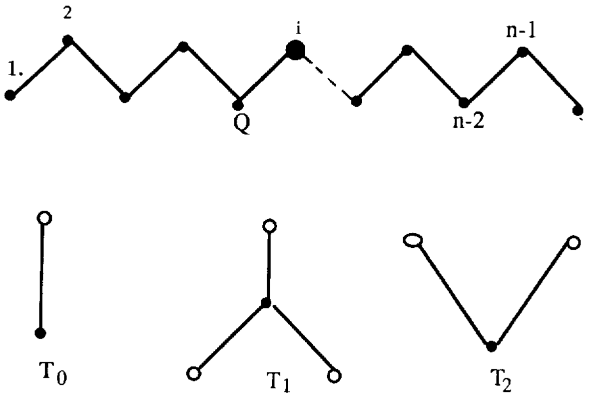

Figure 2 : Subgraphs Q, T_0, T_1 and T_2 obtained from the partition of the molecular stereograph shown in Figure 1.

The molecular graph shown in figure 1 consists of n-2 subgraphs T_2 , two subgraphs T_1 and the subgraph Q. We claim that the molecular graph shown in figure 1 can be constructed by attaching the roots of (n-2) subgraphs of type T_2 and (n-2) internal vertices of Q and then attaching 2 subgraphs of type T_1 and the two external vertices of Q. Let Q be the subgraph consisting of n labeled vertices of degree 4 (C atoms) joined by (n-1) edges (C-C bonds) to form a linear chain. The vertices in Q are permutable by the group of permutation S_2 whose cycle index is :

$$Z(S_2) = \frac{1}{2}\left(s_1^n + s_2^{n/2}\right) \quad \text{(if n even)} \tag{1}$$

or

$$Z(S_2) = \frac{1}{2}\left(s_1^n + s_1 s_2^{\frac{n-1}{2}}\right) \quad \text{(if n odd)} \tag{2}$$

Let T_0,T_1 and T_2 be the subgraphs consisting respectively of 1, 3 and 2 unspecified vertices attached to a central root (C atom). The 3 equivalent vertices of T_1 are permutable by the symmetric group S_3 whose cycle index is :

$$Z(S_3) = \frac{1}{3}\left(s_1^3 + 2s_3\right) \tag{3}$$

If we place H and OH into the two boxes of T_2 we see that no permutation of the hydrogen atom and the hydroxyl group is possible. Hence the permutation group acting on T_2 is the identity group E_2 of order 2 whose cycle index is :

$$Z(E_2) = s_1^2 \tag{4}$$

The permutation group acting on T_0 is the identity group E whose cycle index is $Z(E) = s_1$.The problem of enumerating stable stereoisomers of deoxypolyalcohols with the formula $C_nH_{n+3}(OH)_{n-1}$ is solvable if we can determine the permutation group acting on the whole molecular skeleton. Let us firstly suppose that the number n of carbon atoms is even. This time we can reverse the diagram in fig. 1 end for end and have an equivalent scheme for the molecule. The group of permutations for H_i atoms and OH_i groups located in the left hand half of the molecule at the sites i = 1, 2, 3, ...,n /2 comprising one type T_1 at one extremity and (n/2-1) types T_2 is given by the direct product[9] :

$$D_p = S_3 \times E_{2,2} \times \ldots E_{2,n/2} = S_3 \,.\, E_{2(n/2-1)} = S_3.E_{n-2} \tag{5}$$

the same equation is true for the right hand half of the molecule. Let g_0 be the group of permutations for the whole molecular graph. Since the two sides of the molecule can be permuted by S_2, one may obtain g_0 by applying the theorem of the generalized composition of the group S_2 with the direct product D_p. Hence :

$$g_0 = S_2[S_3 \otimes E_{n-2}] \tag{6}$$

Such a generalized composition of groups is also called the generalized wreath product of groups[10, 11]. When n is odd each half of the molecule consists of one type T_1,$\frac{n-3}{2}$ types T_2 and one type T_0, then the above remarks apply to all unlabeled vertices except the two attached to the middle carbon atom of the chain. These latter ones can be permuted among themselves by the identity group E independently of what happens to the others. Hence, the group g_0 for the whole molecular stereograph is the generalized wreath product :

$$g_0 = S_2\left[S_3 \otimes E_{2\left(\frac{n-3}{2}\right)} \otimes E\right] = S_2[S_3 \otimes E_{n-2}] \tag{7}$$

From equation (6) and (7) we see that the group g_0 is irrespective of the parity of n. The cycle index of the group g_0 is therefore :

$$Z\left(g_0, s_k^j\right) = \frac{1}{18}\left[s_1^{2(n+1)} + 4s_1^{2(n-2)}s_3^2 + 4s_1^{2n-1}s_3 + 3s_2^{n+1} + 6s_2^{n-2}s_6\right] \text{ (if n odd or even)} \tag{8}$$

Let us associate the weights x and y to the hydrogen atom H and the OH group, respectively. Thus the figure counting series in this case is (x+y) and for any s_k^j the figure counting series is $\left(x^k + y^k\right)^j$. The transformation :

$$Z\left(g_0, s_k^j\right) \rightarrow Z\left(g_0, \left(x^k + y^k\right)^j\right) = f_0(x,y) \tag{9}$$

gives the polynomial function $f_0(x,y)$ in which the coefficient of the term $x^{n+3}y^{n-1}$, noted I(0), corresponds to the number of stereoisomers without restriction of the polyalcohol $C_nH_{n+3}(OH)_{n-1}$.

Let $\mathbb{P}_n=\{P_1, P_2, P_3,...,P_i, ..., P_m\}$ be the set of properties where P_i is the property that there are two hydroxy groups on the i^{th} carbon atom and where $m \leq n$. The number of non equivalent ways of choosing exactly r properties from the set $\mathbb{P}_n$ is given by w_r the coefficient of a^r in the expression $Z\left(S_2, s_k^j \rightarrow \left(1+a^k\right)^j\right)$ derived from Polya's theorem. For each molecular system $C_nH_{n+3}(OH)_{n-1}$ the restrictions on w_r are as follows :

$w_0 = 1$ if $r = 0$ and n odd or even; $w_r \neq 0$ if n is odd and $1 \leq r \leq \frac{n-1}{2}$ or if n is even and $1 \leq r \leq \frac{n}{2} - 1$; $w_r = 0$ otherwise.

Given n, r , $\mathbb{P}_n$ and w_r one must firstly generate the r-tuples of properties P_i and draw up the molecular stereograph of each of the w_r r-tuples(or invalid polyalcohols) and secondly determine its appropriate permutation group g_r according to the numbers and the position of the subgraphs T_0 , T_1 and T_2 along the carbon chain. If n is odd and the molecular stereograph permutable end for end by S_2 two cases have to be considered generally :

- the type T_1 is located in the 1st and the nth carbon atoms while those of type T_2 are located in the internal carbon atoms except the central carbon atom which is linked to two types T_0 (see fig. 3a).

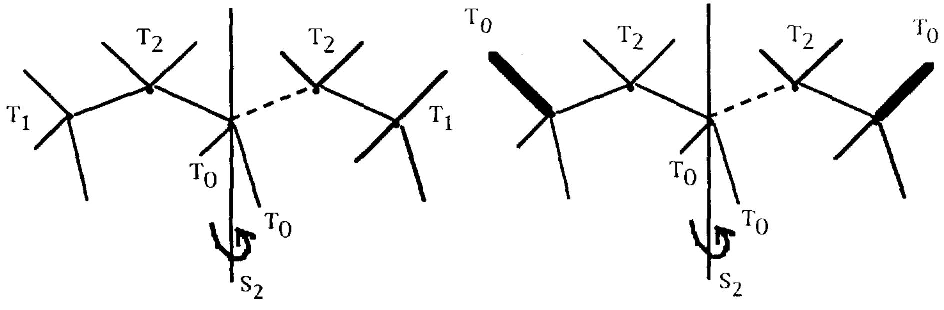

Figure 3a

Figure 3b

Figure 3 : Distributions of the subgraphs T_0 , T_1 and T_2 along the carbon chain ,when n is odd.

The permutation group of such a system is obtained from the generalized wreath product given in equation 7.

- The types T_0 are located in the 1st, nth and the central carbon atoms, while those of type T_2 are placed in the internal carbon atoms (see fig. 3b). The permutation group in this case is given by the relation :

$$g_1 = S_2\left[E_{2\beta+2}\right] \tag{10}$$

If n is even and the molecular stereograph permutable end for end by S_2,two cases have to be considered :

- two types T_1 are located in the 1st and nth carbon atoms while the types T_2 are placed in the internal carbon atoms (fig. 4a). Therefore, the permutation group is :

$$g_2 = S_2\left[S_3 \cdot E_{2\beta}\right] \tag{11}$$

- two types T_0 are located on the 1st and nth carbon atoms while the types T_2 are placed in the internal carbon atoms (see fig. 4b). The permutation group is obtained from equation (12) in this case.

$$g_3 = S_2\left[E_{2\beta+1}\right] \tag{12}$$

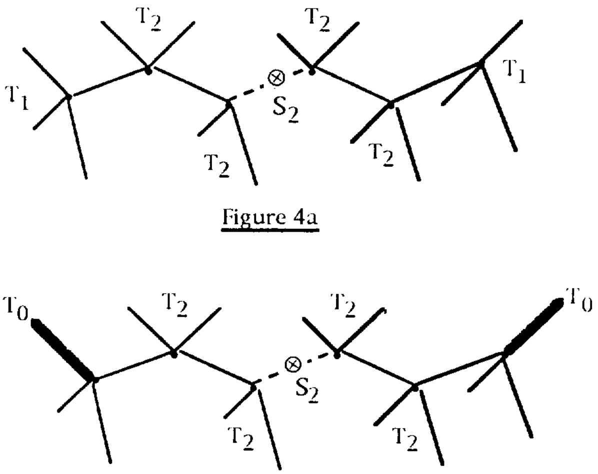

Figure 4a

Figure 4b

Figure 4 : Distributions of the subgraphs T_0, T_1 and T_2 along the carbon chain, when n is even.

In equations (10), (11) and (12) the parameter β is the number of subgraphs of type T_2 contained in one half of the molecular stereograph. When the molecular stereograph has no symmetry the permutation group is obtained from the relation :

$$g_4 = S_3^{\alpha} \ . \ E_{2\beta} \ . \ E_{\gamma} \tag{13}$$

where the parameters α, β and γ are respectively the numbers of subgraphs of types T_1, T_2 and T_0 present in the system. The cycle indices derived from the permutations groups g_1, g_2, g_3 and g_4 are given in equations (14), (15), (16) and (17) respectively :

$$\begin{aligned} Z(g_1) &= Z\left(S_2\left[E_{2\beta+2}\right]\right) \\ &= Z\left(S_2\left[\left(s_1^2\right)^{\beta+1}\right]\right) = \frac{1}{2}\left[s_1^{4(\beta+1)} + s_2^{2(\beta+1)}\right] \end{aligned} \tag{14}$$

$$\begin{aligned} Z(g_2) &= Z\left(S_2\left[S_3 \ . \ E_{2\beta}\right]\right) = Z\left(S_2\left[\frac{1}{3}\left(s_1^3 + 2s_3\right)\left(s_1^{2\beta}\right)\right]\right) \\ &= \frac{1}{18}\left[s_1^{4\beta+6} + 4s_1^{4\beta}s_3^2 + 4s_1^{4\beta+3}s_3 + 3s_2^{2\beta+3} + 6s_2^{2\beta}s_6\right] \end{aligned} \tag{15}$$

$$\begin{aligned} Z(g_3) &= Z\left(S_2\left[E_{2\beta+1}\right]\right) = Z\left(S_2\left[s_1^{2\beta+1}\right]\right) \\ &= \frac{1}{2}\left[s_1^{4\beta+2} + s_2^{2\beta+1}\right] \end{aligned} \tag{16}$$

$$
\begin{aligned}
Z(g_4) &= Z\left(S_3^{\alpha}\right) . Z\left(E_{2\beta}\right) . Z\left(E_{\gamma}\right) \\
&= \left(\frac{1}{3}\right)^{\alpha}\left(s_1^3 + 2s_3\right)^{\alpha}\left(s_1^{2\beta+\gamma}\right)
\end{aligned} \tag{17}
$$

One may notice that the terms s_k^j in equation (8) and (14) -(17) correspond to j permutations cycles of length k. Let us associate to the hydrogen atom H and the functional group (OH) the weights x and y respectively. By replacing in equations (14), (15), (16) and (17) the terms s_k^j by the figure counting series $\left(x^k + y^k\right)^j$ and expanding the resulting algebraic expression one may obtain the polynomial function :

$$
f_r(x,y) = \sum_u \sum_v C(u,v)x^u y^v \tag{18}
$$

in which if u = n+3 and v = n-1-2r, the coefficient C(n+3, n-1-2r) of the term $x^{n+3}y^{n-1-2r}$ is the number of stereoisomers of the polyalcohol statisfying exactly r properties of the set $\mathbb{P}_n$. Let I(r) be the total number of stereoisomers of all polyalcohols obtained according the w_r ways of satisfying r properties of the set $\mathbb{P}_n$.

$$
I(r) = \sum_{1}^{w_r} C(n+3, n-1-2r) \quad \text{if n even} \tag{19}
$$

and

$$
I(r) = \sum_{1}^{w_r} C(n+3, n-1-2r) - \sum_q e_q \text{ if n odd} \tag{20}
$$

where the parameter e_q is the number of stereoisomers issued from the choices q having pairs of properties (P_i, P_j) located at equivalent sites i and j. The carbon atoms i and j will be equivalent if they are transformable into each other under the action of the group S_2.

Finally the number I_n of stable stereoisomers of the system $C_nH_{n+3}(OH)_{n-1}$ that have no two hydroxy groups on the same carbon atom (i.e stereoisomers that satisfy none of the properties of the set $\mathbb{P}_n$), is obtained from the sieve formula based on the principle of inclusion and exclusion of I(r) which is the number of stereoisomers that have not been included and excluded up to r-1 terms in equations (21) or (22).

$$
I_n = I(0) + \sum_{r=1}^{(n-2)/2} (-1)^r I(r) \quad \text{if n even} \tag{21}
$$

or

$$
I_n = I(0) + \sum_{r=1}^{(n-1)/2} (-1)^r I(r) \quad \text{if n odd} \tag{22}
$$

APPLICATIONS

The results of the different steps of the enumeration procedure presented in this paper are given in Table I for deoxyheptitol (n=7).The procedure has been extended to all deoxyalditols $C_nH_{n+3}(OH)_{n-1}$ where $3 \leq n \leq 10$ and for illustration the calculated numbers I_n of stable stereo and position isomers of these chemical

Table I : Results of the enumeration procedure applied to the series of deoxyheptitols ($C_7H_{10}(OH)_6$).

r	w_r	r - tuples	g_r	$x^{n+3}y^{n-1-2r}$	C(n+3,n-1-2r)	e_q	I(r)
0	1	-	$S_2[S_3 . E_5]$	$x^{10}y^6$	1003	-	1003
1	4	P_1	$S_3 . E_{11}$	$x^{10}y^4$	561		1319
		P_2	$S_3^2 . E_8$		301		
		P_3	$S_3^2 . E_8$		301		
		P_4	$S_2[S_3 . E_5]$		156		
2	12	$P_1 P_2$	$S_3 . E_9$	$x^{10}y^2$	46		406
		$P_1 P_3$	$S_3 . E_9$		46		
		$P_1 P_4$	$S_3 . E_9$		46		
		$P_1 P_5$	$S_3 . E_9$		46		
		$P_1 P_6$	$S_3 . E_9$		46		
		$P_1 P_7$*	$S_2[E_6]$		36	6	
		$P_2 P_6$*	$S_2[S_3 . E_3]$		17	4	
		$P_3 P_5$*	$S_2[S_3 . E_3]$		17	4	
		$P_2 P_3$	$S_3^2 . E_6$		30		
		$P_2 P_4$	$S_3^2 . E_6$		30		
		$P_2 P_5$	$S_3^2 . E_6$		30		
		$P_3 P_4$	$S_3^2 . E_6$		30		
3	19	$P_1 P_2 P_3$	$S_3 . E_7$	x^{10}	1		16
		$P_1 P_2 P_4$	$S_3 . E_7$		1		
		$P_1 P_2 P_5$	$S_3 . E_7$		1		
		$P_1 P_2 P_6$	$S_3 . E_7$		1		
		$P_1 P_2 P_7$	E_{10}		1		
		$P_1 P_3 P_4$	$S_3 . E_7$		1		
		$P_1 P_3 P_5$	$S_3 . E_7$		1		
		$P_1 P_3 P_6$	$S_3 . E_7$		1		
		$P_1 P_3 P_7$	E_{10}		1		
		$P_1 P_4 P_5$	$S_3 . E_7$		1		
		$P_1 P_4 P_6$	$S_3 . E_7$		1		
		$P_1 P_4 P_7$*	$S_2[E_5]$		1	1	
		$P_2 P_3 P_4$	$S_3^2 . E_4$		1		
		$P_2 P_3 P_5$	$S_3^2 . E_4$		1		
		$P_2 P_3 P_6$	$S_3^2 . E_4$		1		
		$P_2 P_3 P_7$	$S_3 . E_7$		1		
		$P_2 P_4 P_5$	$S_3^2 . E_4$		1		
		$P_2 P_4 P_6$*	$S_2[S_3 . E_2]$		1	1	
		$P_3 P_4 P_5$*	$S_2[S_3 . E_2]$		1	1	

* r-tuples containing a pair of properties (P_i,P_j) located in equivalent carbon atoms i and j.

compounds are indicated in Table II. One may notice that the summation of all I(r) values (columns 2-6) leads to the numbers I_n of the last column. For the sake of comparison the number $I_5 = 15$ obtained for deoxypentitols in this study is the same result presented earlier by Balasubramanian[6]. It should be mentioned that all $I_n/2$ constitutional isomers for even n are chiral, whereas for odd n some of the constitutional isomers with median CH_2 are meso forms.

Table II : Numbers I_n of stable stereo and position isomers of deoxyalditols $C_nH_{n+3}(OH)_{n-1}$ where $3 \leq n \leq 10$.

n	I(0)	-I(1)	I(2)	-I(3)	I(4)	I_n
3	5	-2				3
4	16	-10				6
5	67	-56	4			15
6	251	-274	55			32
7	1003	-1319	406	-16		74
8	3900	-6098	2624	-266		160
9	15402	-27685	15108	-2529	60	356
10	60418	-123336	82124	-19656	1218	768

CONCLUSION

The focus of this paper has been to develop a general method for counting stable stereo and position isomers of acyclic deoxyalditols $C_nH_{n+3}(OH)_{n-1}$. This procedure can be extended to multideoxyalditols.

ACKNOWLEDGEMENTS

The authors thank University of Yaounde I for providing the research grant N° FS/87002 and Prof. G. LEROY of the Laboratory of Quantum Chemistry, UCL, Belgium, for his permanent support.

REFERENCES

1. LINDBERG B. and MEIR H., *Acta Chem. Scand.*; (1962), 16, 543.

2. BOLLENBACK and UNDERKOFLER, *J. Am. Chem. Soc.*, (1950), 72, 241.

3. TIPSON R. S., *Adv. Carbohydrate Chem.*, (1953), 8, 164.

4. De BELDER A. N. and WEIGEL H., *Chem. and Ind.*, (1964), 1689.

5. FOUQUEY C., POLONSKY J. and LEDERER E., *Bull. Soc. Chim. Fr.*, (1959), 803.

6. BALASUBRAMANIAN K., *Ann. New York Acad. of science*, (1979),56, pp. 33-36.

7 POLYA G. , READ R. C.., *Combinatorial Enumeration of Graphs, Groups and Chemical Compounds*, Springer Verlag; New York, (1987), pp. 58-74.

8. BALABAN A. T. , KENNEDY J. W. and QUINTAS L. V., *J. Chem. Educ.*, (1988), 65, 304.

9. BERGE C., *In Principles of Combinatorics* , Academic Press, London, (1971), pp. 149-173.

10. READ R. C., *in Chemical Applications of Graph Theory* , BALABAN A.T., Academic Press, London, (1976), pp. 25-60.

11 ROBINSON R. W., *J. Comb. Theory*, (1970), 9, 327.

(*Received in UK* 21 *November* 1994; *revised* 30 *January* 1995; *accepted* 3 *February* 1995)