Application of the Sieve Method to the Enumeration of Stable Stereo and Position Isomers of a Series of Deoxycyclitols

Robert Martin Nemba* and Maurice Fah

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

Received August 16, 1996[⊗]

A pattern inventory is developed for the enumeration of stable stereo and position isomers of any series of deoxycyclitols with the formula $(CH_2)_p(CHOH)_q$ where $p(CH_2)$ and $q(CHOH)$ groups form a ring system of size $n = (p + q)$. The counting method consists of two steps: firstly the application of POLYA's counting theorem gives the number of permutation isomers of the series; and secondly, the sequential enumeration of stable stereo and position isomers is carried out by applying Balasubramanian's procedure to each permutation isomer of the series. An illustration is shown in the series of deoxycycloheptitols where $p + q = 7$, $1 \le p \le 5$ and $2 \le q \le 6$.

INTRODUCTION

This paper is a continuation of a previous work where we have presented a mathematical method for counting stable stereo and position isomers of deoxyalditols.[1] The emphasis here is the enumeration of stable stereoisomers of deoxycyclitols, and the term "stable stereoisomers" refers to chemical structures which have no two −OH groups at one site. The deoxycyclitols symbolized with the formula $(CH_2)_p(CHOH)_q$ are polyhydroxy-substituted cycloalkane rings with the characteristics typical of sugar molecules.[2] Any molecule of these series presents permutation isomers due to the possible displacement of the position of the CH_2 groups along the cyclic chain, and each permutation isomer is convertible to a multitude of stereoisomers due to the alternant positions of the hydroxy groups located on asymmetric carbon atoms. The occurrence of some deoxycyclitols in nature has been reported in the literature.[3] For instance the (*s*,*s*,*s*,*s*)-cyclohexane-1,2,4,5-tetrol has been isolated from sugar beet molasses.[4] Deoxycyclitols are conveniently synthesized by nonselective dihydroxylation of the double bonds in medium and large ring olefins. For memory we recall the recent syntheses of cyclohexane-1,2,4,5-tetrol by Suemene et al. and 1,2,5,6,9,10-hexahydroxycyclododecanes by Becker et al.[6] But the nonselective dihydroxylation of all double bonds in medium and large ring olefins should lead to a mixture of diastereoisomers and give rise to the problem of enumeration of distinct structures.

Let the system $(CH_2)_p(CHOH)_q = (n,p,q)$. The three nonnegative integers *n*, *p*, and *q*, which denote respectively the ring size and the numbers of CH_2 and CHOH groups, obey the following restrictions: $n = p + q$ and $3 \le n \le \infty$, while $1 \le p \le n - 2$ and $2 \le q \le n - 1$. To determine with exactness the number $I_n(p,q)$ of stable stereo and position isomers of a series of deoxycyclitols $(CH_2)_p(CHOH)_q$ with a ring size $n = (p + q)$, we have developed a pattern inventory which consists of two steps. In the first step, one must carry out the enumeration of permutation isomers to those structures that have the same cyclic skeleton and the same numbers *p* and *q* of CH_2 and CHOH groups, respectively, but differ only by the skeletal distribution of those groups. In the second step, the computation of $I_n(p,q)$ is carried out through Balasubramanian's method[7] which uses the sieve formula based on the principle of inclusion and exclusion of nonvalid structures of polyols i.e., cyclic polyols having two −OH groups at one site or more pairs of −OH groups located at distinct sites. An illustration is shown in the series of deoxycycloheptitols $C_7H_{2p+q}(OH)_q$ where $n = p + q = 7$, $1 \le p \le 5$ and $2 \le q \le 6$.

MATHEMATICAL FORMULATION

Let us represent the tridimensional graph or stereograph of a deoxycyclitol as 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 $2p + q$ hydrogen atoms −H and *q* hydroxy groups −OH are to be put. The problem is to find the number of ways we can map the unspecified vertices into the set consisting of $2p + q$ hydrogen atoms and *q* hydroxy groups. Thus we have an enumeration problem of Polya type.[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 Figure 1 into 3 subgraphs Q, T_1, and T_2 represented in Figure 2. If we place H and OH into two boxes, we see that no permutation of 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 $Z(E_2) = s_1^2$. The permutation group acting on T_1 is the identity group E whose cycle index is $Z(E) = s_1$. The molecular graph shown in Figure 1 can be constructed by attaching the roots of $(p + q)$ subgraphs of type T_2 and $p + q$ vertices of Q, and one subgraph T_2 can be obtained by attaching the roots of two subgraphs T_1. The vertices in Q are permutable by the group of permutation D_n whose cycle index is given hereafter:

$$Z(D_n) = \frac{1}{2n}\left[\sum_{d|n}\varphi(d)s_d^{n/d} + ns_1s_2^{(n-1)/2}\right] \quad (n \text{ odd}) \qquad (1)$$

$$Z(D_n) = \frac{1}{2n}\left[\sum_{d|n}\varphi(d)s_d^{n/d} + \frac{n}{2}s_2^{n/2} + \frac{n}{2}s_1^2s_2^{(n-2)/2}\right] \quad (n \text{ even}) \qquad (2)$$

[⊗] Abstract published in *Advance ACS Abstracts*, May 1, 1997.

S0095-2338(96)00129-1 CCC: $14.00

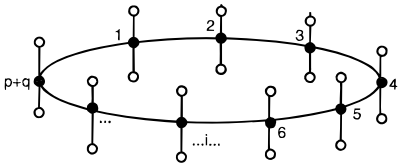

Figure 1. Molecular graph of a deoxycyclitol with a ring size $n = p + q$.

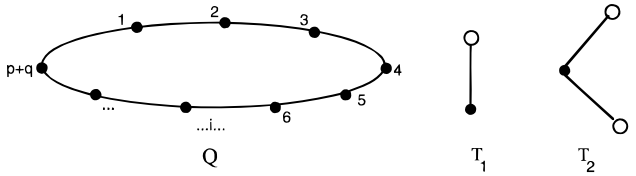

Figure 2. Subgraphs Q, T_1, and T_2 derived from the partition of the parent graph shown in Figure 1.

In these expressions the integer n is the ring size as previously defined; the terms of type s_k^j correspond to j permutation cycles of length k, and the coefficient $\varphi(d)$ is the Euler totient function[9] which is the number of integers from 1 to d that have no factors in common with d and the summation is over all integers d that are factors of n.

Let $\mathbf{M}_{p+q} = \{M_1, M_2, ..., M_i, ..., M_{p+q}\}$ be the set of positions that may be taken in the cyclic chain by CH_2 groups and where M_i indicates that the methylene group is located on the ith carbon atom. The number of nonequivalent ways of placing exactly $p CH_2$ groups among $p + q$ distinct sites in the cyclic chain is given by the transformation hereafter

$$Z(D_n,s_k^j) \rightarrow Z[D_n,(1 + x^k)^j] = \sum_{i=0}^{n} \omega_i x^i \qquad (3)$$

in which if $i = p$, the coefficient ω_p of the term $\omega_p x^p$ is the number of nonequivalent ways of placing exactly $p CH_2$ groups among $p + q$ distinct sites in the cyclic chain. Hence each molecular system $(CH_2)_p(CHOH)_q$ will have ω_p permutation isomers which are symbolized by the p-tuples of properties M_i chosen from the set $\mathbf{M}_{p+q}$. At the end of this first step one can generate the complete list and chemical graphs of the ω_p permutation isomers. Let $G_r(j)$ be the permutation group acting on the whole skeleton of the permutation isomer numbered j and $1 \le j \le \omega_p$. If the molecular stereograph is permutable end for end by the permutation group S_2, hence

$$G_r(j) = S_2[E_{2\alpha+\beta}] \qquad (4)$$

In this equation the parameters α and β are the numbers of subgraphs of types T_2 and T_1 contained in one-half of the molecular stereograph. When this latter has no symmetry, $G_r(j)$ is obtained from the relation

$$G_r(j) = [E_{2\alpha+\beta}] \qquad (5)$$

and this time α and β are the numbers of subgraphs of types T_2 and T_1 contained in the whole molecular stereograph. Let $\mathbf{P}_{p+q} = \{P_1, P_2, ..., P_i, ..., P_m\}$ be the set of properties where P_i is the property that there are two hydroxy groups on the ith carbon atom and where $m \le p + q$. The number of nonequivalent ways of choosing exactly r properties from the set $\mathbf{P}_{p+q}$ is given by $a_r(j)$ the coefficient of x^r in the expression

$$Z[G_r(j),s_k^l] \rightarrow Z[G_r(j),(1 + x^k)^l] = \sum_r a_r(j) x^r \qquad (6)$$

derived from Polya's theorem. For each permutation isomer of the series $(CH_2)_p(CHOH)_q$ the restrictions on $a_r(j)$ are as follows: $a_r(j) \neq 0$ if $0 \le r \le q/2$ (q even) or $0 \le r \le (q-1)/2$ (q odd), and $a_r(j) = 0$ otherwise. Let NID(p,q) be the total number of "invalid deoxycyclitols" i.e., cyclic polyols having two $-$OH groups at one site or more pairs of $-$OH groups located at distinct sites, for the series $(CH_2)_p(CHOH)_q$. The parameter NID(p,q) is obtained from the following relations

$$\mathrm{NID}(p,q) = \sum_{j=1}^{\omega_p} \sum_{r=1}^{\frac{q-1}{2}} a_r(j)$$

if q is odd or

$$\mathrm{NID}(p,q) = \sum_{j=1}^{\omega_p} \sum_{r=1}^{\frac{q}{2}} a_r(j)$$

if q is even. The cycle indices deduced from the appropriate permutation groups $G_r(j)$ defined in eqs 4 and 5 are as follows:

$$Z(S_2[E_{2\alpha+\beta}]) = \frac{1}{2}[s_1^{4\alpha+2\beta} + s_2^{2\alpha+\beta}] \qquad (7)$$

and

$$Z[E_{2\alpha+\beta}] = s_1^{2\alpha+\beta} \qquad (8)$$

Let us associate to the hydrogen atom H and the functional group OH the weights α_1 and α_2, respectively. By replacing in eqs 7 and 8 the terms s_k^l by the figure counting series $(\alpha^k + \alpha^k)^l$ and expanding the resulting algebraic expression one may obtain the polynomial function

$$f(\alpha_1,\alpha_2) = \sum_u \sum_v C(u,v) \alpha_1^u \alpha_2^v \qquad (9)$$

If $u = q$ and $v = q - 2r$, the coefficient $C(q,q-2r)$ of the term $(\alpha_1^q \alpha_2^{q-2r})$ is the number of stereoisomers of the deoxycyclitol satisfying exactly r properties of the set P_{p+q}. Let $I_j(r)$ be the total number of stereoisomers of all polyalkohols obtained according the $a_r(j)$ ways of satisfying r properties of the set P_{p+q} applied to the permutation isomer j.

$$I_j(r) = \sum_1^{a_r(j)} C(q,q-2r) - \sum_t e_t(j) \qquad (10)$$

The parameter $e_t(j) \neq 0$ is the number of stereoisomers issued from the t choices having pairs of properties (P_i,P_k) located at equivalent sites i and k. The carbon atoms i and k will be equivalent if they are transformable into each other under the action of the permutation group S_2. In eq 10 the value $e_t(j) \neq 0$ will be applied only for *meso*-isomers and $e_t(j) = 0$ otherwise. The number $I_j(p,q)$ of stable stereoisomers of

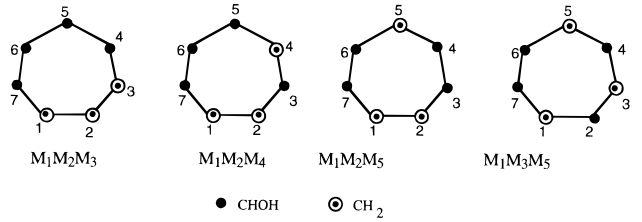

Figure 3. Planar graphs of permutation isomers of cycloheptane tetrols $(CH_2)_3(CHOH)_4$.

the permutation isomer j that have no two hydroxy groups on the same carbon atom is obtained from the sieve formula based on the principle of inclusion and exclusion of $I_j(r)$ which is the number of stereoisomers of the permutation isomer j that have not been included or excluded up to $r-1$ terms in eqs 11 and 12 hereafter

$$I_j(p,q) = I_j(0) + \sum_{r=1}^{\frac{q-1}{2}} (-1)^r I_j(r) \quad (q \text{ odd}) \qquad (11)$$

or

$$I_j(p,q) = I_j(0) + \sum_{r=1}^{\frac{q}{2}} (-1)^r I_j(r) \quad (q \text{ even}) \qquad (12)$$

Finally the number $I_n(p,q)$ of stable stereoisomers of ω_p permutation isomers generated from the cyclic system $(CH_2)_p(CHOH)_q$ is obtained by summing up the numbers $I_j(p,q)$. Hence

$$I_n(p,q) = \sum_{j=1}^{\omega_p} [I_j(0) + \sum_{r=1}^{\frac{q-1}{2}} (-1)^r I_j(r)] \quad (q \text{ odd}) \qquad (13)$$

or

$$I_n(p,q) = \sum_{j=1}^{\omega_p} [I_j(0) + \sum_{r=1}^{\frac{q}{2}} (-1)^r I_j(r)] \quad (q \text{ even}) \qquad (14)$$

APPLICATIONS

Our illustration is based on the case of deoxycycloheptane polyols where $p+q=7$, $1 \le p \le 5$, and $2 \le q \le 6$. If we consider the series of cycloheptane tetrols $C_7H_{10}(OH)_4$ by applying the previous methodology one may find the cycle index $Z(D_7)$ and its associated generating function as follows:

$$f(\omega_p) = 1 + x + 3x^2 + 4x^3 + 4x^4 + 3x^5 + x^6 + x^7$$

The coefficient ω_p of x^m in the counting series $f(\omega_p)$ indicates the number of permutation isomers for each pair of numbers (p,q) which satisfy the relation $p+q=7$ in the series $(CH_2)_p(CHOH)_q$. From this correspondence one can easily deduce that for $p=3$ and $q=4$ the molecular system $(CH_2)_3(CHOH)_4$ leads to four permutation isomers designated by the 3-tuples $M_1M_2M_3$, $M_1M_2M_4$, $M_1M_2M_5$, and $M_1M_3M_5$, represented by the planar graphs in Figure 3 hereafter.

After this first step one must carry out the task of sequential enumeration of stable stereo and position isomers of the series by applying Balasubramanian's procedure to each of these four permutation isomers. The results presented in Table 1 show that the four permutation isomers $M_1M_2M_3$, $M_1M_2M_4$, $M_1M_2M_5$, and $M_1M_3M_5$ give rise to the sequence of numbers $I_j(3,4) = 10$, 16, 10, and 10 stereoisomers, respectively assuming $j = 1$, 2, 3, and 4. The total number of stereoisomers of the series of cycloheptane tetrols is $I_7(3,4) = 46$. The different values of the numbers $I_n(p,q)$ computed in the series of deoxycycloheptane polyols are indicated in Table 2.

Table 1. Results of the Enumeration Procedure Applied to the Series of Cycloheptane Tetrols $(CH_2)_3(CHOH)_4$

PI	r	r-tuples	$G_r(j)$	$\alpha_1^q\alpha_2^{q-2r}$	$C(q,q-2r)$	$e_t(j)$	$I_j(r)$	$I_j(p,q)$
$M_1M_2M_3$	0		$S_2[E_4]$	$\alpha_1^4\alpha_2^4$	38		38	
	1	P_4	E_6	$\alpha_1^4\alpha_2^2$	15			
		P_5	E_6		15		30	
	2	P_4P_5	E_4	α_1^4	1			
		P_4P_6	E_4		1			
		P_4P_7[a]	$S_2[E_2]$	α_1^4	1	1		
		P_5P_6[a]	$S_2[E_2]$		1	1	2	10
$M_1M_2M_4$	0		E_8	$\alpha_1^4\alpha_2^4$	70		70	
	1	P_3	E_6	$\alpha_1^4\alpha_2^2$	15			
		P_5	E_6		15			
		P_6	E_6		15			
		P_7	E_6		15		60	
	2	P_3P_5	E_4	α_1^4	1			
		P_3P_6	E_4		1			
		P_3P_7	E_4		1			
		P_5P_6	E_4		1			
		P_5P_7	E_4		1			
		P_6P_7	E_4		1		6	16
$M_1M_2M_5$	0		$S_2[E_4]$	$\alpha_1^4\alpha_2^4$	38		38	
	1	P_3	E_6	$\alpha_1^4\alpha_2^2$	15			
		P_4	E_6		15		30	
	2	P_3P_4	E_4	α_1^4	1			
		P_3P_6	E_4		1			
		P_3P_7[a]	$S_2[E_2]$	α_1^4	1	1		
		P_4P_6[a]	$S_2[E_2]$		1	1	2	10
$M_1M_3M_5$	0		$S_2[E_4]$	$\alpha_1^4\alpha_2^4$	38		38	
	1	P_2	E_6	$\alpha_1^4\alpha_2^2$	15			
		P_4	E_6		15		30	
	2	P_4P_6	E_4	α_1^4	1			
		P_4P_7	E_4		1			
		P_2P_4[a]	$S_2[E_2]$	α_1^4	1	1		
		P_6P_7[a]	$S_2[E_2]$		1	1	2	10

[a] r-Tuples containing a pair of properties (P_i,P_k) located on equivalent carbon atoms i and k. PI = permutation isomers; $I_n(p,q) = 46$ for the series $(CH_2)_3(CHOH)_4$.

Table 2. Numbers of Stable Stereo and Position Isomers of the Series of Deoxycycloheptanes Polyols

deoxycycloheptane polyols	permutation isomers	$I_j(0)$	$I_j(1)$	$I_j(2)$	$I_j(3)$	$I_j(p,q)$	$I_n(p,q)$
$(CH_2)_5(CHOH)_2$	$M_1M_2M_3M_4M_5$	4	−1			3	
	$M_1M_2M_3M_4M_6$	4	−1			3	9
	$M_1M_2M_3M_5M_6$	4	−1			3	
$(CH_2)_4(CHOH)_3$	$M_1M_2M_3M_6$	20	−12			8	
	$M_1M_2M_3M_4$	10	−6			4	
	$M_1M_2M_4M_6$	10	−6			4	20
	$M_1M_4M_5M_7$	10	−6			4	
$(CH_2)_3(CHOH)_4$	$M_1M_2M_3$[a]	38	−30	2		10	
	$M_1M_2M_4$	70	−60	6		16	
	$M_1M_2M_5$[a]	38	−30	2		10	46
	$M_1M_3M_5$[a]	38	−30	2		10	
$(CH_2)_2(CHOH)_5$	M_1M_2	126	−140	28		14	
	M_1M_3	126	−140	28		14	42
	M_1M_4	126	−140	28		14	
$(CH_2)(CHOH)_6$	M_1	472	−630	210	−10	42	42

[a] Permutation isomers exhibiting some meso-forms which have $e_t(j) = 1$ as indicated in Table 1.

CONCLUSION

The emphasis of this paper has been to develop a general method for counting stable stereoisomers of deoxycyclitols. This method is a double enumeration process which comprises for each series of deoxycyclitols: firstly the enumeration of permutation isomers and secondly the sequential enumeration of stable stereoisomers using the sieve formula based on the principle of inclusion and exclusion of invalid structures of polyols.

ACKNOWLEDGMENT

The author gratefully thanks Professor Georges Leroy of the Laboratory of Quantum Chemistry, Université Catholique de Louvain (Belgium) for his permanent support within this work.

REFERENCES AND NOTES

(1) Nemba, R. M.; Fah, M. On the application of sieve formula to the enumeration of stable stereo and position isomers of deoxyalditols. *Tetrahedron* **1995**, *51*, No. 13, 3831−3840.

(2) (a) Codd, R.; Hambley, T. W.; Lay, P. Synthesis, Characterization and Crystal Structure of a Vanadium (V)−Vanadium (IV)−Vanadium (V) mixed valence trinuclear complex with a tridentate quinato ligand, $\{(NH_4)_2[V^V(O)_2]_2[V^{IV}(O)](\mu\text{-}(-)\text{-quinato}(3-))_2\}\text{-}H_2O$. *Inorg. Chem.* **1995**, *34*, 877−882. (b) Chapelle, S.; Sauvage, J. P.; Koll, P; Verchere, J. F. W-NMR studies of Tungstate Complexes of carbohydrates. 3-Species formed with all-threo alditols acting as tridentate, tetradentate or pentadentate ligands. *Inorg. Chem.* **1995**, *34*, 918−923.

(3) Anderson, L. In *The Carbohydrates*; Academic Press: London, 1972; Vol. 1A, p 521.

(4) Ramanathan, J. D.; Craigie, J. S.; McLehlan, J.; Smith, D. G.; McInnes, A. G. *Tetrahedron Lett.* **1966**, 14**,** 1527−1535.

(5) Suemene, H.; Hasegawa, A; Sakai, K. Highly enantio- and diastereoselective Synthesis of C_2-symmetric 3,5-Cyclohexadiene-1,2-diol and D_2-symmetric Cyclohexane-1,2,4,5-tetrol. *Tetrahedron Asym.* **1995**, *6*, No. 1, 55−58.

(6) Becker, H.; Soler, M. A.; Sharpless, K. B. Selective asymmetric dihydoxylation of Polyenes. *Tetrahedron* **1995**, *55*, No. 5, 1345−1376.

(7) Balasubramanian, K. Enumeration of stable stereo and position isomers of polyalkohols. *Ann. New York Acad. Sci.* **1979**, *319*, 33−36.

(8) Polya, G.; Read, R. C. *Combinatorial Enumeration of Graphs, groups and chemical Compounds*; Springer-Verlag: New York, 1987; pp 58−74.

(9) Polya, G.; Tarjan, R. E.; Woods, D. R. Notes on Introductory Combinatorics; Birkhäuser: Boston, 1983; pp 55−85.

CI960129L