

MATHEMATICAL MODELING OF POLYMERS. PART II.<sup>1</sup>  
IRREDUCIBLE SEQUENCES IN  $n$ -ARY COPOLYMERS

Cristian Artemi <sup>a</sup> and Alexandru T. Balaban <sup>b,\*</sup>

<sup>a</sup> Faculty of Mathematics, Bucharest University

<sup>b</sup> Polytechnic Institute, Department of Organic Chemistry,  
Splaiul Independentei 313, 76206 Bucharest, Roumania

(Received: February 1987)

*Abstract.* The number of irreducible sequences in  $n$ -ary copolymers ( $n = 2, 3, 4$ ) is shown to be related to the graph-theoretical necklace problem, with supplementary conditions involving presence of all  $n$  monomer types at least once, invariance to relabelling of vertices, and impossibility of decomposition into smaller repeating subsequences. It is shown that by combining the Pólya-de Bruijn theorem with the Möbius function of the lattice of divisors one may solve this problem. A general formula is also presented for the number of irreducible sequences in  $n$ -ary copolymers. A computer program was devised for generating and enumerating all irreducible sequences involving  $n = M$  types of comonomers wherein the repeating irreducible sequence contains  $m = N$  monomer units.

## Introduction

In the previous part of this series,<sup>1</sup> we defined and discussed irreducible sequences in binary copolymers, and showed that the same sequences would be found in stereoregular homopolymers or polybutenamers ; these irreducible sequences were formed from a binary alphabet. The present paper generalizes the problem to irreducible sequences corresponding to n-ary copolymers, wherein the alphabet consists of n letters. In chemical practice,<sup>2</sup> higher alphabets than n = 5 are not interesting.

## The necklace problem and irreducible sequences

In Part I of the present series<sup>1</sup> it was shown that the enumeration of irreducible sequences in binary copolymers is related to the enumeration of diastereomeric cycloalkanes bearing one type of substituent at each carbon atom (ignoring enantiomerism), and to the "necklace problem" with beads of two colors.

A hierarchical order of the colors is accepted (corresponding to priority rules for letters of the alphabet, namely  $R < S < T < U$ ), giving rise to a lexicographically ordered list of all irreducible sequences. The enumeration of irreducible sequences in n-ary copolymers as discussed above is equivalent to enumerating the distinct necklaces formed from beads of n colors, provided that three supplementary conditions are added to the well-known<sup>3</sup> necklace problem :

1) Two or more non-isomorphic necklaces correspond to one and the same irreducible sequence if, on relabeling the colors, one obtains the same necklace (i.e. if on permuting the letters of the given alphabet one obtains the same sequence). The necessary and sufficient condition for the existence of such distinct necklaces which give rise to the same irreducible sequence is that they contain at least two types of differently colored beads in equal numbers, with the exception of necklaces containing three types of beads with two single beads of different colors.

For instance the partition  $R^2S^2T$  of  $m = 5$  beads into  $n = 3$  colors gives rise according to Pólya's theorem to four necklaces which according to the lexicographic order can be linearised as RRSST, RSRST, RSRST and RSSRT ; the fourth necklace is converted, however, into the third one by permutation (RS)(T) so that only the first three necklaces correspond to irreducible

sequences. As an exception, all necklaces with  $n$  beads of 3 colors partitioned into 1,1, and  $n-2$  beads of the same color correspond to irreducible sequences.

The special position for this partition  $R^{n-2}ST$  is connected to the facts that the two faces of such necklaces correspond to a permutation of the  $S$  and  $T$  symbols, and that the macromolecular chain of repeating sequences has no privileged direction from one end to the other.

2) No irreducible sequence should consist of smaller repeating subsequences. Whenever  $m$  is not a prime number, such repeating subsequences may be possible. Thus, there are two necklaces with  $m =$  four beads of  $n =$  two colors, but only one irreducible sequence, namely  $RRSS$ , because the sequence  $RSRS$  can be decomposed into two repeating subsequences. Similarly,  $RSRSRS$ ,  $RRSRRS$  and  $RSTRST$  are not irreducible sequences with  $m = 6$  because they can be decomposed into smaller, repeated subsequences.

3) Any sequence of  $m$  symbols which does not contain all  $n$  symbols of the alphabet is reducible to a sequence corresponding to a lower alphabet.

When all these conditions for non-equivalence are fulfilled, we call such chemically non-equivalent sequences : "irreducible sequences".

The necklace problem can be solved by means of Polya's theorem.<sup>4</sup> A diagram with the number of necklaces with up to three colors and eight beads may be found in Table 13 of reference<sup>5</sup>. For the general case, the number of necklaces with  $r_1$  beads of color 1,  $r_2$  beads of color 2, ...,  $r_n$  beads of color  $n$  ( $\sum_i r_i = m$ ), is given by coefficient of the term  $x_1^{r_1} x_2^{r_2} \dots x_n^{r_n}$  in the polynomial obtained, according to Polya's theorem, on substituting the figure-counting series

$$y_k = \sum_{i=1}^n x_i^k$$

into the cycle index of the dihedral group:

$$Z(D_m) = \frac{1}{2}Z(C_m) + \frac{1}{2}y_1 y_2^{(m-1)/2} \quad \text{for } m \text{ odd}$$

$$Z(D_m) = \frac{1}{2}Z(C_m) + \frac{1}{4}(y^{m/2} + y_1^2 y_2^{(m-2)/2}) \quad \text{for } m \text{ even}$$

where the cycle index of the cyclic group is :

$$Z(C_m) = \frac{1}{m} \sum_{k|m} \phi(k) y_k^{m/k}$$

The symmetry operation  $y_k$  permutes  $k$  points of the  $m$ -gon ;  $\phi(k)$  denotes the Euler  $\phi$ -function (i.e. the number of positive integers less than  $k$  and relatively prime to  $k$ ) ;  $k|m$  indicates that  $k$  divides into  $m$ .

Results are presented in Table 1.

TABLE 1. Cycle indices of dihedral groups.

$$Z(D_4) = (y_1^4 + 2y_1^2y_2 + 3y_2^2 + 2y_4)/8$$

$$Z(D_5) = (y_1^5 + 5y_1y_2^2 + 4y_5)/10$$

$$Z(D_6) = (y_1^6 + 4y_2^3 + 2y_3^2 + 3y_1^2y_2^2 + 2y_6)/12$$

$$Z(D_7) = (y_1^7 + 7y_1y_2^3 + 6y_7)/14$$

$$Z(D_8) = (y_1^8 + 2y_4^2 + 4y_8 + 5y_2^4 + 4y_1^2y_2^3)/16$$

$$Z(D_9) = (y_1^9 + 2y_3^3 + 9y_1y_2^4 + 6y_9)/18$$

A list of configuration-counting series (polynomials in  $y_k$  for various values of  $m$ ) is to be found in references<sup>6,7</sup>. The coefficient of each term  $x_1^r x_2^s \dots x_n^v$  (where  $m = \sum_i r_i = r + s + \dots + v$  is the number of beads, some of the terms  $s, \dots, v$  may be zero, and  $n$  is the number of colors,  $n \leq m$ ), is the number of distinct necklaces.

In Table 2 one can see the numbers NK of necklaces (left-hand figures for each partition of  $m$ ), according to Polya's theorem. The numbers IS of irreducible sequences for each partition, obtained according to the computer program (last section of this paper) and the total number  $N(m, n)$  of irreducible sequences obtained either according to the Polya -de Bruijn theorem (following section of this paper) or to a generalized formula (presented in a subsequent section) are also displayed in Table 2 for  $3 \leq m \leq 8$  and  $2 \leq n \leq 5$ . Results for binary copolymers ( $n = 2$ ) agree with those described in Part I of the present series.<sup>1</sup> One should note that, in agreement with condition 1 indicated above, partition  $R^2S^2TU$  may stand for  $R^2ST^2U$ , or  $R^3S^2T^2$  for  $R^2S^3T^2$ , etc. All irreducible sequences corresponding to the same partition are isomeric with one another. A list of irreducible sequences with  $n = 2, 3$ , and 4, and  $m = 2$  through 7, arranged according to partitions, is to be found in Part III of the present series.<sup>8</sup>

TABLE 2. Numbers NK of necklaces (left-hand figure) and irreducible sequences IS (right-hand figure) for each partition  $R^r S^s \dots V^v$  ( $r + s + \dots + v = m$ ), and total number  $N(m, n)$  of irreducible sequences for given values of the number  $n$  of monomer types and the length  $m$  of the sequence.

$\frac{n}{m}$	2		3		4		5	
	NK	Partition IS $N(m, n)$	NK	Partition IS $N(m, n)$	NK	Partition IS $N(m, n)$	NK	Partition IS $N(m, n)$
2	1	RS 1	-	-	-	-	-	-
3	1	R <sup>2</sup> S 1	1	RST 1	-	-	-	-
4	1	R <sup>3</sup> S 1	2	R <sup>2</sup> ST 2	3	RSTU 1	-	-
5	1	R <sup>4</sup> S 1	3	R <sup>3</sup> ST 2	6	R <sup>2</sup> STU 2	12	RSTUV 1
6	1	R <sup>5</sup> S 1	5	R <sup>4</sup> ST 3	13	R <sup>3</sup> STU 3	30	R <sup>2</sup> STUV 3
7	1	R <sup>6</sup> S 1	8	R <sup>5</sup> ST 3	31	R <sup>4</sup> STU 4	60	R <sup>3</sup> STUV 4
8	1	R <sup>7</sup> S 1	14	R <sup>6</sup> ST 4	80	R <sup>5</sup> STU 5	105	R <sup>4</sup> STUV 8
2	2	R <sup>3</sup> S <sup>2</sup> 2	4	R <sup>2</sup> S <sup>2</sup> T 3	5	R <sup>2</sup> S <sup>2</sup> T <sup>2</sup> U 2	210	R <sup>3</sup> S <sup>2</sup> TUV 38
3	3	R <sup>4</sup> S <sup>2</sup> 2	6	R <sup>3</sup> S <sup>2</sup> T 6	13	R <sup>2</sup> S <sup>2</sup> TU 8	1155	R <sup>2</sup> S <sup>2</sup> T <sup>2</sup> UV 39
4	3	R <sup>5</sup> S <sup>2</sup> 2	11	R <sup>4</sup> S <sup>2</sup> T <sup>2</sup> 4	44	R <sup>3</sup> S <sup>2</sup> T <sup>2</sup> U 12	171	R <sup>2</sup> S <sup>2</sup> T <sup>2</sup> U <sup>2</sup> 16
5	3	R <sup>6</sup> S <sup>2</sup> 3	9	R <sup>5</sup> ST 9	31	R <sup>4</sup> STU 9	60	R <sup>3</sup> STUV 12
6	3	R <sup>7</sup> S <sup>2</sup> 4	10	R <sup>6</sup> S <sup>2</sup> T 7	44	R <sup>5</sup> STU 17	90	R <sup>4</sup> STUV 12
7	4	R <sup>8</sup> S <sup>2</sup> 4	18	R <sup>7</sup> S <sup>2</sup> T <sup>2</sup> 12	60	R <sup>6</sup> S <sup>2</sup> T <sup>2</sup> U 12	105	R <sup>5</sup> STUV 8
8	4	R <sup>9</sup> S <sup>2</sup> 5	19	R <sup>8</sup> S <sup>2</sup> T 19	80	R <sup>7</sup> S <sup>2</sup> TU 33	210	R <sup>6</sup> S <sup>2</sup> TUV 38
9	5	R <sup>10</sup> S <sup>2</sup> 5	33	R <sup>9</sup> S <sup>2</sup> T <sup>2</sup> 20	115	R <sup>8</sup> S <sup>2</sup> T <sup>2</sup> U 25	1155	R <sup>7</sup> S <sup>2</sup> T <sup>2</sup> UV 39
10	8	RS 4	38	R <sup>8</sup> S <sup>3</sup> T <sup>2</sup> 25	171	R <sup>7</sup> S <sup>3</sup> T <sup>2</sup> U 16	171	R <sup>6</sup> S <sup>3</sup> T <sup>2</sup> UV 16

A formula for calculating the number  $h(m,n)$  of irreducible sequences based upon the Polya-de Bruijn theorem

Let  $n \geq 2$  and  $m \geq n$  be two natural numbers, and  $S_n$  the symmetric group of order  $n$ . For any two natural numbers  $a$  and  $b$ , we denote by  $[a,b]$  the range of natural numbers in this interval. The set of surjections  $\alpha: [0,a-1] \rightarrow [0,b-1]$  is denoted by  $F(a,b)$ . This set  $F(a,b)$  may be considered as the set of sequences with length  $a$  over an alphabet containing  $b$  symbols such that all these symbols are contained at least once in any sequence. In this context we shall consider a sequence as a function. The cardinal of this set is :

$$s_{a,b} = |F(a,b)| = \sum_{i=0}^{b-1} (-1)^i \binom{b}{i} (b-i)^a$$

For any natural number  $p \geq n$  which divides the number  $m$ , we denote by  $A_p$  the set of functions in  $n^m$  having  $p$  as primitive period ( $A_m$  is the set of non-periodic functions).

Let  $X$  and  $Y$  be two finite sets :  $X = \{a_1, \dots, a_m\}$  and  $Y = \{b_1, \dots, b_n\}$ . The set  $X$  may represent the objects and the set  $Y$  may represent their colors. Let  $G$  be a permutation group on  $X$  and  $H$  a permutation group on  $Y$ . A function  $f: X \rightarrow Y$  is called a coloring of objects in  $X$  with colors in  $Y$ . On the set  $Y^X$  of functions defined on  $X$  and with values in  $Y$  we define the equivalence relationship (denoted by  $\div$ ) as follows :  $f, g \in Y^X$ ,  $f \div g$  iff there exists a permutation  $\pi \in G$  and a permutation  $\sigma \in H$  and such that  $g = \sigma f \pi$ , i.e. for any  $a \in X$ , we have  $g(a) = \sigma(f(\pi(a)))$ . We observe that the equivalence  $\div$  depends on the choice of permutation groups  $G$  and  $H$ .

We denote with  $\Lambda$  the set of classes of  $\div$  equivalence, called the set of coloring schemes ; only the structure of coloring matters, and not the actual colors.

de Bruijn, by generalizing Polya's results, obtained the so-called "Polya-de Bruijn theorem" (formulated here for the particular case of finite sets and of weights equal to one).<sup>9</sup>

The number of coloring schemes (defined as above) is :

$$B_{G,H}(m,n) = |\Lambda| = P_G \left( \frac{\partial}{\partial z_1}, \frac{\partial}{\partial z_2}, \dots, \frac{\partial}{\partial z_m} \right) P_H(\rho_1, \rho_2, \dots, \rho_n)$$

evaluated in  $z_1 = z_2 = \dots = z_m = 0$ , where

$\rho_i = \exp \{i(z_1 + z_2 + z_3 + \dots)\}$ , and  $P_G$  and  $P_H$  are the configuration-count-

ing series of the permutation groups G and H, respectively. For simplifying the notation, whenever G and H are evident from the context, we shall abbreviate  $B_{G,H}(m,n)$  to  $B(m,n)$ .

*Example.* Let  $|X| = 6$  and  $|Y| = 3$ , therefore  $G = D_6$ , the dihedral group of order six, and  $H = S_3$ , the symmetric group of order 3.

The corresponding configuration-counting series are :

$$Z(D_6) = \frac{1}{2} \left[ \frac{1}{6} \sum_{k|6} \phi(k) y_k^{6/k} + \frac{1}{2} (y_2^3 + y_1^2 y_2^2) \right]$$

leading to the expression from Table 1.

From the expression of the  $Z(S_k)$  :

$$Z(S_k) = \sum_{\lambda_1 + 2\lambda_2 + \dots + k\lambda_k = k} \frac{1}{\lambda_1! \lambda_2! \dots \lambda_k! 1^{\lambda_1} 2^{\lambda_2} \dots k^{\lambda_k}} \cdot y_1^{\lambda_1} y_2^{\lambda_2} \dots y_k^{\lambda_k}$$

we obtain for  $k = 3$

$$Z(S_3) = y_1^3/6 + y_1 y_2/2 + y_3/3.$$

On applying the Polya-de Bruijn theorem one obtains :

$$B(6,3) = \frac{1}{12} \left( \frac{\partial^6}{\partial z_1^6} + 4 \frac{\partial^3}{\partial z_2^3} + 3 \frac{\partial^2}{\partial z_1^2} \frac{\partial^2}{\partial z_2^2} + 2 \frac{\partial^2}{\partial z_3^2} + \frac{\partial}{\partial z_6} \right) \cdot$$

$$\cdot \left( \frac{1}{6} e^{3(z_1+z_2+\dots+z_6)} + \frac{1}{2} e^{z_1+3z_2+z_3+3z_4+z_5+3z_6} + \frac{1}{3} e^{3z_3+3z_6} \right)$$

for  $z_1 = z_2 = \dots = z_6 = 0$ .

We obtain, therefore :

$$B(6,3) = \frac{1}{12} \left( \frac{3^6}{6} + \frac{1}{2} + 0 + \frac{4 \cdot 3^3}{6} + \frac{4 \cdot 3^3}{2} + 0 + \frac{3 \cdot 3^2 \cdot 3^2}{6} + \frac{3 \cdot 3^2}{2} + 0 + \frac{2 \cdot 3^2}{6} + \frac{2}{2} + \frac{2 \cdot 3^2}{3} + \frac{2 \cdot 3}{6} + \frac{2 \cdot 3}{2} + \frac{2 \cdot 3}{3} \right) = 22.$$

If, in the Polya-de Bruijn theorem, we consider the particular case  $G = D_m$  and  $H = S_n$ , then we observe that the restriction of the  $\div$  relationship to the set of non-periodic and surjective functions  $f \in Y^X$ , denoted by  $F(m,n) \cap A_m$ , is exactly the chemical equivalence defined earlier.<sup>1</sup>

We observe that the equivalence class of a nonperiodic or a surjective

function is formed only from nonperiodic or surjective functions, respectively. Furthermore, the property of surjectivity is transmitted to any restriction to a period. Therefore we may write (denoting by  $N'(m)$  the number of  $\div$  equivalence classes from  $F(m,n)$ , and by  $N(m)$  the number of  $\div$  equivalence classes from  $F(m,n) \cap A_n$ , i.e. the number of classes of chemical equivalence we wish to obtain) :

$$N'(m,n) = B_{D_m, S_n}(m,n) - B_{D_m, S_{n-1}}(m,n-1)$$

$$N'(m,n) = \sum_{q|n} N(q)$$

By using the Möbius inversion theorem (described in the previous Part!) we obtain :

$$N(m,n) = \sum_{q|m} \mu(q,m) \left[ B_{D_q, S_n}(q,n) - B_{D_q, S_{n-1}}(q,n-1) \right]$$

where  $\mu(q,m)$  is the Möbius function of the lattice of divisors.

*Example.* For  $m = 6$  and  $n = 3$ , we calculate first by means of the Polya-de Bruijn theorem  $B(6,2) = 8$  ;  $B(3,3) = 3$  ;  $B(3,2) = B(2,2) = B(2,3) = 2$  ;  $B(1,2) = B(1,3) = 1$ . On using the result of the previous example,  $B(6,3) = 22$ , and on performing the calculations, we obtain

$$N(6,3) = (22 - 8) - (3 - 2) - (2 - 2) + (1 - 1) = 13.$$

This value coincides with that obtained for irreducible sequences of length 6 in ternary copolymers according to the computer program presented below ; the same value is displayed in Table 2.

An explicit formula (based on the Möbius inversion theorem) for computing the number  $N(m,n)$  of irreducible sequences

We denote by  $P_k$  the set of permutations  $\pi$  with order  $k$  in  $S_n$ , where  $k$  is a natural number. If  $e$  denotes the identity permutation, we have :  $\pi^k = e$ , and  $\pi^u \neq e$  for any  $u \in [1, k-1]$ . It may be observed that  $k$  is the lowest common multiple of the lengths of cycles in the unique decomposition of  $\pi$  into disjoint cycles.

A permutation is said to be of type  $(\lambda_1, \lambda_2, \dots, \lambda_n)$  if it contains  $\lambda_i$  ( $i \in [1, n]$ ) cycles of length  $i$  in its decomposition into disjoint cycles.



Evidently,  $\sum_{i=1}^n i\lambda_i = n$ .

According to Cauchy's formula, the number  $h(\lambda_1, \lambda_2, \dots, \lambda_n)$  of permutations of type  $(\lambda_1, \lambda_2, \dots, \lambda_n)$  is :

$$h(\lambda_1, \lambda_2, \dots, \lambda_n) = \frac{n!}{\lambda_1! \lambda_2! \dots \lambda_n! 1^{\lambda_1} 2^{\lambda_2} \dots n^{\lambda_n}}$$

As a simplified notation, whenever  $n$  is known from the context, we shall write  $h(\lambda_1, \lambda_2)$  instead of  $h(\lambda_1, \lambda_2, 0, \dots, 0)$ .

The set of  $n$ -uples indicating the permutation types from  $S_n$  which have order  $k$  is denoted by  $W_k$ . The cardinal  $|W_k|$  is the number of permutation types with order  $k$ , and the number of permutations with order  $k$  is :

$$\sum_{(\lambda_1, \dots, \lambda_n) \in W_k} h(\lambda_1, \dots, \lambda_n)$$

Given two functions  $\alpha : [0, a-1] \rightarrow [0, b-1]$  and  $\beta : [0, c-1] \rightarrow [0, b-1]$ , we define another function  $\alpha\beta$  called "concatenation of  $\alpha$  with  $\beta$ " as follows :

$$\alpha\beta : [0, a+c-1] \rightarrow [0, b-1]$$

$$(\alpha\beta)(i) = \begin{cases} \alpha(i) & \text{if } i \in [0, a-1] \\ \beta(i-a) & \text{if } i \in [a, a+c-1] \end{cases}$$

As an example, let for an alphabet of  $n = 3$  symbols the two functions be :  $\alpha = 1021$  and  $\beta = 21$ . Then we have the concatenations  $\alpha\beta = 102121$  and  $\beta\alpha = 211021$ .

We shall use definitions and notation employed in the first part of this series. Thus, cyclic equivalence will be denoted by  $\approx$ . If a function  $\alpha$  remains in the same class of cyclic equivalence after composition with another function (e.g. a permutation), we say that  $\alpha$  is cyclically invariant with respect to this function. We denote by  $M_k$  the set of non-periodic functions from  $F(m, n)$  which are invariant with respect to a permutation from  $P_k$ . In mathematical form,

$$M_k = \{ \alpha \in F(m, n) \cap A_m \mid \exists \pi \in P_k \text{ such that } \alpha \approx \pi(\alpha) \}$$

The set of functions in  $M_k$  which do not belong to  $M_p$  (for any  $p$ , multiple of  $k$ ) is denoted by  $L_k$ . A composition of  $\alpha$  with the permutation

$\pi = (a-1, a-2, \dots, 0)$  represents a function, which will be denoted by  $\hat{\alpha}$ , that we call the symmetrical transpose of  $\alpha$ . We define  $L'_k$  as the set of functions in  $L_k$  such that their symmetrical transpose are cyclically invariant with respect to a permutation. In mathematical form,

$$L'_k = \{ \alpha \in L_k \mid \exists \pi \in S_n \text{ such that } \pi(\alpha) = \hat{\alpha} \}$$

As an example,  $\alpha = 012102 \in F(6,3)$  ;  $\pi = (1,0,2) \in P_2$  ;  $\pi(\alpha) = 102012 \simeq \alpha$ , hence  $\alpha \in M_2$ . Taking into account that  $P_{2k} = \emptyset$  for any  $k \geq 2$  (according to Fermat's theorem from the theory of finite groups, the order of any permutation from  $S_n$  must divide  $n!$ ),  $\alpha$  must belong also to  $L_2$ . Now we examine the symmetrical transpose of  $\alpha$  :

$$\hat{\alpha} = 201210 \text{ ; } \pi(\alpha) = 102012 \simeq \hat{\alpha}$$

Therefore we find that  $\alpha \in L'_2$ .

*Observations.*

1. If  $\alpha \in L_k$ , then  $\pi(\alpha) \in L_k$  for any  $\pi \in S_n$ .
2.  $\{P_k\}_{k|n!}$  is a partition of  $S_n$ .
3. If  $\alpha \in F(m,n)$ , for any two different permutations  $\pi$  and  $\sigma$  from  $S_n$  we have  $\pi(\alpha) \neq \sigma(\alpha)$ .
4. If  $\alpha \in L_k$  then there exists a unique permutation  $\pi \in P_k$  such that  $\pi^0(\alpha) \simeq \pi(\alpha) \simeq \dots \simeq \pi^{k-1}(\alpha)$ .

For any  $\sigma \in S_n$  such that  $\sigma^k \neq e$ ,  $\alpha$  is not cyclically equivalent to  $\sigma(\alpha)$ , and we have :  $\sigma(\alpha) \simeq \sigma\pi(\alpha) \simeq \dots \simeq \sigma\pi^{k-1}(\alpha)$ . Thus for  $\alpha \in L_k$ , the set  $\{\pi(\alpha) \mid \pi \in S_n\}$ , having cardinal  $n!$ , may be partitioned in disjoint subsets of  $k$  elements each. Each of these subsets is included in a class of cyclic equivalence. As an example, again for an alphabet of three symbols, we define the function  $\alpha = 011220 \in F(6,3)$ , and we observe that  $\alpha \in L_3$ . The symmetric group of order 3 consists of  $e = (0,1,2)$  ;  $\pi_1 = (1,0,2)$  ;  $\pi_2 = (2,1,0)$  ;  $\pi_3 = (0,2,1)$  ;  $\sigma_1 = (1,2,0)$  and  $\sigma_2 = (2,0,1)$ . The set  $\{\xi(\alpha) \mid \xi \in S_n\}$  is partitioned into subsets  $\{e(\alpha), \sigma_1(\alpha), \sigma_2(\alpha)\}$ ,  $\{\pi_1(\alpha), \pi_2(\alpha), \pi_3(\alpha)\}$  such that :  $e(\alpha) = 011220 \simeq \sigma_1(\alpha) = 122001 \simeq \sigma_2(\alpha) = 200112$ , and  $\pi_1(\alpha) = 100221 \simeq \pi_2(\alpha) = 211002 \simeq \pi_3(\alpha) = 200110$ .

*Proposition 1.* The number, denoted by  $N(m,n)$ , of non-periodic, chemically non-equivalent sequences with length  $m$  over an alphabet with  $n$  symbols is :

$$(1) \quad N(m,n) = \frac{1}{2m \cdot n!} \sum_{k|m} k(|L_k| + |L'_k|)$$

*Proof.* Let  $k$  be a natural number which divides  $m$ . For  $\alpha \in L_k$  we have the set  $\{\xi(\alpha) \mid \xi \in S_n\}$  which may be partitioned into  $n!/k$  subsets from different classes of cyclic equivalence (see Observation 4). Thus the set  $L_k$  contains  $|L_k|/m$  classes of cyclic equivalence. For obtaining the cardinal of a system of representatives which are independent of the  $S_n$  permutation group (i.e. which cannot be obtained from one another by a permutation within  $S_n$  and, possibly, by a cyclic permutation) we have to divide further by  $n!/k$ .

Taking into account that  $L_k' \subseteq L_k$ , it results that  $|L_k - L_k'| = |L_k| - |L_k'|$ .

In order to make the representatives of classes of chemical equivalence independent relative to symmetry, we have to consider that :

- (i) in  $L_k'$  the number of classes of chemical equivalence is  $|L_k'| \cdot k / (m \cdot n!)$ .
- (ii) in  $L_k$  we have  $(|L_k| - |L_k'|) \cdot k / (2m \cdot n!)$  classes.

Therefore in  $L_k$  we have  $(|L_k| + |L_k'|) \cdot k / (2m \cdot n!)$  classes. By summing over  $k|m$ , we obtain the formula (1). An example for  $n = 2$  and  $m = 4$  with  $\pi = (1,0)$  is :

$$\begin{aligned} \alpha_1 &= 0001 ; \pi(\alpha_1) = 1110 ; \hat{\alpha}_1 = 1000 ; \pi(\hat{\alpha}_1) = 0111 \\ \alpha_2 &= 0010 ; \pi(\alpha_2) = 1101 ; \hat{\alpha}_2 = 0100 ; \pi(\hat{\alpha}_2) = 1011 \\ \alpha_3 &= 0011 ; \pi(\alpha_3) = 1100 ; \hat{\alpha}_3 = 1100 ; \pi(\hat{\alpha}_3) = 0011 \\ \alpha_4 &= 0100 ; \pi(\alpha_4) = 1011 ; \hat{\alpha}_4 = 0010 ; \pi(\hat{\alpha}_4) = 1101 \\ \alpha_5 &= 0110 ; \pi(\alpha_5) = 1001 ; \hat{\alpha}_5 = 0110 ; \pi(\hat{\alpha}_5) = 1001 \\ \alpha_6 &= 0111 ; \pi(\alpha_6) = 1000 ; \hat{\alpha}_6 = 1110 ; \pi(\hat{\alpha}_6) = 0001 \\ \alpha_7 &= 1000 ; \pi(\alpha_7) = 0111 ; \hat{\alpha}_7 = 0001 ; \pi(\hat{\alpha}_7) = 1110 \\ \alpha_8 &= 1001 ; \pi(\alpha_8) = 0110 ; \hat{\alpha}_8 = 1001 ; \pi(\hat{\alpha}_8) = 0110 \\ \alpha_9 &= 1011 ; \pi(\alpha_9) = 0100 ; \hat{\alpha}_9 = 1101 ; \pi(\hat{\alpha}_9) = 0010 \\ \alpha_{10} &= 1100 ; \pi(\alpha_{10}) = 0011 ; \hat{\alpha}_{10} = 0011 ; \pi(\hat{\alpha}_{10}) = 1100 \\ \alpha_{11} &= 1101 ; \pi(\alpha_{11}) = 0010 ; \hat{\alpha}_{11} = 1011 ; \pi(\hat{\alpha}_{11}) = 0100 \\ \alpha_{12} &= 1110 ; \pi(\alpha_{12}) = 0001 ; \hat{\alpha}_{12} = 0111 ; \pi(\hat{\alpha}_{12}) = 1000 \\ L_1^+ &= \{\alpha_1, \alpha_2, \alpha_4, \alpha_6, \alpha_7, \alpha_9, \alpha_{11}, \alpha_{12}\} = L_1 \\ L_2^+ &= \{\alpha_3, \alpha_5, \alpha_8, \alpha_{10}\} = L_2 \end{aligned}$$

According to the above formulas, we obtain that in  $L_1^+$  there is  $|L_1^+|/2 \cdot 4 = 1$  class of chemical equivalence, and that in  $L_2^+$  there is  $|L_2^+| \cdot 2/2 \cdot 4 = 1$  class of chemical equivalence, therefore formula (1) yields  $N(4,2) = 2$ .

*Lemma 1.* If  $\alpha \in L_k$ , then  $k$  divides into  $m$  and there exist : permutation  $\pi \in P_k$  and subsequence  $\gamma$  of length  $m/k$  such that  $\alpha = \pi^0(\gamma)\pi^1(\gamma)\dots\pi^{k-1}(\gamma)$ .

For the above-mentioned example  $\alpha = 011220 \in F(6,3)$  we observe that

$\alpha \in L_3$ ,  $\alpha = \pi^{\sharp}(\gamma)\pi^1(\gamma)\pi^2(\gamma)$  where  $\alpha = 01$ , and  $\pi = (1,2,0) \in P_3$ .

*Lemma 2.*

$$|F(m,n) \cap A_m| = \sum_{k|m} |L_k|$$

The proof is based on reductio ad absurdum, by supposing the existence of  $\alpha \in L_i \cap L_j$ ,  $i < j$ . From the definition of  $L_k$  it results that  $i$  is not a divisor of  $j$ , and including also Lemma 1 one obtains a contradiction.

For the sake of the next Lemma we make use of the set of triples  $H(m,n) = \{(\alpha,\pi,t) \in (F(m,n) \cap A_m) \times S_n \times [0,m-1]$  with the property that  $\alpha(i) = \pi(\alpha(t \oplus (-i)))$  for any  $i \in [0,m-1]\}$ .

The operation  $\oplus$  (sum modulo  $m$ ) is :

$$t \oplus (-i) = \begin{cases} t - i & \text{for } i \in [0,t] \\ m + t - i & \text{for } i \in [t+1,m-1] \end{cases}$$

For example, if  $\alpha = 010020 \in F(6,3) \cap A_6$ ,  $\pi = (0,2,1) \in S_3$  ; we choose  $t_1 = 2$ ,  $t_2 = 5$ . Hence,  $\{(\alpha,e,t_1),(\alpha,\pi,t_2)\} \subset H(6,3)$ . We observe that  $\alpha \in L_2$ .

*Lemma 3.*

$$|H(m,n)| = \sum_{k|m} k|L'_k|$$

The proof consists in showing that for each  $\alpha \in L'_k$  there exist exactly  $k$  pairs  $(\pi,t)$  that satisfy the condition  $(\alpha,\pi,t) \in H(m,n)$ .

*Lemma 4.* For any natural number  $k$

$$|M_k| = \sum_{k|d} |L_d|$$

*Proof.* From the definition of  $M_k$  it results that  $\{L_d\}_{k|d} = \{L_d$  with the property that  $k$  divides into  $d\}$  is a partition of  $M_k$ .

*Proposition 2.*

$$(2) \quad N(m,n) = \frac{1}{2m \cdot n!} \left( |F(m,n) \cap A_m| + |H(m,n)| + \sum_{\substack{k|n \\ k \neq 1}} (k-1) \sum_{\substack{p|m \\ k|p}} \mu(k,p) |M_p| \right)$$

where  $\mu$  is the Möbius function of the lattice of divisors.

*Proof.* From Lemmas 2 and 3 and formula (1) we obtain :

$$(3) \quad N(m,n) = \frac{1}{2m \cdot n!} \left( |F(m,n) \cap A_m| + |H(m,n)| + \sum_{\substack{k|m \\ k \neq 1}} (k-1) |L_k| \right)$$

With the help of the  $\mu$  function of the lattice of divisors, a new Möbius function  $\mu^*$ , is introduced such that :

$\mu^* : [1, m] \times [1, m] \rightarrow \{-1, 0, 1\}$ ,  $\mu^*(p, q) = \mu(q, p)$

for any  $q$  and  $p$  from the set of the divisors of  $m$ . One observes that  $[1, m]$  is a locally finite set, which is partially ordered according to the order relationship defined as :  $x, y \in [1, m]$ ,  $x \leq y$  iff  $y|x$ . The universal minorant is  $m$ .

One demonstrates easily that the function  $\mu^*$  verifies the conditions of Möbius functions.

From Lemmas 1 and 4 we obtain  $|M_k| = \sum_{\substack{k|p \\ p|m}} |L_p|$ , therefore  $|M_k| =$   
 $= \sum_{m \leq p \leq k} |L_p|$ , where  $\leq$  denotes the partial order relationship defined above

on  $[1, m]$ . By applying the Möbius inversion theorem we obtain :

$$|L_k| = \sum_{\substack{k|p \\ p|m}} \mu^*(p, k) \quad |M_p| = \sum_{\substack{k|p \\ p|m}} \mu(k, p) |M_k|$$

On replacing this result in formula (3) we obtain formula (2).

*Lemma 5.*

$$|F(m, n) \cap A_m| = \sum_{q|m} \mu(q, m) \cdot s_{q, n}$$

As indicating earlier,  $\mu(q, m)$  is the Möbius function of the lattice of divisors, and  $s_{q, n}$  is the number of surjections from a set with  $q$  elements to a set with  $n$  elements.

*Proof.* Let  $P(m, n)$  be a set of functions with a property  $\Xi$  ; let this property be transmitted to any restriction of the function to a period. Then we may compute the number of non-periodic functions which possess property  $\Xi$  by using the Möbius function  $\mu$  and Möbius's inversion theorem:

$$|P(m, n) \cap A_m| = \sum_{q|m} \mu(q, m) |P(q, n)|$$

If we make the assignment  $P(m, n) = F(m, n)$ , and if we assume that property  $\Xi$  is surjectivity (knowing that a surjective function restricted to one period is also a surjective function), we obtain Lemma 5.

In order to obtain the next lemma, we start from a natural number  $r$ , and from the  $n$ -uple of natural numbers  $(\lambda_1, \dots, \lambda_n)$ . We assume that there exist  $\lambda_i$  boxes, each containing  $i$  distinct numbers from  $[0, r-1]$ . The number of functions  $\gamma : [0, r-1] \rightarrow [0, n-1]$  with the property that the set of values of these functions contains at least one element from each box is denoted by  $v(r ; \lambda_1, \dots, \lambda_n)$ .

*Lemma 6.*

$$v(r; \lambda_1, \dots, \lambda_n) = n^r - \sum_{l=1}^z (-1)^{l+1} \sum_{\substack{j_1 + \dots + j_n = l \\ 0 \leq j_i \leq \lambda_i}} \binom{\lambda_1}{j_1} \dots \binom{\lambda_n}{j_n} \left( n - \sum_{i=1}^n i j_i \right)^r$$

where  $z = \sum_{i=1}^n \lambda_i$ .

*Proof.* One calculates the cardinal of the set of functions with the property that there exists at least one box whose elements are not present in the set of values of functions, then one may obtain the cardinal  $v(r; \lambda_1, \dots, \lambda_n)$  of the complement of this set.

*Lemma 7.* For two natural numbers  $p$  and  $m$  ( $p$  divides into  $m$ , i.e.  $p|m$ ), we have :

$$|M_p| = \sum_{\substack{p|q|m \\ q \nmid \frac{m}{p}}} \mu(q, m) \sum_{(\lambda_1, \dots, \lambda_n) \in W_p} h(\lambda_1, \dots, \lambda_n) \cdot v\left(\frac{q}{p}; \lambda_1, \dots, \lambda_n\right)$$

where  $p|q|m$  is read as  $p|q$ , and  $q|m$ .

*Proof.* When  $p|q$ , then  $\alpha \in M_p$  iff there exists a unique permutation  $\alpha \in P_p$ , and if there exists a function  $\gamma : [0, q/p-1] \rightarrow [0, n-1]$  such that  $\alpha = \pi^0(\gamma)\pi^1(\gamma)\dots\pi^{p-1}(\gamma)$ , and  $\alpha \in A_m$ . We therefore have a bijection between  $M_k$  and the set of pairs  $(\pi, \gamma)$  from the previous sentence ; the numbers of functions  $\gamma$ , which, for a given permutation  $\pi \in P_p$  of type  $(\lambda_1, \lambda_2, \dots, \lambda_n)$  have the property that  $\pi^0(\gamma)\pi^1(\gamma)\dots\pi^{p-1}(\gamma)$  is a surjection, is exactly  $v(q/p; \lambda_1, \dots, \lambda_n)$ . On following the reasoning from the proof of Lemma 6 for the elimination of periodic sequences by means of the Möbius function  $\mu$ , we obtain Lemma 7.

*Lemma 8.* Let  $\alpha \in F(m, n) \cap A_m$ . If there exists a permutation  $\pi \in S_n$  such that  $\pi(\alpha) = \hat{\alpha}$ , then  $\pi^2 = e$ .

*Proof.* If  $\pi(\alpha) = \hat{\alpha}$ , then there exists  $t \in [0, m-1]$  such that  $\alpha(i) = \pi(\alpha(t \oplus (-i)))$  for any  $i \in [0, m-1]$ . Hence  $\alpha(i) = \pi^2(\alpha(i))$  for any  $i \in [0, m-1]$ . Since  $\alpha$  is a surjection, it results that  $\pi^2 = e$ .

*Lemma 9.*

$$|H(m, n)| = \sum_{\lambda_1 + 2\lambda_2 = r} h(\lambda_1, \lambda_2) \cdot m \left[ \sum_{q|p} \mu(q, m) \lambda_1 v\left(\left\lfloor \frac{q-1}{2} \right\rfloor; \lambda_1-1, \lambda_2\right) + \right. \\ \left. + \frac{1}{2} \sum_{2|q|m} \mu(q, m) \left( v\left(\frac{q}{2}; \lambda_1, \lambda_2\right) - \lambda_1 v\left(\frac{q}{2} - 1; \lambda_1-1, \lambda_2\right) + \lambda_1(\lambda_1-1) v\left(\frac{q}{2} - 1; \lambda_1-2, \lambda_2\right) \right) \right]$$

We have denoted by  $v(r ; \lambda_1, \lambda_2)$  the cardinal  $v(r ; \lambda_1, \lambda_2, 0, \dots, 0)$ .

*Proof.* By using Lemma 8 and the transmittance, over any restriction to one period, of the invariance to a permutation from  $S_n$  and the symmetry, we obtain :

$$(4) \quad |H(m, n)| = \sum_{q|m} \mu(q, m) \sum_{\lambda_1 + 2\lambda_2 = n} h(\lambda_1, \lambda_2) \frac{m}{q} v'(q ; \lambda_1, \lambda_2)$$

where  $v'(q ; \lambda_1, \lambda_2)$  represents the number of triplets  $(\alpha, \pi, t) \in F(q, n) \times S_n \times [0, q-1]$  with the property that  $\pi$  is of type  $(\lambda_1, \lambda_2, 0, \dots, 0)$  and  $\alpha(i) = \pi(\alpha(t \oplus (-i)))$  for any  $i \in [0, q-1]$ . We have two cases to consider, namely :

(A) if  $q$  is even,

$$v'(q ; \lambda_1, \lambda_2) = \frac{q}{2} \left[ v\left(\frac{q}{2} ; \lambda_1, \lambda_2\right) + \lambda_1 v\left(\frac{q}{2} - 1 ; \lambda_1 - 1, \lambda_2\right) + \lambda_1(\lambda_1 - 1) \cdot v\left(\frac{q}{2} - 1 ; \lambda_1 - 2, \lambda_2\right) \right]$$

(B) if  $q$  is odd,

$$v'(q ; \lambda_1, \lambda_2) = q \lambda_1 v\left(\frac{q-1}{2} ; \lambda_1 - 1, \lambda_2\right).$$

On replacing in (4) we obtain Lemma 9.

*Theorem.* For  $n \geq 2$  and  $m \geq n$ , the number  $N(m, n)$  of non-periodic, chemically non-equivalent sequences with length  $m$  over an alphabet of  $n$  symbols is (for  $|H(m, n)|$  Lemma 9 gives its explicit form) :

$$N(m, n) = \frac{1}{2m \cdot n!} \left[ \sum_{q|m} \mu(q, m) s_{q, n} + |H(m, n)| + \sum_{k|m} (k-1) \sum_{k|p|m} \mu(k, p) \cdot \sum_{\substack{p|q|m \\ q \nmid \frac{m}{p}}} (q, m) \cdot \sum_{(\lambda_1, \dots, \lambda_n) \in W_p} h(\lambda_1, \dots, \lambda_n) v\left(\frac{q}{p} ; \lambda_1, \dots, \lambda_n\right) \right]$$

*The proof* is obtained by replacing in formula (2) the results of Lemmas 5, 7, and 9.

*Corollary.* For the particular case when  $m$  is odd,

$$N(m, n) = \frac{1}{2m \cdot n!} \left[ \sum_{q|m} \mu(q, m) \left( s_{q, n} + m \sum_{\lambda_1 + 2\lambda_2 = n} h(\lambda_1, \lambda_2) \lambda_1 v\left(\frac{q-1}{2} ; \lambda_1 - 1, \lambda_2\right) \right) + \sum_{k|m} (k-1) \sum_{k|p|m} \mu(k, p) \sum_{\substack{p|q|m \\ q \nmid \frac{m}{p}}} \mu(q, n) \sum_{(\lambda_1, \dots, \lambda_n) \in W_p} h(\lambda_1, \dots, \lambda_n) v\left(\frac{q}{p} ; \lambda_1, \dots, \lambda_n\right) \right]$$

We shall now examine a few particular cases for binary, ternary and quaternary copolymers.

1°. The case  $n = 2$  (binary copolymers, and other isomorphic sequences such as stereoregular homopolymers or elastomers, which were examined in the first part of this series<sup>1</sup>).

$$N(m,n) = \frac{1}{4m} \left( \sum_{q|m} \mu(q,m)2^q + m \sum_{q|m} \mu(q,m)2^{\lceil q/2 \rceil} + m \sum_{\substack{2|q|m \\ q \neq \frac{m}{2}}} \mu(q,m)2^{q/2} \right)$$

One obtains thus the result of Theorem 1 from Part 1 of the present series, in slightly different notation. As in the corollaries of that paper, one may further particularize the formula for even or odd  $m$  values.

2°. The case  $n = 3$  (ternary copolymers) :

$$N(m,n) = \frac{1}{12m} \left[ \sum_{q|m} \mu(q,m)(s_{q,3} + m(2 \cdot 3^{\lceil q/2 \rceil} - 3 \cdot 2^{\lceil q/2 \rceil})) + \sum_{\substack{2|q|m \\ q \neq \frac{m}{3}}} \mu(q,m)m(2 \cdot 3^{q/2} - 3 \cdot 2^{q/2}) + 4 \sum_{\substack{3|q|m \\ q \neq \frac{m}{3}}} \mu(q,m) \cdot 3^{q/2} + 3 \sum_{\substack{2|q|m \\ q \neq \frac{m}{2}}} \mu(q,m)(3^{q/2} - 2^{q/2} - 1) \right]$$

3°. The case  $n = 4$  (quaternary copolymers) :

$$N(m,n) = \frac{1}{48m} \left[ \sum_{q|m} \mu(q,m) \left[ s_{q,4} + 4m(4^{\lceil q/2 \rceil} - 2 \cdot 3^{\lceil q/2 \rceil} + 2) \right] + 2 \sum_{\substack{2|q|m \\ q \neq \frac{m}{2}}} \mu(q,m)(3 \cdot 4^{q/2} - 4 \cdot 3^{q/2}) + 3 \sum_{\substack{2|q|m \\ q \neq \frac{m}{2}}} \mu(q,m)(3 \cdot 4^{q/2} - 4 \cdot 3^{q/2} - 2 \cdot 2^{q/2} + 4) + \right]$$



$$+ 16 \sum_{\substack{3|q|m \\ q \nmid \frac{m}{3}}} \mu(q,m) (4^{q/3} - 3^{q/3} - 1) + 12 \sum_{\substack{4|q|m \\ q \nmid \frac{m}{4}}} \mu(q,m) \cdot 4^{q/4} \Big]$$

In comparison with applying the Polya-de Bruijn theorem, the above expressions, which do not contain partial derivatives, present the advantage that they may be easily implemented by means of a pocket calculator.

For computing values of  $s_{m,n}$  (the number of surjections from a set with  $m$  elements to a set with  $n$  elements), one may use the recurrence relationship :

$$s_{m+1} = ns_{m,n-1} + s_{m,n}$$

(where  $s_{m,1} = 1$ , and  $s_{m,n} = m!$ ).

This relationship is obtained from the two formulas indicated below where  $S(m,n)$  is the Stirling number of the second kind, i.e. the number of partitions of a set with  $m$  elements into  $n$  classes :

$$S(m,n) = \frac{1}{n!} s_{m,n}$$

$$S(m+1,n) = S(m,n-1) = nS(m,n)$$

(where  $S(m,1) = S(m,m) = 1$ ).

#### Asymptotic behaviour of the numbers $N(m,n)$ of irreducible sequences

In the preceding part<sup>1</sup> it was shown that for  $m \rightarrow \infty$  the asymptotic value for  $N(m,2)$  is  $2^{m-2}/m$ .

One can easily demonstrate that in the general case the asymptotic limit for  $N(m,n)$  is :

$$\lim_{n \rightarrow \infty} = n^m / 2m \cdot n!$$

One sees that for  $n = 2$  this formula becomes identical to that indicated above.

Computer program for generating an ordered list of irreducible sequences

The SEQV program, written in PASCAL using structured programming, and displayed in Table 3, generates a system of representatives of classes of chemical equivalence for sequences of length  $N \leq NMAX$  which contain symbols from a set with cardinal  $M$ . It will be observed that from now onwards the letters  $M$  and  $N$  have been permuted relatively to the previous text and formulas (where  $n$  and  $m$ , respectively, had been used)\* ; the total number of irreducible sequences will therefore be  $N(m,n) = N(N,M)$ .

The program was implemented on a Roumanian-made microcomputer M-216 (IBM-PC comparable) based upon the INTEL-8086 microprocessor.

The main program calls the following procedures :

INCREASE, which furnishes the next sequence (in lexicographic order) which is a candidate for being an irreducible sequence.

MINIMAL, which tests if the sequence being analysed is minimal in its equivalence class (from all sequences of this equivalence class, only one representative, namely the minimal one, is displayed).

PERIOD, which tests if the sequence is decomposable into repeating subsequences (i.e. if it is periodical).

DISPLAY which prints or displays the representatives.

The algorithm selects all minimal sequences in their equivalence classes, and displays them as representatives of these classes in lexicographic order ; it considers as alphabet the set  $\{1,2,\dots,M\}$  with the order relationship from the set of the first  $M$  natural numbers.

Only those sequences will be considered which contain at least one from each symbol of the alphabet, i.e. only surjections will be considered. Consequently, the search of irreducible sequences will have as range  $N \in [M, NMAX]$ . For this purpose the array IND is used ; it indicates the position (rank) of the first appearance of symbols in the sequence.

Taking into account that all irreducible sequences start with a string

---

\*We wish to conserve in the computer print-out (Table 4) the notation  $N$  for the length of the sequence, as in Part I<sup>1</sup> ; on the other hand, in the title of the present paper we wished to refer to  $n$ -ary copolymers, hence the correspondence  $n = M$  and  $m = N$ .

(block) of  $\ell$  symbols 1 (translated as RR...R), the sequences which contain a longer block of any other symbol cannot act as representatives because a permutation of the corresponding symbols followed by a circular permutation of the positions leads to a smaller sequence. The array BLOCK is used for memorizing the position of subsequences forming blocks of length  $\ell$  (when all but the first blocks have smaller lengths than  $\ell$ , the sequence is irreducible).

By using the arrays IND and BLOCK, the procedure INCREASE provides to the main program the next sequence having the property that it is surjective and that all its blocks have lengths smaller than, or equal to,  $\ell$ . Evidently, no irreducible sequence may end in 1 (corresponding to R). Finally the program ignores (by means of the PERIOD procedure) any periodical sequence.

On reaching a sequence which contains in its first M positions all ordered symbols of the alphabet, the algorithm goes on to sequences with length  $N+1$  (if  $N < NMAX$ ) or stops (if  $N = NMAX$ ).

In Table 4 we present the results of the program for  $N \in [3,10]$  and  $M = 3$ ; in Table 5 for  $N \in [4,9]$  and  $M = 4$ ; and in Table 6 for  $N \in [5,9]$  and  $M = 5$ .

The upper limit  $NMAX$  is not inherent in the program and may be easily increased above 10, to any desired number. Also, the  $M$  value may be increased. Of course, such increases will lead to enhanced execution times. A comparison of the present program SEQV with the previous program<sup>1</sup> (which was written in FORTRAN-IV and had  $M = 2$ ) shows that the execution time of SEQV is lower, and that its algorithm is more performing.

TABLE 3. Computer program SEQV

```
program seqv;
const   nmax:=10;
type    s1= array [1..10] of integer;
        i1= array [1..10] of integer;
        ds= array [1..6] of integer;
        data=text;
var     ddd:data;
        seqv,block:s1;
        ind:i1;
        divis:ds;
        test,ovrfl:boolean;
        aux,c,i,ii,j,k,l,ll,m,n,nr,p,q,s,t: integer;
procedure increase(var k,p:integer);
var l,i: integer;
begin
  if seqv[k]=m then
    begin while seqv[k]=m do k:=k-1;
           if ind[m]=k+1 then s:=s-1;aux:=seqv[k];
           while ind[aux]=k do
             begin k:=k-1;s:=s-1;aux:=seqv[k]
             end
           end;
    seqv[k]:=seqv[k]+1;aux:=seqv[k];
    if ind[aux]>k then
      begin s:=s+1;ind[aux]:=k end;
    if seqv[k]=seqv[k-1] then
      begin i:=1;
           while i<=p do
             begin if block[i]<k-1
                    then i:=i+1
                    else if block[i]=k-1 then
                      begin ovrfl:=false;i:=p+1 end
                      else p:=i-1
                    end;
             l:=1;j:=k;
             while seqv[j]=seqv[j-1] do
               begin l:=l+1;j:=j-1 end;
             if l=block[l] then begin p:=p+1;block[p]:=k end
           end
        else
          begin i:=1;
               while i<=p do
                 begin if block[i]<k then i:=i+1
                        else begin if block[i]=k then
                                begin if i=1 then
                                  begin block[l]:=k-1;p:=1 end
                                  else p:=i-1
                                end
                                else p:=i-1;
                                      i:=p+1
                                end
                             end
                end
          end
        if block[l]=1 then
          begin p:=p+1;block[p]:=k;l:=1;i:=k+1;
               while i<=(n-(m-s+1)) do
                 begin seqv[i]:=1;p:=p+1;block[p]:=i;
```

```
        l:=l mod 2+1;i:=i+1
    end;
    l:=s;
    while i<n do
    begin ind[l]:=i;seqv[i]:=l;
        l:=l+1;p:=p+1;block[p]:=i;i:=i+1
    end;
    if seqv[n]=1 then
    begin if seqv[n-1]=2 then if ovrfl then
        begin ovrfl:=false;k:=n end;
            seqv[n]:=2
        end
    end
else
begin l:=0;i:=k+1;
    while i<n-(m-s+1) do
    begin seqv[i]:=1;l:=l+1;
        if l=block[l] then
        begin l:=0;p:=p+1;block[p]:=i;
            if i<n-(m-s+1) then
            begin i:=i+1;seqv[i]:=2 end
            end;
            i:=i+1
        end;
        l:=s;
        while i<n do
        begin ind[l]:=i;seqv[i]:=l;i:=i+1;l:=l+1 end;
        if seqv[n]=1 then
        begin if (block[l]=2) and (seqv[n-1]=2) then
            begin if block[p]=n-1 then
                begin if ovrfl then begin ovrfl:=false;k:=n end
                    end
                else begin p:=p+1;block[p]:=n end
            end
            else if block[p]=n then p:=p-1;
                seqv[n]:=2
            end
        end
    end;
    if ovrfl then k:=n+s-m-1
end;{increase}
procedure minimal(c,dir,j:integer);
    var aux,i,niv:integer;
        perm: array[1..5] of integer;
begin for i:=1 to m do perm[i]:=0;
    niv:=0;
    while c<=n do
    begin aux:=seqv[j];
        if perm[aux]=0 then
        begin niv:=niv+1;perm[aux]:=niv end;
        if perm[aux]=seqv[c] then
        begin c:=c+1;j:=(j+n-1+dir) mod n+1 end
        else
        begin if perm[aux]<seqv[c] then test:=false;
            c:=n+1
        end
        end
    end
end;{minimal}
```

```
procedure period(indc:integer;d:ds);
  var aux,i,k,q,r:integer;
begin if indc<=d[1]] then
  begin i:=1;while d[i]<indc do i:=i+1;
    while (i<=1) and (test=true) do
      begin test:=false;r:=1;aux:=d[i];
        while r<=aux do
          begin q:=1;k:=n div aux;
            while q<k do
              begin if seqv[r]=seqv[aux*q+r] then q:=q+1
                else
                  begin test:=true;q:=k;r:=aux end
                end;
                r:=r+1
              end;
              i:=i+1
            end
          end
        end;{period}
  procedure display(var ddd:data;var ii:integer);
    var alph:array[1..10] of char;
      i:integer;
  begin for i:=1 to n do
    begin if seqv[i]=1 then write(ddd,'R')
      else if seqv[i]=2 then write(ddd,'S')
        else if seqv[i]=3 then write(ddd,'T')
          else if seqv[i]=4 then write(ddd,'U')
            else write(ddd,'V')
          end;if ii=5 then begin ii:=0;writeln(ddd);
            write(ddd,' ')
          end
        else write(ddd,' ')
      end;{display}
  begin{main program}
    assign(ddd,'rez.dta');
    rewrite(ddd);
    read(m);writeln;
    for n:=m to nmax do
      begin write(ddd,' N=',n);writeln(ddd);
        for i:=1 to n-m+1 do seqv[i]:=1;
          for i:=2 to m do
            begin seqv[n-m+i]:=i;ind[i]:=n-m+i end;
            ind[1]:=1;p:=1;block[1]:=n-m+1;ii:=1;
            write(ddd,' ');display(ddd,ii);
            s:=2;test:=true;ovrfl:=true;ll:=0;aux:=n div 2;k:=n-m+1;
            for q:=m+1 to aux do
              begin if n div q*q=n then
                begin ll:=ll+1;divis[ll]:=q end
              end;
              while ind[m]>=m do
                begin increase(k,p);
                  if ovrfl then
                    begin minimal(block[1],-1,1);
                      if test then
                        begin nr:=2;
                          while nr<=p do
                            begin c:=block[1];
```

```

                                j:=block[nr]-block[l]+1;
                                minimal(c,-1,j);
                                if test then
                                minimal(block[l],1,block[nr])
                                else nr:=p;
                                nr:=nr+1
                                end;
                                end;
                                end;
                                if test then
                                begin if ll<>0 then period(ind[m],divis);
                                if test then begin ii:=ii+1;
                                display(ddd,ii)
                                end
                                else test:=true
                                end
                                else test:=true
                                end
                                else ovrf1:=true
                                end;
                                if ii<5 then writeln(ddd);writeln(ddd);writeln(ddd)
                                end;close(ddd)
                                end.{main program}
```

TABLE 4. IRREDUCIBLE SEQUENCES FOR AN ALPHABET OF THREE LETTERS: R,S,T.

N=3	RST				
N=4	RRST	RSRT			
N=5	RRRST	RRSRT	RRSST	RRSTS	RSRST
N=6	RRRRST	RRRSRT	RRRSST	RRRSTS	RRSRRT
	RRRSRT	RRSRTS	RRSRRT	RRSRTT	RRSTST
	RRSTTS	RRSRST	RSRTST		
N=7	RRRRRST	RRRRSRT	RRRRSST	RRRRSTS	RRRSRRT
	RRRSRST	RRRSRTS	RRRSRTT	RRRSSTT	RRRSSTS
	RRRSSTT	RRRSSTT	RRRSSTT	RRRSSTT	RRRSRTT
	RRRSRTT	RRRSSTT	RRRSRTT	RRRSRTT	RRSRTS
	RRSRRT	RRSRRT	RRSRRT	RRSRRT	RRSRRT
	RRSRRT	RRSRRT	RRSRRT	RRSRRT	RRSRRT
	RRSRRT	RRSRRT	RRSRRT	RRSRRT	RRSRRT
	RRSRRT	RRSRRT	RRSRRT	RRSRRT	RRSRRT
	RRSRRT	RRSRRT	RRSRRT	RRSRRT	RRSRRT
	RRSRRT	RRSRRT	RRSRRT	RRSRRT	RRSRRT
	RRSRRT	RRSRRT	RRSRRT	RRSRRT	RRSRRT
N=8	RRRRRST	RRRRSRT	RRRRSST	RRRRSTS	RRRRSRRT
	RRRRSRT	RRRRSRTS	RRRRSRTT	RRRRSSTT	RRRRSTS
	RRRRSSTT	RRRRSSTT	RRRRSSTT	RRRRSSTT	RRRRSRTT
	RRRSRRTS	RRRSRRTT	RRRSRRT	RRRSRRT	RRRSRRTS
	RRRSRTT	RRRSRTS	RRRSRTS	RRRSRTT	RRRSRTS
	RRRSRTT	RRRSRTS	RRRSRTS	RRRSRTT	RRRSRTS
	RRRSSTSS	RRRSSTT	RRRSSTT	RRRSSTT	RRRSRTS
	RRRSSTT	RRRSSTT	RRRSSTT	RRRSSTT	RRRSSTT
	RRRSSTT	RRRSSTT	RRRSSTT	RRRSSTT	RRRSSTT
	RRRSSTT	RRRSSTT	RRRSSTT	RRRSSTT	RRRSSTT
	RRRSSTT	RRRSSTT	RRRSSTT	RRRSSTT	RRRSSTT
	RRRSSTT	RRRSSTT	RRRSSTT	RRRSSTT	RRRSSTT
	RRRSSTT	RRRSSTT	RRRSSTT	RRRSSTT	RRRSSTT
	RRRSSTT	RRRSSTT	RRRSSTT	RRRSSTT	RRRSSTT
	RRRSSTT	RRRSSTT	RRRSSTT	RRRSSTT	RRRSSTT
	RRRSSTT	RRRSSTT	RRRSSTT	RRRSSTT	RRRSSTT
	RRRSSTT	RRRSSTT	RRRSSTT	RRRSSTT	RRRSSTT
	RRRSSTT	RRRSSTT	RRRSSTT	RRRSSTT	RRRSSTT
	RRRSSTT	RRRSSTT	RRRSSTT	RRRSSTT	RRRSSTT
	RRRSSTT	RRRSSTT	RRRSSTT	RRRSSTT	RRRSSTT
	RRRSSTT	RRRSSTT	RRRSSTT	RRRSSTT	RRRSSTT
	RRRSSTT	RRRSSTT	RRRSSTT	RRRSSTT	RRRSSTT
	RRRSSTT	RRRSSTT	RRRSSTT	RRRSSTT	RRRSSTT
	RRRSSTT	RRRSSTT	RRRSSTT	RRRSSTT	RRRSSTT
N=9	RRRRRST	RRRRSRT	RRRRSST	RRRRSTS	RRRRSRRT
	RRRRSRT	RRRRSRTS	RRRRSRTT	RRRRSSTT	RRRRSTS















TABLE G. IRREDUCIBLE SEQUENCES FOR AN ALPHABET OF FIVE LETTERS: R,S,T,U,V.

N=5	RSTUV				
N=6	RRSTUV	RSRTUV	RSTRUV		
N=7	RRRSTUV	RRSKTUV	RRSSTUV	RRSTRUV	RRSTSUV
	RRSTTUV	RRSTUSV	RRSTUTV	RRSTUVS	RSRSTUV
	RSRTRUV	RSRTSUV	RSRTUTV	RSRTUVT	RSTRSUV
	RSTRUSV				
N=8	RRRRSTUV	RRRSRTUV	RRRSSTUV	RRRSTRUV	RRRSTSUV
	RRRRSTTUV	RRRRSTUSV	RRRRSTUTV	RRRRSTUVS	RRSRRTUV
	RRRSRTUV	RRSRTRUV	RRRSRTSUV	RRSRRTTUV	RRSRTRUV
	RRRSRTUSV	RRSRRTUV	RRSRRTUUV	RRSRRTUVS	RRSRRTUVT
	RRSRRTUVU	RRSRRTUVV	RRSRRTUV	RRSRSTUV	RRSRSTUV
	RRRSSTUSV	RRRSSTTUV	RRRSSTUV	RRRSSTUV	RRRSSTUV
	RRRSSTUVS	RRRSSTUVT	RRRSSTUVU	RRRSSTUVV	RRSRTRUV
	RRRSRTSUV	RRRSRTTUV	RRRSRTUSV	RRRSRTUTV	RRSRTRUV
	RRRSRTUVS	RRRSRTUVT	RRRSRTUVU	RRRSRTUVV	RRRSRTUV
	RRRSRTUSV	RRRSRTUTV	RRRSRTUV	RRRSRTUVS	RRRSRTUVT
	RRRSRTUVU	RRRSRTUVV	RRRSRTUV	RRRSRTUVS	RRRSRTUVT
	RRRSSTUSV	RRRSSTTUV	RRRSSTUV	RRRSSTUV	RRRSSTUV
	RRRSSTUVS	RRRSSTUVT	RRRSSTUVU	RRRSSTUVV	RRSRTRUV
	RRRSRTSUV	RRRSRTTUV	RRRSRTUSV	RRRSRTUTV	RRSRTRUV
	RRRSRTUVS	RRRSRTUVT	RRRSRTUVU	RRRSRTUVV	RRRSRTUV
	RRRSRTUSV	RRRSRTUTV	RRRSRTUV	RRRSRTUVS	RRRSRTUVT
	RRRSRTUVU	RRRSRTUVV	RRRSRTUV	RRRSRTUVS	RRRSRTUVT
	RRRSSTUSV	RRRSSTTUV	RRRSSTUV	RRRSSTUV	RRRSSTUV
	RRRSSTUVS	RRRSSTUVT	RRRSSTUVU	RRRSSTUVV	RRSRTRUV
	RRRSRTSUV	RRRSRTTUV	RRRSRTUSV	RRRSRTUTV	RRSRTRUV
	RRRSRTUVS	RRRSRTUVT	RRRSRTUVU	RRRSRTUVV	RRRSRTUV
	RRRSRTUSV	RRRSRTUTV	RRRSRTUV	RRRSRTUVS	RRRSRTUVT
	RRRSRTUVU	RRRSRTUVV	RRRSRTUV	RRRSRTUVS	RRRSRTUVT
	RRRSSTUSV	RRRSSTTUV	RRRSSTUV	RRRSSTUV	RRRSSTUV
	RRRSSTUVS	RRRSSTUVT	RRRSSTUVU	RRRSSTUVV	RRSRTRUV
	RRRSRTSUV	RRRSRTTUV	RRRSRTUSV	RRRSRTUTV	RRSRTRUV
	RRRSRTUVS	RRRSRTUVT	RRRSRTUVU	RRRSRTUVV	RRRSRTUV
	RRRSRTUSV	RRRSRTUTV	RRRSRTUV	RRRSRTUVS	RRRSRTUVT
	RRRSRTUVU	RRRSRTUVV	RRRSRTUV	RRRSRTUVS	RRRSRTUVT
	RRRSSTUSV	RRRSSTTUV	RRRSSTUV	RRRSSTUV	RRRSSTUV
	RRRSSTUVS	RRRSSTUVT	RRRSSTUVU	RRRSSTUVV	RRSRTRUV
	RRRSRTSUV	RRRSRTTUV	RRRSRTUSV	RRRSRTUTV	RRSRTRUV
	RRRSRTUVS	RRRSRTUVT	RRRSRTUVU	RRRSRTUVV	RRRSRTUV
	RRRSRTUSV	RRRSRTUTV	RRRSRTUV	RRRSRTUVS	RRRSRTUVT
	RRRSRTUVU	RRRSRTUVV	RRRSRTUV	RRRSRTUVS	RRRSRTUVT
	RRRSSTUSV	RRRSSTTUV	RRRSSTUV	RRRSSTUV	RRRSSTUV
	RRRSSTUVS	RRRSSTUVT	RRRSSTUVU	RRRSSTUVV	RRSRTRUV
	RRRSRTSUV	RRRSRTTUV	RRRSRTUSV	RRRSRTUTV	RRSRTRUV
	RRRSRTUVS	RRRSRTUVT	RRRSRTUVU	RRRSRTUVV	RRRSRTUV
	RRRSRTUSV	RRRSRTUTV	RRRSRTUV	RRRSRTUVS	RRRSRTUVT
	RRRSRTUVU	RRRSRTUVV	RRRSRTUV	RRRSRTUVS	RRRSRTUVT
	RRRSSTUSV	RRRSSTTUV	RRRSSTUV	RRRSSTUV	RRRSSTUV
	RRRSSTUVS	RRRSSTUVT	RRRSSTUVU	RRRSSTUVV	RRSRTRUV
	RRRSRTSUV	RRRSRTTUV	RRRSRTUSV	RRRSRTUTV	RRSRTRUV
	RRRSRTUVS	RRRSRTUVT	RRRSRTUVU	RRRSRTUVV	RRRSRTUV
	RRRSRTUSV	RRRSRTUTV	RRRSRTUV	RRRSRTUVS	RRRSRTUVT
	RRRSRTUVU	RRRSRTUVV	RRRSRTUV	RRRSRTUVS	RRRSRTUVT
	RRRSSTUSV	RRRSSTTUV	RRRSSTUV	RRRSSTUV	RRRSSTUV
	RRRSSTUVS	RRRSSTUVT	RRRSSTUVU	RRRSSTUVV	RRSRTRUV
	RRRSRTSUV	RRRSRTTUV	RRRSRTUSV	RRRSRTUTV	RRSRTRUV
	RRRSRTUVS	RRRSRTUVT	RRRSRTUVU	RRRSRTUVV	RRRSRTUV
	RRRSRTUSV	RRRSRTUTV	RRRSRTUV	RRRSRTUVS	RRRSRTUVT
	RRRSRTUVU	RRRSRTUVV	RRRSRTUV	RRRSRTUVS	RRRSRTUVT
	RRRSSTUSV	RRRSSTTUV	RRRSSTUV	RRRSSTUV	RRRSSTUV
	RRRSSTUVS	RRRSSTUVT	RRRSSTUVU	RRRSSTUVV	RRSRTRUV
	RRRSRTSUV	RRRSRTTUV	RRRSRTUSV	RRRSRTUTV	RRSRTRUV
	RRRSRTUVS	RRRSRTUVT	RRRSRTUVU	RRRSRTUVV	RRRSRTUV
	RRRSRTUSV	RRRSRTUTV	RRRSRTUV	RRRSRTUVS	RRRSRTUVT
	RRRSRTUVU	RRRSRTUVV	RRRSRTUV	RRRSRTUVS	RRRSRTUVT





KSRSTURVT	RSRSTURVU	RSRSTUTUV	RSRSTUTVT	RSRSTUTVU
RSRSTUVTU	RSRSTUVUT	RSRTRSUV	RSRTRSTUV	RSRTRSURV
RSRTRSUSV	RSRTRSUTV	RSKTRSUVT	RSRTRSUVU	RSRTRRSV
RSRTRUSTV	RSRTRUSUV	RSRTRUSVU	RSRTRUTSV	RSRTRUTUV
RSKTRUTVU	RSRTRSRSUV	RSRTRSTUV	RSRTRSUSV	RSRTRSUTV
RSRTRSUVT	RSRTRSUVU	RSRSTRUV	RSRSTUSV	RSRSTUVU
RSKTSURTV	RSRTSURUV	RSRTSURVT	RSRTSURVU	RSRTSUSTV
RSRTSUSVT	RSRTSUTSV	RSRTSUTUV	RSRTSUTVT	RSRTSUTVU
RSRTSUVST	RSRTSUVTU	RSRTSUVTV	RSRTSUVUT	RSRTURSTV
RSRTURSUV	RSRTURSVT	RSRTURSVU	RSRTURTUV	RSRTURTVT
RSRTURTVU	RSRTURVTU	RSRTURVTV	RSRTURVUT	RSRTUSTUV
RSRTUSTVU	RSRTUSUTV	RSRTUSVTU	RSRTUSVTV	RSRTUSVUT
RSRTUVTVU	RSTRSTRUV	RSTRSTUSV	RSTRSURSV	RSTRSURTV
RSTRSURVU	RSTRSUTVU	RSTRUSRTV	RSTRUSTUV	RSTRUSTVU
RSTRUSVTU	RSTRUTSUV	RSTRUTSVU	RSTURSTUV	

### References

1. Part I, A.T. Balaban and C. Artemi, *Math. Chem.*, submitted for publication ; The general problem and its solution were briefly stated in : A.T. Balaban, *Comput. Math. Applications*, 12B, 999 (1986) reprinted in "*Symmetry Unifying Human Understanding*" (ed. I. Hargittai), Pergamon Press, New York, 1986.
2. T. Alfrey Jr., J.J. Bohrer and H.F. Mark, "*Copolymerization*", Interscience, New York, 1952 ; G.E. Ham (ed.), "*Copolymerization*", Interscience, New York, 1964.
3. F. Harary, in "*Graph Theory and Theoretical Physics*" (ed. F. Harary), chapter 1, Academic Press, London, 1967.
4. G. Polya, *Acta Math.*, 68, 145 (1937).
5. A.T. Balaban, *Studii Cercetari Chim. Acad. RPR.*, 7, 257 (1959).
6. A.T. Balaban, D. Farcasiu and F. Harary, *J. Labelled Comp.*, 6, 211 (1970).
7. A.T. Balaban and F. Harary, *Rev. Roum. Chim.*, 12, 1511 (1967).
8. A.T. Balaban and C. Artemi, submitted for publication.
9. N.G. de Bruijn, *Nederl. Akad. Wetenschap. Proc. Ser. A*, 62, *Indagationes Math.*, 21, 59 (1959).