# Enumeration of Isomers of Polyethers

JIANJI WANG*

Department of Chemistry, Inner Mongolia Normal University, Huhehaote 010022, China

FANGZHI GU

Computer Department, Inner Mongolia Normal University, Huhehaote 010022, China

Received May 29, 1991

In this paper, we report an alternate recurrence method to enumerate the isomers of polyethers containing an arbitrary number of oxygen atoms. Functions determining the number of polyether isomers are described, and some numerical results are tabulated.

## INTRODUCTION

Many enumerations have been reported for hydrocarbons and substituted hydrocarbons[1-4] but enumeration of the isomers of ethers that contain more than two oxygens has not yet been reported. We present three pairs of alternate recurrence formulas for enumerating the isomers of two kinds of alkyls and derive functions for the generation of constitutions, configurations, and achiral configurations of polyethers using Pólya's theorem.

## 1. DEFINITIONS

**1.1. Polyether.** An acyclic saturated ether can contain an arbitrary number of C-O-C units in which no carbon can be adjacent to two or more oxygen atoms. This is called a polyether. Its molecular formula is $C_iH_{2i+2}O_j$, where $j = 0, 1, 2, \ldots$. In this paper, "polyethers" include alkanes and ethers.

A polyether can be represented by a quartic tree consisting of two different kinds of points: one had degree 1-4 and is colored black, representing carbon. The other only has degree 2 and is colored red, representing oxygen. Thus, they can be enumerated by the method of counting trees.[5,6]

**1.2. First Kind of Alkyl-Alkyl(I).** An alkyl(I) is a part of a polyether. It has a root (cut point), the root has a cut bond, and the cut bond joins the root and a carbon atom in the polyether. Except for the cut bond, only one of the other three bonds adjacent to the root can (not must) be adjacent to an oxygen atom.

**1.3. Second Kind of Alkyl-Alkyl(II).** An alkyl(II) may be a part of a polyether. It also has a root (cut point), the root has a cut bond, and the cut bond joins the root and an oxygen atom in a polyether. Except for the cut bond, the other three bonds adjacent to the root may not be adjacent to any other oxygen atom.

In this paper, an alkyl(I) is represented by R and an alkyl(II) by R′, as shown in Figure 1. An alkyl(II) and an oxygen atom can be joined by its cut bond to form an alkoxy group (R′-O-), but an alkyl(I) cannot.

## 2. ENUMERATION

**2.1.** Let $A(x,y)$ be the generating function of constitutions of (I) and $A'(x,y)$ be the generating function of constitutions of (II). These functions obey the following alternate recurrence relations:

$$A'(x,y) = \sum_{i=0}^{\infty}\sum_{j=0}^{k} a'_{ij}x^iy^j = \frac{1}{6}x[A^3(x,y) + 3A(x,y)A(x^2,y^2) + 2A(x^3,y^3)] \quad (1)$$

and

$$A(x,y) = \sum_{i=0}^{\infty}\sum_{j=0}^{l} a_{ij}x^iy^j = 1 + A'(x,y) + \frac{1}{2}x[yA'(x,y)A^2(x,y) + yA'(x,y)A(x^2,y^2)] \quad (2)$$

where $a_{oo} = 1$. If $i$ is odd, $k = 1/2(i-1)$, and $l = 1/2(i-1)$. If $i$ is even, $k = i/2$ and $l = 1/2i - 1$. $a_{ij}$ is the number of constitutions of (I) containing $i$ carbons and $j$ oxygens, while $a'_{ij}$ is the number of constitutions of (II) containing $i$ carbons and $j$ oxygens.

**2.2.** Let $C(x,y)$ be the generating function of configurations of (I) and $C'(x,y)$ be the generating functions of configurations of (II). These functions obey the following alternate recurrence relations:

$$C'(x,y) = \sum_{i=0}^{\infty}\sum_{j=0}^{k} c'_{ij}x^iy^j = \frac{1}{3}x[C^3(x,y) + 2C(x^3,y^3)] \quad (3)$$

and

$$C(x,y) = \sum_{i=0}^{\infty}\sum_{j=0}^{l} c_{ij}x^iy^j = 1 + C'(x,y) + xyC'(x,y)C^2(x,y) \quad (4)$$

where $c_{oo} = 1$, $c_{ij}$ is the number of configurations of (I) containing $i$ carbons and $j$ oxygens, and $c'_{ij}$ is the number of configurations of (II) with $i$ carbon and $j$ oxygens.

**2.3.** Let $E(x,y)$ be the generating function of achiral configurations of (I) and $E'(x,y)$ be the generating function of achiral configurations of (II). These functions obey the following alternate recurrence relations:

$$E'(x,y) = \sum_{i=0}^{\infty}\sum_{j=0}^{k} e'_{ij}x^iy^j = xE(x,y)C(x^2,y^2) \quad (5)$$

and

$$E(x,y) = \sum_{i=0}^{\infty}\sum_{j=0}^{l} e_{ij}x^iy^j = 1 + E'(x,y) + xyE'(x,y)C(x^2,y^2) \quad (6)$$

where $e'_{oo} = 1$, $e_{ij}$ is the number of achiral configurations of (I) containing $i$ carbons and $j$ oxygens, and $e'_{ij}$ is the number of achiral configurations of (II) with $i$ carbons and $j$ oxygens.

**2.4.** Let $B(x,y)$ be the generating function of constitutions of polyethers. We use the relation $p - q + s = 1$[3], where $q$ is not the number of equivalence classes of C-C bonds, but the number of equivalence classes of C-C bonds and C-O-C

0095-2338/91/1631-0552$02.50/0 

**Table I.** Numbers of Constitutions of Polyethers ($b_{ij}$)

| | i | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| j | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
| 0 | 1 | 1 | 1 | 2 | 3 | 5 | 9 | 18 | 35 | 75 | 159 | 355 | 802 | 1858 |
| 1 | | 1 | 1 | 3 | 6 | 15 | 33 | 82 | 194 | 482 | 1188 | 2988 | 7528 | 19181 |
| 2 | | | | 1 | 3 | 12 | 37 | 119 | 355 | 1062 | 3093 | 8974 | 25736 | 73520 |
| 3 | | | | | | 2 | 10 | 53 | 221 | 887 | 3259 | 11552 | 39345 | 130712 |
| 4 | | | | | | | | 5 | 38 | 252 | 1333 | 6347 | 27534 | 112341 |
| 5 | | | | | | | | | | 14 | 151 | 1239 | 7886 | 43746 |
| 6 | | | | | | | | | | | | 45 | 648 | 6252 |
| 7 | | | | | | | | | | | | | | 161 |

**Table II.** Numbers of Configurations of Polyethers ($d_{ij}$)

| | i | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| j | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
| 0 | 1 | 1 | 1 | 2 | 3 | 5 | 11 | 24 | 55 | 136 | 345 | 900 | 2412 | 6563 |
| 1 | | 1 | 1 | 3 | 7 | 19 | 49 | 139 | 384 | 1104 | 3180 | 9306 | 27390 | 81373 |
| 2 | | | | 1 | 4 | 18 | 66 | 242 | 846 | 2919 | 9897 | 33271 | 110937 | 368028 |
| 3 | | | | | | 2 | 17 | 112 | 578 | 2739 | 11885 | 49171 | 195434 | 755408 |
| 4 | | | | | | | | 8 | 90 | 771 | 5085 | 29088 | 150132 | 721395 |
| 5 | | | | | | | | | | 29 | 505 | 5410 | 43350 | 292570 |
| 6 | | | | | | | | | | | | 140 | 3056 | 38859 |
| 7 | | | | | | | | | | | | | | 720 |

**Table III.** Numbers of Achiral Configurations of Polyethers ($g_{ij}$)

| | i | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| j | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
| 0 | 1 | 1 | 1 | 2 | 3 | 5 | 7 | 14 | 21 | 40 | 61 | 118 | 186 | 355 |
| 1 | | 1 | 1 | 3 | 5 | 11 | 19 | 39 | 68 | 134 | 238 | 458 | 818 | 1553 |
| 2 | | | | 1 | 2 | 8 | 16 | 44 | 86 | 205 | 393 | 871 | 1651 | 3500 |
| 3 | | | | | | 2 | 5 | 22 | 52 | 155 | 347 | 881 | 1894 | 4410 |
| 4 | | | | | | | | 4 | 12 | 59 | 153 | 516 | 1242 | 3481 |
| 5 | | | | | | | | | | 9 | 29 | 160 | 458 | 1640 |
| 6 | | | | | | | | | | | | 20 | 70 | 429 |
| 7 | | | | | | | | | | | | | | 48 |

units—a C-O-C unit being treated as a labeled C-C bond. In this way, we get the function:

$$B(x,y) = \sum_{i=1}^{\infty}\sum_{j=0}^{k} b_{ij}x^i y^j = \frac{1}{24}x[A^4(x,y) + 4yA'(x,y)A^3(x,y) + 6A^2(x,y)A(x^2,y^2) + 12yA'(x,y)A(x,y)A(x^2,y^2) + 3A^2(x^2,y^2) + 8A(x,y)A(x^3,y^3) + 8yA'(x,y)A(x^3,y^3) + 6A(x^4,y^4)] - \frac{1}{2}[A(x,y) - 1]^2 + \frac{1}{2}[A(x^2,y^2) - 1] - \frac{1}{2}y[A'(x,y) - 1]^2 + \frac{1}{2}y[A'(x^2,y^2) - 1] \quad (7)$$

where $b_{ij}$ is the number of constitutions of polyethers. Some results from these relationships are shown in Table I.

**2.5.** Let $D(x,y)$ be the generating function of constitutions of polyethers. Then

$$D(x,y) = \sum_{i=1}^{\infty}\sum_{j=0}^{k} d_{ij}x^i y^j = \frac{1}{12}x[C^4(x,y) + 4yC'(x,y)C^3(x,y) + 3C^2(x^2,y^2) + 8C(x,y)C(x^3,y^3) + 8yC'(x,y)C(x^3,y^3)] - \frac{1}{2}[C(x,y) - 1]^2 + \frac{1}{2}[C(x^2,y^2) - 1] - \frac{1}{2}y[C'(x,y) - 1]^2 + \frac{1}{2}y[C'(x^2,y^2) - 1] \quad (8)$$

where $d_{ij}$ is the number of configurations of polyethers. Some results from this relation are given in Table II.

**2.6.** Let $G(x,y)$ be the generating function of achiral configurations of polyethers and $D'(x,y)$ be the generating function of configurations containing one labeled carbon atom. Let $P(x,y)$ be the generating function of achiral configurations of polyethers containing one labeled carbon and $D_2(x,y)$ be the generating function of chiral configurations of polyethers containing one labeled carbon atom. Then


R, alkyl(I) R', alkyl(II)

**Figure 1.** An alkyl(I) and an alkyl(II).

$$D'(x,y) = P(x,y) + D_2(x,y) \quad (9)$$

Let

$$D''(x,y) = P(x,y) + \frac{1}{2}D_2(x,y) \quad (10)$$

Then

$$D'(x,y) = \frac{1}{12}x[C^4(x,y) + 4yC'(x,y)C^3(x,y) + 3C^2(x^2,y^2) + 8C(x,y)C(x^3,y^3) + 8yC'(x,y)C(x^3,y^3)] \quad (11)$$

$$D''(x,y) = \frac{1}{24}x[C^4(x,y) + 4yC'(x,y)C^3(x,y) + 3C^2(x^2,y^2) + 8C(x,y)C(x^3,y^3) + 8yC'(x,y)C(x^3,y^3) + 6E^2(x,y)C(x^2,y^2) + 12yE'(x,y)E(x,y)C(x^2,y^2) + 6c(x^4,y^4)] \quad (12)$$

and

$$P(x,y) = 2D''(x,y) - D'(x,y) = \frac{1}{2}xC(x^4,y^4) + \frac{1}{2}xE^2(x,y)C(x^2,y^2) + xyE'(x,y)E(x,y)C(x^2,y^2) \quad (13)$$

Care must be taken that $P(x,y)$ does not contain meso ($R_+$-$R_-$ or $R'_+$-O-$R'_-$) forms.

Let $Q(x,y)$ be the generating function of achiral configurations of polyethers containing a labeled C-C bond or a C-O-C unit, in which meso forms are not included. Meso forms of polyethers containing one labeled C-C bond or a C-O-C unit can be achiral. Therefore

$$Q(x,y) = \sum_{i=2}^{\infty}\sum_{j=0}^{k} q_{ij}x^iy^j = \frac{1}{2}[E(x,y) - 1]^2 + \frac{1}{2}[E(x^2,y^2) - 1] + \frac{1}{2}y[E'(x,y) - 1]^2 + \frac{1}{2}y[E'(x^2,y^2) - 1] \quad (14)$$

Let $S(x,y)$ be the generating function of achiral configurations of R-R forms and of R'-O-R' forms in which the meso forms are also not included. So

$$S(x,y) = [E(x^2,y^2) - 1] + y[E'(x^2,y^2) - 1] \quad (15)$$

Meso forms however are achiral and must be counted. Let $M(x,y)$ be the generating function of the meso forms of polyethers. Then

$$M(x,y) = \frac{1}{2}[C(x^2,y^2) - 1] - \frac{1}{2}[E(x^2,y^2) - 1] + \frac{1}{2}y[C'(x^2,y^2) - 1] - \frac{1}{2}y[E'(x^2,y^2) - 1] \quad (16)$$

and

$$G(x,y) = \sum_{i=1}^{\infty}\sum_{j=0}^{k} g_{ij}x^iy^j = P(x,y) - Q(x,y) + S(x,y) + M(x,y) = \frac{1}{2}xC(x^4,y^4) + \frac{1}{2}xE^2(x,y)C(x^2,y^2) + xyE'(x,y)E(x,y)C(x^2,y^2) - \frac{1}{2}[E(x,y) - 1]^2 + \frac{1}{2}[C(x^2,y^2) - 1] - \frac{1}{2}y[E'(x,y) - 1]^2 + \frac{1}{2}y[C'(x^2,y^2) - 1] \quad (17)$$

where $g_{ij}$ is the number of achiral configurations of polyethers. Some results from this algorithm are given in Table III.

## REFERENCES AND NOTES


(1) Slanina, Z. *Contemporary Theory of Chemical Isomerism*; D. Reidel Publishing Co.: Dordrecht, 1986.
(2) Fujita, S. *J. Math. Chem.* **1990**, *5*, 99-121.
(3) Fujita, S. *Theor. Chem. Acta* **1990**, *77*, 307; **1989**, *76*, 247.
(4) Muller, W. R. *J. Comput. Chem.* **1990**, *11*, 223.
(5) Wang, J.; Wang, Q. *Tetrahedron* **1991**, *47*, 2969.
(6) Balaban, A. T. *Chemical Applications of Graph Theory*; Academic Press: London, 1976; pp 25-62.


# A Canonical Representation of Multistep Reactions


RAÚL E. VALDÉS-PÉREZ

School of Computer Science, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213


Received May 28, 1991


A canonical representation of multistep reactions or reaction networks is introduced. This representation has been applied to the efficient generation of reaction pathways for computer-assisted elucidation. Other potential applications are to chemical information systems and to intelligible depiction of pathways in publications. An algorithm is presented that canonicalizes a multistep reaction from a specification of the starting materials.


## INTRODUCTION

An ideal canonical representation of a multistep reaction or network would have the following consequences. First, it would facilitate information retrieval within chemical databases that store information on multistep reactions. Second, it would enhance the intelligibility of multistep reactions as they are depicted in publications. Third, it would find application in computer algorithms that generate reaction networks for theoretical investigations or for modeling a chemical reaction.

This paper proposes a canonical network representation that can serve the three purposes above. The canon was discovered in the context of the design of a program to generate pathway hypotheses for building models of a chemical reaction. In this application there was a critical need to avoid redundant generation of multiple pathways that are identical except for permutations among steps and among reactants and products within single steps.

A desirable feature of the canon is that a given reaction network can easily be rearranged to follow the canon without rewriting the network in another representation, e.g., as a matrix. A second feature is that canonicalizing a network is conveniently done by hand. Lastly, canonicalized networks convey well the notion of flow, since steps appear in the network "soon" after their reactants become available; this will become clearer below.

**Related Work.** Brandt and von Scholley[1] describe a canonical numbering of reaction matrices in which a matrix represents the changes in the covalent bonds of molecules that occur as a result of a single reaction step. Fujita[1] likewise describes a method to represent the molecular-structural changes induced by individual reactions. However, our search in the literature for canonical representations of multistep reactions has not turned up any precursors to this work.

## A CANONICAL REPRESENTATION

The canon consists of three conditions that an ordered reaction network must fulfill; these are discussed in what follows. For convenience, we assume that a given network is to be represented on a page, one step per line, so that it makes sense to speak of one step being higher or lower than another. For now, we also assume that the starting materials of the reaction network are known; later in the paper we describe an algorithm to infer a set of starting materials from a given network. Finally, we assume that each reaction step is irreversible; the handling of reversible steps will be discussed separately later.

0095-2338/91/1631-0554$02.50/0 