

EIN WÖRTERBUCH DER CHEMIE - DIE SEQUENZLÄNGEN-  
VERTEILUNG. THEORIE, RECHENVERFAHREN, ANWENDUNG.

I. Moțoc

Chemie-Forschungszentrum  
Bul.Mihai Viteazu 24,  
1900 Timișoara, Rumänien

( Received : Revised Version September 1980)

Synopsis

The present paper proposes to consider the copolymer chain as a sentence with sequences  $(-M_I-)_{L>1}$  ( $M_I$  are the mers of the chain) as words and  $(-M_I-)_{1}$  as punctations. Using this chemical language, one can "read" from the macromolecule the reactivity of monomers  $M_I$ . For the case of binary copolymerization we developed a computing procedure to determine  $r_1$ ,  $r_2$  ratios using qualitative informations concerning sequence distribution. The program MEMORY-7 (which implements the discussed computing procedure) is presented and applied to the methyl methacrylate/chloroprene copolymers.

1. Einleitung

Die Arbeit stellt eine informationelle Erläuterung der Begriffe Makromolekül und Sequenzlängenverteilung vor.

Es wird gezeigt, dass nicht die Zusammensetzung der Copolymeren, sondern, notwendigerweise, die Sequenzlängenverteilung ein Wörterbuch der Chemie darstellt. In diesem Sinne wurde ein Rechenverfahren für den Fall der binären irreversiblen Copolymerisation mit Endeffekt erarbeitet (dieses Verfahren kann ohne Schwierigkeit auch an eine andere Copolymerisationskinetik angepasst werden), womit die Copolymerisationskonstanten  $r_1$  und  $r_2$  aufgrund von qualitativen Angaben bezüglich der Sequenzlängenverteilung bestimmt

werden<sup>3</sup>. Das vorgeschlagene Verfahren wird auf den numerischen Rechner mittels des MEMORY-7 Programms implementiert. Dieses Programm löst das umgekehrte des vom MEMORY-3 Programm gelösten Problems, ein Programm, welches die binäre irreversible Copolymerisation mit Endeffekt simuliert<sup>1,2,5</sup>.

Der in dieser Arbeit gewählte Gesichtspunkt wird häufig in der molekularen Biologie<sup>4</sup> verwendet und wurde neuestens auch bei der Behandlung der Konformation der Biopolymeren<sup>6,7</sup> berücksichtigt.

## 2. Ein Wörterbuch der Chemie : die Sequenzlängenverteilung

Da offenbar die Reaktivität der Monomeren die Sequenzlängenverteilung bedingt (sowohl als Länge als auch als gegenseitige Anordnung), ist folgende Erläuterung selbstverständlich (konkretisiert am Fall der linearen Copolymeren) :

1) Die Kette des Copolymeren wird als Satz, P, betrachtet.  
2) Folglich, muss es ein logisches Set von Wörtern geben. Als Wörter werden die Sequenzen (Blocks)  $(-M_I-)_I$ ,  $L_I \geq 2$ ,  $M_I$  sind Meren, betrachtet.

Die Gesamtheit aller Wörter bildet das Wörterbuch DC der Chemie.

3) Die Sequenzen  $(-M_I-)_I$ ,  $L_I = 1$ , werden als Satzzeichen betrachtet.

4) Die grammatikalischen Regeln, RG, für die P einen Sinn hat, können mittels der Beziehung zwischen der Struktur (des Monomeren) und der Reaktivität festgesetzt werden.

5) DC und RG bilden die chemische Sprache, LC, in welcher P einen Sinn hat, oder, mit anderen Worten, LC stellt die Sprache dar, in der das betrachtete Makromolekül die Reaktivität der Monomeren aus der chemischen Reaktion, aus der es hervorgegangen ist, "memoriert" (gespeichert) hat.

Abbildung 1 systematisiert die vorhin erläuterten Begriffe.

Die Gültigkeit des in Abbildung 1 gegebenen Schemas kann durch ein Rechenverfahren, welches die vom Makromolekül

gespeicherten Informationen effektiv "entziffert", geprüft werden. Im nächsten Abschnitt der Arbeit wird ein derartiges Rechenverfahren vorgeschlagen, sowie auch dessen Implementierung durch das MEMORY-7 Programm diskutiert. Die vorgeschlagene Methode beruht auf der Monte Carlo Technik.

### 3. Rechenverfahren, MEMORY-7 Programm

Da der vorgeschlagene Algorithmus verhältnismässig komplex ist, ist es nötig, zugleich sowohl die angenommene mathematische Methode sowie auch die Art und Weise in der diese auf den Rechner implementiert wurde vorzustellen. Die Charakteristika des Programms sind :

Benennung : MEMORY-7  
Sprache : FORTRAN IV  
Hardware : IRIS 50, FELIX C-256, IBM etc.  
Quelle : 180 Lochkarten  
Speicher : 60 KØ  
Berechnungszeit : zwischen 20' und 2 h, je nach dem Polymerisationsgrad und der spezifizierten Zusammensetzung des Copolymeren.

Das logische Schema des Programms ist in Abbildung 2 gegeben.

Die vom Programm beanspruchten Eingangsdaten sind folgende : N - der Polymerisationsgrad ; NA und NB - die Anzahl der Meren  $M_1$  und  $M_2$  in der Kette (offensichtlich ist  $NA + NB = N$ ) ;  $C = [M_2]/[M_1]$ , wobei  $[M_1]$  und  $[M_2]$  die Konzentrationen der Monomeren im Reaktionsgemisch sind. Die Daten werden durch einen DATA-Block eingeführt.

Die Generierung der aleatorischen Zahlen wird mittels der Subroutine ALEAT verwirklicht. Um die Performanz des Programms im Verhältnis zur Berechnungszeit zu steigern, wird die Tabelle T eingeführt, welche für den restlichen Rechenvorgang zum sekundären Generatoren von aleatorischen Zahlen wird. In T wird die beste Sequenz von 1000 aleatorischen Zahlen, die durch einen Vorsprung um 6000 Zahlen des

Generatoren ALEAT erhalten wurde, festgehalten.

Die Werte für  $r_1$  und  $r_2$  werden folgendermassen bestimmt :

1) Es wird ein Set von M  $r_1$ -Werten, die mit der Beziehung :

$$r_1 = (NA + C)/(1 - NA)$$

berechnet wurden, getestet. Für jeden betrachteten  $r_1$ -Wert wird an die Sektion für die Bestimmung der  $r_1$ -Werte appelliert.

2) Für die M  $r_1$ -Werte können höchstens M verschiedene  $r_2$ -Werte erhalten werden.  $r_2$  wird mittels der Beziehung :

$$r_2 = (1 - B)/(C + B)$$

bestimmt, wobei B-J/N und J-LP, 2LP, ... Der ursprüngliche Wert des Schritts LP ist LP = 1.  $r_2$  wird für jeden der M betrachteten Werte von  $r_1$  berechnet.

Danach wird für jedes Wertepaar  $r_1$ ,  $r_2$  ein Makromolekül nach dem durch das MEMORY-3 Programm implementierten Verfahren<sup>1</sup> generiert. Um die Effizienz des Programms zu steigern, wurden die beiden folgenden (übrigens selbstverständlichen) Vorgänge eingeführt :

1) Wenn in der Testzeit von einem  $r_2$ -Wert die Anzahl der in die Kette eingegliederten Meren  $M_1$  den Wert NA überschreitet, wird der Rechenvorgang unterbrochen und ein neuer  $r_2$ -Wert bestimmt.

2) Wenn für zwei aufeinanderfolgende Werte  $r_2'$  und  $r_2''$  der vorgeschriebene NA Wert nicht erreicht wird, so werden auch die übrigen  $r_2$ -Werte unterlassen und es wird zu einem neuen  $r_1$ -Wert übergegangen.

Der Wählschritt LP ist ursprünglich LP = 1, aber nachher wird er gemäss folgender Beziehung berechnet :

$$LP = (NA - NA_{old})/3$$

wobei  $NA_{old}$  die Anzahl der Meren  $M_1$ , die in die Kette des Copolymeren, unter Anwendung des dem vorhergegangenen

Schrittwertes entsprechenden  $r_2$ -Wertes, eingegliedert wurden.

Das Programm gibt die Liste der Werte der Konstanten  $r_1$  und  $r_2$ , die zu Copolymeren mit vorgeschriebener Zusammensetzung führen, sowie auch die entsprechenden Sequenzlängenverteilung in den Copolymeren.

Es wird an einer verbesserten Version des MEMORY-7 Programms gearbeitet, welche, durch Anwendung einer Methode zur Beschleunigung der Konvergenz, eine erhebliche Verkürzung der Berechnungszeit<sup>8</sup> ermöglicht.

#### 4. Anwendung : Copolymeren Methylmethacrylat/Chloropren

Das vorgeschlagene Rechenverfahren wird am Beispiel der Copolymeren Methylmethacrylat ( $M_1$ )/Chloropren ( $M_2$ ) veranschaulicht. Die Daten betreffend die Zusammensetzung und Sequenzlängenverteilung entstammen einer Arbeit<sup>9</sup> von Ebdon.

Für das mit 85%  $M_1$  und 15%  $M_2$  im Reaktionsgemisch (Mol-%) erhaltene Copolymere erhält man eine Zusammensetzung von 40%  $M_1$  und 60%  $M_2$  (Mol-%). In der Annahme eines Polymerisationsgrades  $N = 1000$ , erhält man mit Hilfe des MEMORY-7 Programms folgende Ergebnisse :

Das erste Set von  $r_1$ ,  $r_2$ -Werten, die die Zusammensetzung des Copolymeren wiedergeben, enthält Werte geschaart um :

$$(A) r_1 = 0,005, r_2 = 2,774 \text{ (40,4\% } M_1 \text{ und 59,6\% } M_2 \text{)}.$$

Das zweite Set von  $r_1$ ,  $r_2$ -Werten schaart sich um :

$$(B) r_1 = 0,058, r_2 = 5,6 \text{ (39,5\% } M_1 \text{ und 60,5\% } M_2 \text{)}.$$

Die Sequenzlängenverteilung in den mit den beiden Sets von Copolymerisationskonstanten erhaltenen Copolymeren ist in Tabelle 1 gegeben. ( $k$  bezeichnet die Länge der Sequenz,  $n_I(k)$ ,  $I = 1,2$ , bezeichnet die Anzahl der Sequenzen  $-M_1-$  der Länge  $k$ .)

Durch eine qualitative Auswertung der Informationen betreffend<sup>9</sup> die Sequenzlängenverteilung (eine qualitative Auswertung dieser Informationen ist ausreichend, da im Be-

reich der Sequenzen sehr grosse Unterschiede zwischen den beiden Copolymeren sind) wird das Set (B) von Copolymerisationskonstanten gewählt. Die aus den Daten betreffend die Zusammensetzung bestimmten Werte sind<sup>10</sup> :

(C)  $r_1 = 0,08$ ,  $r_2 = 5,1$  (41,8%  $M_1$  und 58,2%  $M_2$ , durch die Monte Carlo Methode<sup>2</sup> berechnet). Die (berechnete) Sequenzlängenverteilung in den diesem Set von Copolymerisationskonstanten entsprechenden Copolymere ist in Tabelle 1 zusammengestellt.

Man beobachtet eine sehr gute Übereinstimmung der berechneten  $r_1$ ,  $r_2$  Werte mit den experimentell bestimmten  $r_1$ ,  $r_2$  Werten.

##### 5. Schlussfolgerungen

Mit einem performanten Programm kann das in dieser Arbeit vorgestellte Rechenverfahren zu einer nützlichen und zugleich einfachen Methode für die Bestimmung der Konstanten  $r_1$  und  $r_2$  werden, indem man ausschliesslich durch spektroskopische Methoden erhaltene Daten verwendet, ohne an Methoden der Kinetik zu appellieren. Oder, kann die Methode zur Überprüfung der durch die anderen Methoden erhaltenen Werte der Copolymerisationskonstanten verwendet werden.

Tabelle 1. Copolymeren Methylmethacrylat/Chloropren  
(Sequenzlängenverteilung).

| k   | A        |          | B        |          | C        |          |
|-----|----------|----------|----------|----------|----------|----------|
|     | $n_1(k)$ | $n_2(k)$ | $n_1(k)$ | $n_2(k)$ | $n_1(k)$ | $n_2(k)$ |
| 1.  | 379      | 254      | 224      | 145      | 211      | 146      |
| 2.  | 10       | 84       | 51       | 72       | 52       | 75       |
| 3.  | -        | 36       | 18       | 46       | 23       | 40       |
| 4.  | -        | 11       | 2        | 15       | 5        | 16       |
| 5.  | 1        | 3        | -        | 6        | -        | 6        |
| 6.  |          | -        | -        | 4        | -        | 5        |
| 7.  |          | 1        | 2        | 3        | 2        | 1        |
| 8.  |          |          |          | 1        |          | -        |
| 9.  |          |          |          | 1        |          | 1        |
| 10. |          |          |          | 1        |          | 1        |
| ... |          |          |          | -        |          | -        |
| 16. |          |          |          | 1        |          | 1        |

A, B und C beziehen sich auf die drei Sets von  $r_1, r_2$ -Werten, so wie im Text erläutert wird.

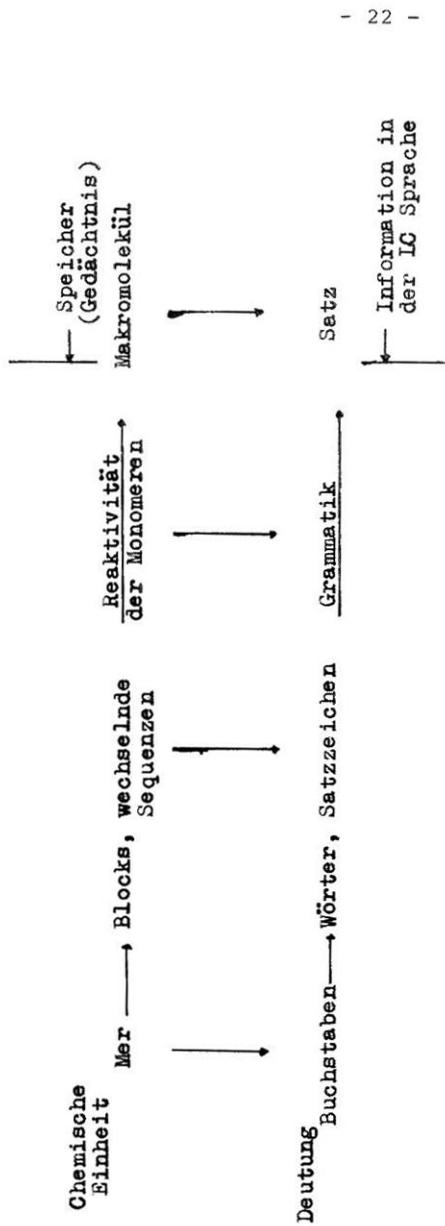


Abb.1. Die chemische Sprache IC.

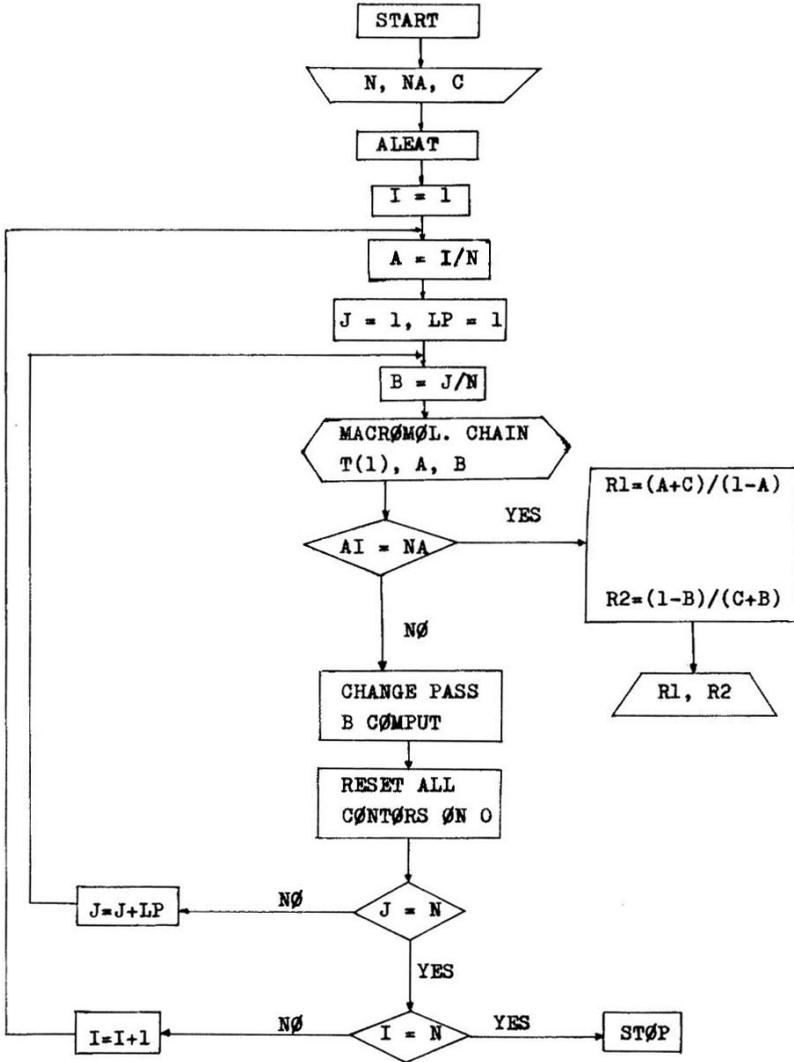


Abb.2. Logisches Schema des MEMORY-7 Programms.

Literaturverzeichnis

1. I.MOTOC, St.HOLBAN und C.CIUBOTARIU, J.Polym.Sci., 15, 1465 (1977).
2. I.MOTOC, St.HOLBAN und R.VANCEA, J.Polym.Sci., 16, 1601 (1978).
3. R.CHUJO, H.UBARA und A.NISHIOKA, Polym.J., 2, 670 (1972).
4. O.GUREL, Z.Naturforsch., 30, 295 (1975) (und die in der Arbeit angeführte Literatur).
5. I.MOTOC und I.MUSCUTARIU, J.Macromol.Sci.-Chem., in Druck.
6. J.A.DE BASETH und T.L.ISENHOUR, "Computer-Assisted Structure Elucidation", Ed.D.H.Smith, ACS Symposium Series No.54, 1977.
7. A.M.LIQUORI, "The Stereochemistry of Macromolecules", Ed.A.D.Ketley, Band 3, Dekker, New York, 1967.
8. I.MOTOC und St.HOLBAN, wird veröffentlicht.
9. J.R. EBDON, Polymer, 15, 782 (1974).
10. D.BRAUN, W.BRENDLEIN und G.MOTT, Eur.Polym.J., 9, 1007 (1973).