

strangablesemechanismus ist daher vielleicht als eine der Möglichkeiten nicht auszuschließen.

Eingegangen am 8. Juli 1972 [A 902a]

- [1] M. Calvin: *Chemical Evolution*. Clarendon Press, Oxford 1969.
[2] M. G. Rutten: *The Origin of Life*. Elsevier, Amsterdam, London, New York 1971.
[3] C. Ponnamperuma: *Exobiology*. North-Holland, Amsterdam, London 1972.
[4] R. Bucet u. C. Ponnamperuma: *Molecular Evolution I. Chemical Evolution and the Origin of Life*. North-Holland, Amsterdam, London 1971.
[5] F. H. C. Crick, J. Mol. Biol. 38, 367 (1968); L. J. Orgel, ibid. 38, 381 (1968).
[6] J. Monod: *Le hasard et la nécessité*. Edition du Senteil, Paris 1970; Zufall und Notwendigkeit. Piper, München 1971.
[7] J. v. Neumann: *Theory of self-reproducing automata*. University of Illinois Press, Urbana 1966.
[8] A. M. Turing, Mind 59, 433 (1950).
[9] W. A. Beyer, M. L. Stein u. S. M. Ulam: Los Alamos Report 1971, 4822.
[10] I. Prigogine in: *Proceedings of the Fifth Nobel Symposium*. Almqvist und Wiksell, Stockholm 1967.
[11] G. Oster, A. Perelson u. A. Katchalsky, Nature 234, 393 (1971).
[12] M. Eigen, Naturwissenschaften 58, 465 (1971).
[13] J. D. Watson: *Molecular Biology of the Gene*. 2. Aufl. Benjamin, New York 1970; E. Harbers: *Nucleinsäuren*. Thieme, Stuttgart 1969; G. Schreiber, Angew. Chem. 83, 645 (1971); Angew. Chem. internat. Edit. 10, 638 (1971).
[14] R. Naylor u. P. H. Gilham, Biochemistry 8, 2722 (1966).
[15] M. Eigen u. D. Pörschke, J. Mol. Biol. 53, 123 (1970); D. Pörschke u. M. Eigen, ibid. 62, 361 (1971); D. Pörschke, Biopolymers 10, 1989 (1971); D. Pörschke u. F. Eggers, Eur. J. Biochem. 26, 490 (1972).
[16] R. L. Baldwin, Accounts Chem. Res. 4, 265 (1971); I. E. Scheffler, E. L. Elson u. R. L. Baldwin, J. Mol. Biol. 36, 291 (1968).
[17] O. C. Uhlenbeck, J. Boller u. P. Doty, Nature 225, 508 (1970); J. B. Lewis u. P. Doty, ibid. 225, 510 (1970).

- [18] P. Debye, J. Phys. Chem. 53, 1 (1949); Ann. N. Y. Acad. Sci. 51, 575 (1949).
[19] F. Cramer, Progr. Nucleic Acid Res 11, 391 (1971).
[20] H. D. Försterling u. H. Kuhn: *Physikalische Chemie in Experimenten*. Verlag Chemie, Weinheim 1971; H. D. Försterling, H. Kuhn u. K. H. Tews, Angew. Chem. 84, 862 (1972); Angew. Chem. internat. Edit. 11, Heft 9 (1972).
[21] P. J. G. Butler u. A. Klug, Nature, New Biol. 229, 47 (1971); M. Nomura, Scientific American, Dezember 1969 und in F. O. Schmitt: *The Neurosciences. 2. Study Program*, Rockefeller University Press, New York 1970, S. 913; R. A. Garret u. H. G. Wittmann in T. B. Anfinsen: *Aspects of Protein Biosynthesis*. Academic Press, New York 1971.
[22] M. Paecht-Horowitz, J. Berger u. A. Katchalsky, Nature 228, 636 (1970); M. Paecht-Horowitz, [4], S. 245.
[23] S. W. Fox, Science 132, 200 (1960); Naturwissenschaften 56, 1 (1969).
[24] C. D. Laird, Chromosoma 32, 378 (1971).
[25] F. M. Perutz, Nature 219, 902 (1968).
[26] F. M. Perutz, Nature 228, 728 (1970).
[27] M. O. Dayhoff, [3], S. 266, besonders Tabelle 2, S. 287.
[28] E. Harbers, [13], S. 171.
[29] E. C. Cox u. C. Yanofsky, Proc. Natl. Acad. Sci. USA 58, 1895 (1967).
[30] E. L. Wollman, F. Jacob u. W. Hayes, Cold Spring Harbor Symp. Quant. Biol. 21, 141 (1964).
[30a] H. Dertinger u. H. Jung: *Molekulare Strahlenbiologie*. Springer, Berlin 1969.
[31] T. Ohta u. M. Kimura, Nature 233, 118 (1971); M. Kimura u. T. Ohta, J. Mol. Evolution 1, 1 (1971); M. Kimura, Proc. Natl. Acad. Sci. USA 63, 1181 (1969); J. L. King u. T. H. Jukes, Science 164, 788 (1969).
[32] T. Lanz u. G. Neuhäuser, Z. Anatomie Entwicklungsgesch. 123, 462 (1963).
[33] J. Langman: *Medizinische Embryologie*. Thieme, Stuttgart 1970.
[34] J. L. King u. T. H. Jukes, Science 164, 788 (1969), besonders S. 794.
[35] J. W. Fristom, Ann. u. Rev. Genet. 4, 325 (1970).
[36] H. Riv u. D. F. Kubai, Ann. u. Rev. Genet. 4, 263 (1970); D. Werner u. M. Petersen, unveröffentlicht.
[37] F. Crick, Nature 234, 25 (1971).
[38] M. Nei, Nature 221, 40 (1969).
[39] J. F. Pardon, M. H. F. Wilkins u. B. M. Richards, Nature 215, 508 (1967).

Computermodell zur Bildung selbstorganisierender Systeme

Von H. D. Försterling, H. Kuhn und K. H. Tews^[*]

Es wird ein Modell eines einfachen sich selbst organisierenden Systems untersucht. An dem Modell werden Mechanismen studiert, die für ein molekular-biologisches Verständnis der Evolution wichtig sind.

1. Einleitung

Im Folgenden wird das Verhalten einfacher selbstorganisierender Systeme^[1] durch ein Computermodell simuliert. Es wird ein System mit bereits vorhandenem genetischem Apparat betrachtet und vorausgesetzt, daß die Bildung von Proteinen an der Nucleinsäurematrix möglich ist. Es wird untersucht, wie die weitere Evolution eines solchen Systems verläuft.

[*] Dr. H. D. Försterling
Physikalisch-chemisches Institut der Universität
355 Marburg, Biegenstraße 12
Prof. Dr. H. Kuhn und Dr. K. H. Tews
Max-Planck-Institut für Biophysikalische Chemie
34 Göttingen-Nikolausberg, Postfach 968

Ein Protein, das vom System gebildet wird, kann in dreifacher Weise auf das System zurückwirken: es kann die Zeit τ_R , die im Mittel zwischen zwei aufeinanderfolgenden Replikationen vergeht, die Zeit τ_L , nach der das System im Mittel zerstört wird, oder die Wahrscheinlichkeit W , daß bei der Replikation ein Fehler auftritt, beeinflussen (Abb. 1). Dabei kann das Protein entweder direkt den Replikationsvorgang oder die Proteinsynthese begünstigen (Verkleinerung von τ_R und W) oder das System vor der Zerstörung schützen (Vergrößerung von τ_L). Eine Beeinflussung von τ_R , W und τ_L kann aber auch indirekt erfolgen: die gebildeten Proteine können beispielsweise das System nach außen abkapseln und dadurch die Wegdiffusion wichtiger Reaktionsteilnehmer verhindern, das System gegen äußere Eingriffe schützen oder die Synthese eines

für die Vervielfältigung des Systems nötigen Bausteines erleichtern. Tritt durch ein Protein, das im Lauf der Generationen durch Ablesefehler bei der Replikation der DNA entstanden ist, eine solche Erleichterung auf, so hat das

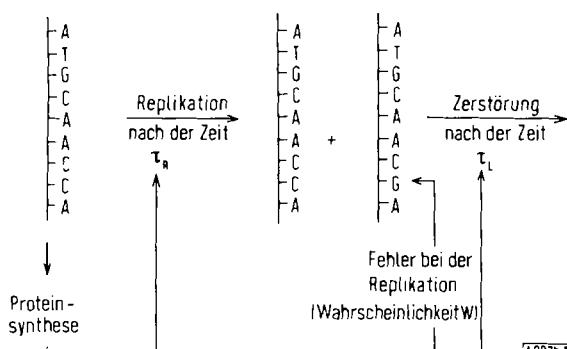


Abb. 1. Beeinflussung von τ_R , τ_L und W durch Proteine. – A = Adenin. T = Thymin, G = Guanin, C = Cytosin.

System, wenn einmal gebildet, einen Selektionsvorteil gegenüber Systemen mit anderer Basensequenz. Im Verlauf vieler Generationen ändern sich durch die Akkumulation von Ablesefehlern die Basensequenzen der DNA-Stränge und die Aminosäuresequenzen der entsprechenden Proteine allmählich, und es treten durch Selektion Formen auf mit Proteinen, deren Aminosäuresequenz (Realsequenz) einem Optimum (Idealsequenz) zustrebt.

Ist die Wahrscheinlichkeit W , daß im Elementarschritt des Replikationsprozesses ein Fehler auftritt, sehr gering, dann dauert es sehr lange, bis eine Mutation erscheint und durch Selektion vorteilhafter Mutanten eine Annäherung an die Idealsequenz stattfinden kann. Ist W relativ groß, dann treten sehr oft Mutationen auf, und es kann recht schnell die Idealsequenz oder eine ihr nahestehende Sequenz zufällig erreicht werden; im Verlauf der weiteren Replikation ergeben sich dann jedoch so häufig Fehler, daß sich die vorteilhaften Nachkommen trotz ihrer Selektionsvorteile gegenüber den vielen anderen nicht durchsetzen können. Es findet also keine Konvergenz zur Idealsequenz hin statt. Um eine möglichst schnelle Selektion zu erreichen, muß daher W einen optimalen Wert haben.

2. Computerrechnung

Um den Mechanismus der Selektion eines Proteins leichter zu überblicken, nehmen wir zunächst an, daß W und τ_R konstant sind und nur τ_L von der Aminosäuresequenz beeinflußt wird. Wir betrachten ein einfaches Modell mit einem DNA-Strang aus nur 9 Basen, also mit der Information für ein Protein aus drei Aminosäuren. Von den $4^9 = 262144$ Vertauschungsmöglichkeiten der vier möglichen Basen soll eine Austauschmöglichkeit, die wir willkürlich vorgeben, zu einem Protein führen, das das System optimal schützt, τ_L also maximal vergrößert (Idealsequenz). Je weniger sich die tatsächlich vorhandene Sequenz (Real-

sequenz) von der Idealsequenz unterscheidet, um so größer soll τ_L sein. Wir legen willkürlich fest:

$$\tau_L \sim r^{-F} \quad (1)$$

F soll die Anzahl der Fehlerstellen, also der Nichtübereinstimmungen zwischen Realsequenz und Idealsequenz sein; r ist der Faktor, um den sich die Überlebenschance eines Individuums bei einer vorteilhaften Mutation erhöht; wir setzen willkürlich $r = 1.5$. Je größer F ist, um so kleiner ist τ_L , um so früher muß das herausgegriffene Individuum also sterben.

Wir gehen davon aus, daß zu Beginn der Evolution 50 gleiche DNA-Stränge mit einer zufälligen Anfangssequenz vorliegen: wir lassen die 50 Individuen replizieren (Fehlerhäufigkeit $W = 1\%$, Fehlerstellen durch Zufall festgelegt) und merzen anschließend wieder 50 Exemplare aus. Dadurch bleibt die Gesamtzahl der Individuen konstant. Das Ausmerzen geschieht durch Zufall, jedoch so, daß relativ mehr Exemplare mit großem τ_L überleben als Exemplare, für die τ_L klein ist. Anschließend erfolgen erneut Replikation, Ausmerzen usw. Das durch den Zufall bestimmte Anbringen von Fehlerstellen im Replikationsschritt und das zufällige Ausmerzen von Individuen erfolgt im Computerprogramm dadurch, daß vor jedem Schritt eine Zufallszahl den weiteren Ablauf steuert.

Die zu Beginn vorliegenden DNA-Stränge werden sich von der Idealsequenz unterscheiden, da es unwahrscheinlich ist, durch Zufall von den 4^9 Möglichkeiten die richtige zu realisieren. Führt ein Fehler bei der Replikation zu einer besseren Übereinstimmung mit der Idealsequenz, dann wird die mutierte Form länger leben, sich bei gleicher Zeit τ_R öfter replizieren und sich damit bevorzugt vermehren; im umgekehrten Fall führt der Fehler zum Aussterben der Art. Auf diese Weise wird der Anteil an Individuen, die sich von der Idealsequenz nur wenig oder nicht unterscheiden, immer größer.

Tabelle 1. Selektive Anpassung eines Systems von 50 DNA-Strängen (Anfangszufallssequenz C G T A T C G T C) an ein Protein mit der Idealsequenz A G T T C C G A G; Fehlerhäufigkeit bei der Replikation $W = 1\%$ (d.h. von den 450 Basen der 50 Individuen werden im Replikationsschritt 4 bis 5 falsch kopiert); zu Beginn (Generation 0) stimmen alle DNA-Stränge an 4 Stellen mit der Idealsequenz überein ($F = 5$).

Generation	Zahl der Exemplare mit Fehlerstellen									
	0	1	2	3	4	5	6	7	8	9
0	50									
10					4	42	4			
20				14	23	12	1			
30			5	45						
40		3	34	13						
50	14	36								
60	23	26	1							
70	1	33	15	1						
80	14	31	5							
90	41	8	1							
100	46	4								
110	47	3								
120	41	9								
130	49	1								

Die Idealsequenz wird nach freiem Ermessen vorgegeben. Die Zufallssequenz der Ausgangsmoleküle stellt man durch Würfeln fest. Um einen wichtigen Gesichtspunkt der Evolution nicht außer acht zu lassen, gibt man mehrere Idealsequenzen vor: es gibt nicht nur eine Sequenz, die einem besonders vorteilhaften Protein entspricht, sondern sehr viele. Je nachdem, welche Zufallssequenz anfänglich gewürfelt wurde, streben die evolvierenden Formen zur einen oder anderen Idealsequenz. Daraus geht hervor, daß für den Gang der Evolution die anfänglichen Zufälle entscheidend sind.

Das Resultat einer solchen Computerrechnung^[2] ist für $W = 1\%$ in Tabelle 1 für die Idealsequenzen (A = Adenin, C = Cytosin, G = Guanin, T = Thymin)

```

G C A T T C T A A
A T G C A G G G T
A G T T C C G A G
T G G G A C C A C

```

und die anfängliche Zufallssequenz

```
C G T A T C G T C
```

aufgeführt; der Computer steuert die dritte Idealsequenz an, für die F den kleinsten Wert besitzt ($F = 5$).

Tabelle 2 enthält das Resultat für dieselben Idealsequenzen, wenn von der anfänglichen Zufallssequenz

```
G C T A T G A A C
```

ausgegangen wird; in diesem Fall wird die erste Idealsequenz angesteuert ($F = 5$).

Tabelle 2. Selektive Anpassung eines Systems von 50 DNA-Strängen (Anfangszufallssequenz G C T A T G A A C) an ein Protein mit der Idealsequenz G C A T T C T A A; Fehlerhäufigkeit $W = 1\%$, wie in Tabelle 1.

Generation	Zahl der Exemplare mit Fehlerstellen									
	0	1	2	3	4	5	6	7	8	9
0	50									
10	7	8	35							
20	40	10								
30	2	48								
40	27	23								
50	43	7								
60	2	45	3							
70	3	42	5							
80	4	22	21	3						
90	17	27	6							
100	34	15	1							
110	38	10	2							
120	48	2								
130	46	4								

Man sieht aus beiden Tabellen, wie die selektive Anpassung an die jeweilige Idealsequenz im Verlauf von 100 bis 120 Generationen allmählich vor sich geht.

Wiederholt man die Rechnung für den ersten Fall (Tab. 1) mit anderen Werten für W ($W = 0.25\%$ oder $W = 4\%$), dann werden die in den Tabellen 3 und 4 dargestellten Resultate erhalten. Während nach Tabelle 1 nach 100 Ge-

nerationen 92% aller 50 DNS-Stränge Idealsequenz besitzen ($W = 1\%$), ist bei $W = 4\%$ die Anpassung viel schlechter

Tabelle 3. Selektive Anpassung eines Systems von 50 DNA-Strängen wie in Tabelle 1, jedoch mit $W = 0.25\%$.

Generation	Zahl der Exemplare mit Fehlerstellen									
	0	1	2	3	4	5	6	7	8	9
0	50									
50	47	3								
100	49	1								
150	49	1								
200	2	48								
250	46	4								
300	43	7								
350	48	2								
400	50									
450	44	6								
500	49	1								
550	1	50								

Tabelle 4. Selektive Anpassung eines Systems von 50 DNA-Strängen wie in Tabelle 1, jedoch mit $W = 4\%$.

Generation	Zahl der Exemplare mit Fehlerstellen									
	0	1	2	3	4	5	6	7	8	9
0	50									
10	11	29	6	3	1					
20	1	12	33	4						
30	6	34	9	1						
40	4	10	23	13						
50	19	17	13	1						
60	18	19	12	1						
70	15	25	10							
80	29	17	3	1						
90	30	13	7							
100	33	15	1	1						
110	28	20	2							
120	30	16	3	1						
130	28	16	5	1						

(nach 100 Generationen besitzen nur 66% aller Exemplare Idealsequenz, obwohl die Evolution zu Beginn viel schneller verläuft als mit $W = 1\%$); bei $W = 0.25\%$ geht die Evolution sehr langsam voran (nach 100 Generationen ist noch kein einziges Exemplar mit Idealsequenz entstanden), nach genügend langer Zeit (500 Generationen) sind jedoch praktisch 100% aller Exemplare ideal angepaßt.

3. Abschätzung der optimalen Mutationshäufigkeit

Im Folgenden wird der optimale Wert von W für das im Abschnitt 2 betrachtete Modell abgeschätzt.

Wir betrachten Ahnenketten aus n Generationen (Abb. 2); n soll so groß sein, daß auf eine Ahnenkette im Mittel gerade 1 Replikationsfehler entfällt. Ist N die Zahl der Basen in der DNA ($N = 9$ in unserem Modell), dann gilt $N \cdot n \cdot W = 1$, also

$$W = 1/(N \cdot n) \quad (2)$$

Wir wollen zunächst n berechnen und gehen davon aus, daß zu Beginn des Vermehrungsprozesses a Individuen mit gleicher Basenfrequenz vorliegen ($a = 50$ in unserem Modell). Im Verlauf von n Generationen treten also a Replikationsfehler auf; bei F Nichtübereinstimmungen ist die

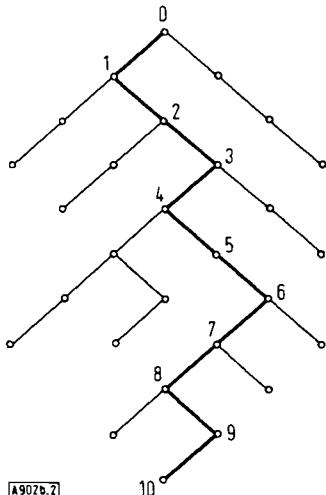


Abb. 2. Ahnenkette mit $n = 10$ Generationen (stark ausgezogene Linie). An jeder Verzweigungsstelle findet Replikation statt (weitere Verzweigungen sind der Übersicht halber fortgelassen worden).

Wahrscheinlichkeit, daß einer dieser Fehler vorteilhaft ist, gleich $F/(3 \cdot N)$ [der Austausch einer Base durch eine andere führt nur dann zu einer vorteilhaften Mutation, wenn eine der nicht übereinstimmenden Basen verändert wird (Wahrscheinlichkeit F/N), und wenn gleichzeitig die richtige Base eingesetzt wird (Wahrscheinlichkeit $1/3$)]. Es ist also die Anzahl a' der vorteilhaften Fehler

$$a' = a \cdot F/(N \cdot n) \quad (3)$$

In unserem Modell ($F = 5$, $N = 9$, $a = 50$) ist also

$$a' = 5 \cdot a/(3 \cdot 9) = 0.19a$$

es führen also nur etwa 20% aller Fehler zu vorteilhaften Mutanten. Wir nehmen vereinfachend an, daß jeder der a' Fehler, die im Mittel nach n Generationen gemacht werden, genau nach n Generationen auftritt. Die dann plötzlich vorhandenen a' vorteilhaften Mutanten haben gegenüber den anderen Individuen eine r -fache Überlebenswahrscheinlichkeit ($r = 1.5$ in unserem Modell), so daß sie sich stärker als alle anderen vermehren. Nach n Generationen existieren also a' spontan entstandene Individuen der vorteilhaften Mutante oder auf

$$Z = a; a' = 3 \cdot N/F \quad (4)$$

Individuen in der Population entfällt gerade ein Individuum der vorteilhaften Mutante. Das Verhältnis zwischen der Zahl der Individuen der vorteilhaften Mutante (m_v) und der Zahl der übrigen Individuen (m) ist also

$$m_v/m = a'/(a - a') = 1/(Z - 1)$$

Nach $n + 1$ Generationen ist das Verhältnis m_v/m etwa um den Faktor r größer (jede vorteilhafte Mutante hat gegenüber einem nicht mutierten Individuum die r -fache Überlebenswahrscheinlichkeit), nach $n + 2$ Generationen etwa um den Faktor r^2 und nach $2n - 1$ Generationen etwa um den Faktor r^{n-1} (nach genau $2n$ Generationen treten zum zweiten Mal Replikationsfehler auf). Es gilt also für das Verhältnis m_v/m nach $2n - 1$ Generationen

$$m_v/m = r^{n-1}/(Z - 1) \quad (5)$$

Bevor zum zweiten Mal Replikationsfehler auftreten, muß die vorteilhafte Mutante alle übrigen Individuen praktisch völlig verdrängt haben; wir wollen annehmen, daß nach den $2n - 1$ Generationen 10-mal so viele Exemplare der Mutante wie der ursprünglichen Form vorliegen, also

$$m_v/m = 10 \quad (6)$$

ist. Nach Gleichung (5) und (6) ist $r^{n-1} = 10 \cdot (Z - 1)$ oder

$$n = 1 + [1 + \lg(Z - 1)]/\lg r \quad (7)$$

Nach Gleichung (4) dauert die Selektion umso länger, je kleiner F ist; am längsten dauert sie für $F = 1$. Dafür gilt nach Gl. (4) $Z = 3N$, und im untersuchten Modell ($r = 1.5$, $N = 9$) ergibt sich daraus $n = 15$ und damit nach Gl. (2) $W = 0.8\%$.

In Übereinstimmung mit dem Computerergebnis liegt der optimale Wert von W also bei etwa 1%; nach 15 Generationen sollte eine vorteilhafte Mutante gebildet und nach 15 weiteren Generationen selektiert sein. Da sich die anfängliche Zufallssequenz von der ähnlichsten Idealsequenz in den betrachteten Beispielen an 5 Stellen unterscheidet, müßte die selektive Anpassung nach ungefähr $5 \cdot 2 \cdot 15 = 150$ Generationen erfolgt sein. Das trifft (Tabellen 1 und 2) tatsächlich zu.

Bei großen Werten von Z und n geht Gleichung (7) in die einfache Beziehung $n = \lg Z/\lg r$ über^[1].

Anmerkung: Es wurde davon ausgegangen, daß das Verhältnis m_v/m in einer Generation um den Faktor r ansteigt. Das gilt für $r - 1 \ll 1$ und trifft für größere Werte von r nur näherungsweise zu, wie sich aus folgender Überlegung ergibt: Nach n Generationen liegen a' Exemplare der vorteilhaften Mutante und $a - a'$ Exemplare der Normalform vor, nach der Replikation $2a'$ bzw. $2(a - a')$ Exemplare. Von diesen läßt man insgesamt a absterben, damit wieder die alte Population erhalten wird. Ist x die Sterbezahl bei der Normalform und y die Sterbezahl bei der Mutante, so gilt daher $x + y = a$. Da die Sterbewahrscheinlichkeit der Normalform um den Faktor r größer ist als die der Mutante, folgt weiter:

$$x/y = [(a - a')/a'] \cdot r$$

Mit diesen Beziehungen wird:

$$(m_v/m)_{n+1} = (m_v/m)_n \cdot f$$

mit

$$f = [2a' + 2(a - a')r - a]/[2a' + 2(a - a')r - ar]$$

Im Fall $r - 1 \ll 1$ wird $f = r$; setzt man wie in der Computerrechnung $r = 1.5$ und $a' = 0.19a$, so ist $f = 1.38$, also von r etwas verschieden.

Die numerischen Rechnungen wurden am Rechner TR4 der Zentralen Rechenanlage in Marburg ausgeführt. Den Herren cand. chem. G. Lauer und cand. chem. T. Lichtenberg danken wir für ihre Mitarbeit.

Eingegangen am 10. März 1971 [A 902 b]

[1] H. Kuhn, Angew. Chem. 84, 838 (1972); Angew. Chem. internat. Edit. 11, Heft 9 (1972).

[2] H. D. Försterling u. H. Kuhn: Physikalische Chemie in Experimenten. Verlag Chemie, Weinheim 1971.