ISSN 0029-3865

CBPF-NF-065/87

BIOGENESIS: AUTOCATALYTIC POLYMERIZATION OF DNA-LIKE MOLECULES AS A CRITICAL PHENOMENON*

by

Constantino Tsallis

Centro Brasileiro de Pesquisas Fisicas - CNPq/CBPF
Rua Dr. Xavier Sigaud, 150
22290 - Rio de Janeiro, RJ - Brasil

*Lectures given at the "International Workshop on the Theory of Disordered Systems in Biological Models" (Bogotá, September 1987), to be published by World Scientific Publishing, ed. L. Peliti and S. Solla.

ABSTRACT

A fundamental prebiotic stage of the appearance of life on Earth (and/or possibly elsewhere) is the formation, from a random assembly of oligomers, of information - containing DNA-like macromolecules. By assuming a autocatalytic polymerization mechanism essentially based in the complemetarity of the Crick and Watson base-pairs, we develop theoretically a chemical equilibrium picture which suggests that this stage may have occurred as a critical phenomenon. We review a renormalization group approach worked out in collaboration with R. Ferreira, as well as a computer simulation performed in collaboration with H.J. Herrmann. Both techniques, in spite of their differences, converge onto a Darwinian-like scenario which naturally incorporates biologically relevant aspects of diversity and selection, and which suggests that polymers like ADN double-chains might have been more primitive than proteins.

Key-words: Biogenesis; Critical phenomena; Renormalization group; Computer simulation.

Everything already existed before (Ben Akiba).
Nothing already existed before (Konrad Lorenz).

I- INTRODUCTION

The most primitive living organisms whose traces have been found on Earth are photo-bacterias dated of about 3.8×10^9 years. Since the Earth itself was formed about 10^9 years ealier, several and complex prebiotic stages were overcome during that period until the (possibly spontaneous) appearance of these organisms occurred. Among the various prebiotic stages which were probably followed (see [1] for an overall review) during the transition from inanimate to living matter, one of great importance (and still badly understood) certainly is the growth of codified macromolecules of the DNA-type (i.e., with potentiality for self-replication) starting from a random assembly of oligomers (dimers, trimers, tetramers, etc.) possibly fluctuating in a primordial soup. This is an important step of the chain which joins Organic Chemistry to Biology. It is the central scope of the works [2-6] which are going to be reviewed here. In spite of its crucial role, this step remains quite enigmatic. This is due in part to the fact that little related experimental work has been done (see Refs. [7,8] and references therein), at least not enough to provide an enlightening and comprehensive view of the problem. This situation is, in some sense, in contrast with the present knowledge[9-11] of a more primitive step, namely the formation of nucleotides (which, in long chains, form the nucleic acids) and aminoacids (which, in long chains, form the proteins), starting from H_2O, CO, methane, etc. (the transition from Inorganic to Organic Chemistry, generally speaking). This step is now considered to be based on scientifically reliable grounds, the scenario being violent non-equilibrium phenomena (electrical discharges, light, heat and radiation flashes, etc.) occurring in re-

latively simple (and possibly molecular-oxygen-less) atmosphere. Many experiments done along the lines of S. Miller's pioneering ones support this picture.

The growth of codified self-replicating macromolecules (DNA or RNA or their precursors) has recently attracted quite intensive theoretical effort[12-19] within thermodynamic and/or statistical mechanic frameworks. The central growth mechanism is assumed to be autocatalysis based on a Crick and Watson-like complementarity[20], A-T and C-G thus constituting the two complementary base-pairs. Let us make clear at this point that, although we shall use along the present work the notation A-T and C-G, we do not necessarily refer to the well known nucleotides (adenine, thymine, cytosine and guanine): the notation might refer to their precursors as well. Moreover, if the basic macromolecule is a RNA-like one[10,21-25], rather than a DNA-like one, the notation A-T would then refer to the A-U pair.

Alternative growth mechanisms, using clay[26], protein[27] and aminoacid pairing[28] as basic ingredients, have been proposed. However, autocatalysis based on A-T and C-G complementarity is very appealing[8,10,15-17,21-25] and this is the standpoint we adopt herein.

With the collaboration of R. Ferreira we have developed, since 1983, a chemical equilibrium critical phenomenon picture which enables the understanding of the growth of codified self-replicating polymers. The theoretical framework we have used is the renormalization group (RG)[29,30], more precisely a real space version of it, similar to those available in the literature[31-33] for standard polymerizations. Our first approach[2] assumed a *single* base-pair; it provided polymeric growth consistent with *diversity*, but with *no selection* in it. Our second approach[3] generalized the first one in the sense that the fugacity K_{AT} of the A-T link (Hy-

drogen bridge) in general is *different* from the fugacity K_{CG} of the C-G link (Hydrogen bridge); it came out that if *two* different base-pairs are assumed in the autocatalysis, both *diversity and selection* become possible, thus satisfying in a natural way the basic requirements for Darwinian evolution. However, at that level the treatment was not exempted from unsatisfactory results, namely, although we had selection, we had what we tend to consider the *bad* selection, in the sense that the most priviledged codes were those containing, along the chains, exclusively A-T pairs or exclusively C-G pairs, *which do not exist nowadays!* We then argued that the fugacities *along* the chains (i.e., *not* the complementary pairs, but the strong covalent-like links) being different among them, the plausible codes (*sequences* of A-T and C-G links) could easily become those which indeed exist in nature, i.e. ratios of numbers of A-T and C-G links not too different from unity. In this way, the *good* selection can be recovered in the picture. In fact, the importance of having *four*, and *not two*, different nucleotides (capable consequently of forming *two*, and *not one*, complementary pairs) has also been recently emphasized by Anderson[19]. Reference [4] is a comprehensive review of our RG approach and results.

Very recently, with the collaboration of H.J. Herrmann[5], we have implemented the same physical ideas in a computer by simulating the diffusion and linking of oligomers in a "soup" in order to observe the possible growth of polymers. The growth does occur in a way which is essentially consistent with the results obtained with RG. The second approach (computer simulation, CS) was done in order to check the invariance of the main concepts and results under changement of theoretical techniques. In fact, both RG and SC preserve and reinforce the chemical equilibrium critical phenomenon view-

point. They present, however, some differences, namely the CS approach does not deteriorate the fact that the growth we have worked with is essentially one-dimensional, whereas the approximative RG approach does, as will be seen through the type of phase diagrams that are obtained. However, none of the approaches takes into account the fact that DNA-like macromolecules might be seriously folded and present a certain amount of cross-links. This fact could transform the system into a fractal with fractal dimensionality higher than one. This could have as a consequence, as argued in Ref. [6], for the phase diagrams to be closer to those resulting within the RG approach than to those resulting within the CS one. We are presently working along this line.

The integrated and pedagogical review of the above theoretical studies constitutes the scope of the present lectures. In Section II we present the model, the RG formalism and the corresponding results; in Section III we present the CS approach and the corresponding results; we finally compare both and conclude in Section IV.

II- MODEL AND RENORMALIZATION GROUP APPROACH

The monomers (nucleotides) A,T,C,G can form double strings through *intrachain* strong (covalent-like) bonds (noted ——— in the illustration which follows) and *interchain* weak (hydrogen-bridge-like) bonds (noted ....) as illustrated below:

$$\begin{array}{ccccccccccc} \mathrm{A} & — & \mathrm{G} & — & \mathrm{T} & — & \mathrm{A} & — & \mathrm{C} & — & \mathrm{G} \\ \vdots & & \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\ \mathrm{T} & — & \mathrm{C} & — & \mathrm{A} & — & \mathrm{T} & — & \mathrm{G} & — & \mathrm{C} \end{array} \qquad (1)$$

Let K_{AT} and K_{CG} be the *fugacities* (or bonding constants) respectively associated with the *interchain* A-T and C-G complementary bridges ($K_{AT}, K_{CG} > 0$). K_{AT} and K_{CG} depend, in a complex unknown manner, on all the thermal equilibrium (or quasi-equilibrium) external parameters (temperature, pressure, humidity, concentrations of various salts, etc.) which characterize the *primordial soup*, assumed to contain arbitrary amounts of randomly codified oligomers (dimers, trimers, etc.) like that of scheme (1) (which represents and hexamer) or, more probably, smaller ones. We further assume that oligomers can grow through the autocatalytic process illustrated in Fig. 1. Notice that we have obtained, as a final product, the *hexamer* of scheme (1), whereas at the initial stage, we had nothing longer than *tetramers*. In the present illustration, (A-G-T-A) and (C-G) play the role of *growing fragments*, and (A-T-G-C) plays the role of *catalysing fragment*. We are assuming that the intrachain condensation (characterized in the example of Fig. 1 by the fugacity J_{AC} of the A-C covalent bond) between the two growing fragments is greatly favoured ($J_{AC} >> 1$) in the presence of the catalysing fragment bonded, to *both* growing fragments, through the interchain bridges.

In order to better understand the RG framework within which we shall perform calculations, let us first discuss the *single* base-pair particular case ($K_{AT} = K_{CG} \equiv K$; both A and C denoted by A; both T and G denoted by T). We perform the configurational analysis associated with the growth of a small oligomer (e.g., a *dimer* in Figure 2(a)), according to the following rules: (i) we consider all the growth-active configurations of all the catalysing fragments whose size is not longer than twice the growing fragment under consideration (we want to retain only the most probable mechanisms, and the probability of occurrence of catalysing fragments *much longer* than

the growing fragment is rather poor): (ii) the "weight" equals 1 when the catalysing fragment is unambiguously associated with the growing fragment under consideration, equals 1/2 when it can equally well be associated with the other growing fragment, and equals 0 (and is therefore absent from the figure) when it is unambiguously associated with the other fragment (to be more precise, when the number of non-connected residues at any given end of the catalysing fragment exceeds the number of its residues actually connected to the growing fragment under consideration); (iii) the number of growth-active ends (1 or 2) of the catalysing fragment can be disgarded (*procedure* I) or taken into account (*procedure* II) by introducing a "growth efficiency" which equals the number of growth-active ends; (iv) the interchain bonds are assumed independent (hence the effective fugacity of a given set of simultaneous bonds is just the *product* of the corresponding fugacities); (v) multiple catalysing processes (involving more than one catalysing fragment) or similar complex processes are neglected because of a presumably low probability of occurrence. These set of rules obviously involve a certain degree of arbitrariness; however we believe that any other "reasonable" set of rules would lead to results not essentially different from those we shall present.

Figure 2(a) yields, through the sum of (weight) x (growing efficiency) x (fugacity), the following effective fugacities:

$$R_2^{I}(K) = K + 4K^2 \qquad \text{(procedure I)} \qquad (2)$$

$$R_2^{II}(K) = K + 5K^2 \qquad \text{(procedure II)} \qquad (2')$$

The subscript 2 stand for dimer. We now repeat the configurational

analysis for the growth of a longer oligomer (e.g., a *trimer* in Fig. 2(b)). We obtain the following effective fugacities:

$$R_3^I(K) = K + 3K^2 + 8K^3 \qquad \text{(procedure I)} \qquad (3)$$

and also

$$R_3^{II}(K) = K + 3K^2 + 11K^3 \qquad \text{(procedure II)} \qquad (3')$$

We can now write down the RG recursive equation, namely

$$R_2^\alpha(K') = R_3^\alpha(K) \qquad (\alpha = \text{I or II}) \qquad (4)$$

Both recurrences admit the trivial (stable) fixed points K=0 (corresponding to lack of infinite growth, and characterizing the *finite growth* (FG) *phase*) and $K = \infty$ (characterizing the *infinite growth* (IG) *phase*). They also admit a critical (unstable) fixed point, namely $K^* = 1/8 = 0.125$ for procedure I, and $K^* = 2/11 \simeq 0.18$ for procedure II. The present calculation provides further information: while approaching the critical value K^*, the *mean length* ξ of the growing fragment diverges as $\xi \propto (K^* - K)^{-\nu}$, where the critical exponent ν is given (within the present RG approximation) by

$$\nu = \frac{\ell n(b/b')}{\ell n\left(\frac{dK'}{dK}\right)_{K^*}} = \frac{\ell n(b/b')}{\ell n\left[\frac{dR_b(K)/dK}{dR_{b'}(K)/dK}\right]_{K^*}} \qquad (5)$$

where b(b') is the size of the original (renormalized) oligomer un

der analysis (in our present example, b=3 for the trimer and b'=2 for the dimer) and $R_b(K)$ ($R_{b'}(K')$) the corresponding effective fugacity. We obtain $\nu \simeq 7.0$ for procedure I, and $\nu \simeq 3.3$ for procedure II. The smaller and more satisfactory (because more consistent with related calculations in polymer physics) value of ν obtained through procedure II, is to be attributed to the higher realism introduced by the growth efficiency. Anyhow it is completely out of the scope of the present very crude approximations to obtain reliable numbers for K* or ν: our arguments concern only the qualitative facts of the picture.

Summarizing, we have seen that the autocatalytic mechanism might lead, when approaching a critical value for the interchain fugacity, to the growth of codified self-replicating polymers. This already seems to us a very suggestive conclusion. However, if a *single* base-pair is assumed, all codes grow, and all do so at the same value of K: this is fine regarding diversity, *but*, from the biological standpoint, completely unsatisfactory in what concerns selection! We shall next see that the (realistic) assumptions of *two* (or more) different base-pairs, will lead to a remarkable improvement is this sense.

The parameter space of our problem will now be a two-dimensional one, namely determined by K_{AT} and K_{CG} (all intrachain fugacities are assumed infinite at this level of approximation). The RG flow will now be determined by (explicit or implicit) recursive relations of the following type:

$$K'_{AT} = f_{bb',\sigma}(K_{AT},K_{CG}) \tag{6a}$$

$$K'_{CG} = g_{bb',\sigma}(K_{CG},K_{AT}) \tag{6b}$$

where $f_{bb',\sigma}$ and $g_{bb',\sigma}$ are functions which will in general depend on the respective sizes b and b' of the original and renormalized oligomers we have chosen to work with, *as well as on the particular code which is growing* (and which is denoted by the index σ). Examples of such codes are the following: $\ldots K_{AT}K_{AT}K_{AT}K_{AT}\ldots$ (σ=1), $\ldots K_{CG}K_{CG}K_{CG}K_{CG}\ldots$ (σ=2), $\ldots K_{AT}K_{CG}K_{AT}K_{CG}\ldots$ (σ=3), $\ldots K_{AT}K_{AT}K_{CG}K_{CG}K_{AT}K_{AT}K_{CG}K_{CG}\ldots$ (σ=4), $\ldots K_{AT}K_{AT}K_{CG}K_{AT}K_{AT}K_{CG}\ldots$ (σ=5). The single base-pair particular case can be obtained through three different limits, namely: (i) $K_{AT}=K_{CG}\equiv K$ and arbitrary σ, therefore $f_{bb',\sigma}(K,K) = g_{bb',\sigma}(K,K) \equiv F_{bb'}(K)$; (ii) $K_{AT}\equiv K$, arbitrary K_{CG}, and σ=1, therefore $f_{bb',1}(K,K_{CG}) \equiv F_{bb'}(K)$; (iii) $K_{CG}\equiv K$, arbitrary K_{AT}, and σ=2, therefore $g_{bb',2}(K,K_{AT}) \equiv F_{bb'}(K)$. Furthermore, for codes which are invariant through $K_{AT} \rightleftarrows K_{CG}$ permutation (e g., σ = 3,4, but not σ=5) the following property must be satisfied: $f_{bb',\sigma}(X,Y) = g_{bb',\sigma}(X,Y)$ for arbritary (X,Y). Several of the above properties can be verified on the following examples.

1st example: growth of the $\ldots K_{AT}K_{CG}K_{AT}K_{CG}\ldots$ sequence (σ=3). The RG equations are given by

$$R_5^{II}(K'_{AT},K'_{CG},K'_{AT},K'_{CG}) = R_9^{II}(K_{AT},K_{CG},K_{AT},K_{CG}) \qquad (7)$$

and

$$R_5^{II}(K'_{CG},K'_{AT},K'_{CG},K'_{AT}) = R_9^{II}(K_{CG},K_{AT},K_{CG},K_{AT}) \qquad (8)$$

where

$$R_5^{II}(K_{AT},K_{CG},K_{AT},K_{CG}) = K_{AT} + 3K_{AT}K_{CG} + 5K_{AT}^2K_{CG} + 7K_{AT}^2K_{CG}^2 + 29K_{AT}^3K_{CG}^2 \qquad (9)$$

and

$$R_9^{II}(K_{AT},K_{CG},K_{AT},K_{CG}) = K_{AT} + 3K_{AT}K_{CG} + 5K_{AT}^2K_{CG}$$
$$+ 7K_{AT}^2K_{CG}^2 + 9K_{AT}^3K_{CG}^2 + 11K_{AT}^3K_{CG}^3$$
$$+ 13K_{AT}^4K_{CG}^3 + 15K_{AT}^4K_{CG}^4 + 89K_{AT}^5K_{CG}^4 \qquad (10)$$

See in Figure 3 the associated RG flow, which determines the corresponding critical line, and also exhibits that the two base-pairs case belongs to the same university class as the single base-pair case.

2nd example: growth of the $\ldots K_{AT}K_{AT}K_{CG}K_{CG}K_{AT}K_{AT}K_{CG}K_{CG}\ldots$ sequence (σ=4). The RG equations are given by

$$R_5^{II}(K'_{AT},K'_{AT},K'_{CG},K'_{CG}) = R_9^{II}(K_{AT},K_{AT},K_{CG},K_{CG}) \qquad (11)$$

and

$$R_5^{II}(K'_{CG},K'_{CG},K'_{AT},K'_{AT}) = R_9^{II}(K_{CG},K_{CG},K_{AT},K_{AT}) \qquad (12)$$

where

$$R_5^{II}(K_{AT},K_{AT},K_{CG},K_{CG}) = K_{AT} + \frac{3}{2}K_{AT}^2 + \frac{3}{2}K_{AT}K_{CG} + \frac{5}{2}K_{AT}^2K_{CG}$$
$$+ \frac{5}{2}K_{AT}K_{CG}^2 + 7K_{AT}^2K_{CG}^2 + 29K_{AT}^3K_{CG}^2 \qquad (13)$$

and

$$R_9^{II}(K_{AT},K_{AT},K_{CG},K_{CG}) = K_{AT} + \frac{3}{2}K_{AT}^2 + \frac{3}{2}K_{AT}K_{CG}$$
$$+ \frac{5}{2}K_{AT}^2K_{CG} + \frac{5}{2}K_{AT}K_{CG}^2 + 7K_{AT}^2K_{CG}^2$$

$$+ 9K_{AT}^{3}K_{CG}^{2} + \frac{11}{2} K_{AT}^{4}K_{CG}^{2} + \frac{11}{2} K_{AT}^{3}K_{CG}^{3} + \frac{13}{2} K_{AT}^{4}K_{CG}^{3}$$

$$+ \frac{13}{2} K_{AT}^{3}K_{CG}^{4} + 15K_{AT}^{4}K_{CG}^{4} + 89K_{AT}^{5}K_{CG}^{4} \qquad (14)$$

The corresponding RG flow is similar to that presented in Figure 3.

Equation (9) (Eq. (13)) has been established by making the configurational analysis associated with the sequence $\ldots K_{AT}K_{XX'}K_{YY'}K_{CG}K_{AT}K_{XX'}K_{YY'}K_{CG}\ldots$ (see Figure 4), calculating the effective fugacity $R_5^{II}(K_{AT},K_{XX'},K_{YY'},K_{CG})$, and then taking XX' = CG and YY' = AT (XX' = AT and YY' = CG). We have proceeded analogously to obtain equations (10) and (14).

We have indicated in Figure 5 the critical lines corresponding to various typical sequences. We notice an important improvement with respect to the one base-pair model: the picture presents now *both* diversity and selection! In other words, a microscopic basis for Darwinian evolution is now achieved. However, and in spite of this interesting achievement, the model is not yet free from two important limitations: (i) if we assume a reasonable time evolution of K_{AT} and K_{CG} (see Figure 5), the most priviledged codes are those presenting either very low or very high (A+T)/(C+G) ratios, a fact which is not easily consistent with the values $(1/2 \lesssim (A+T)/(C+G) \lesssim 2$; see [20]) associated with modern living systems (at least in the biosphere); (ii) the critical line is one and the same for all sequences of nucleotides which correspond to a single sequence of bonds (e.g., ...ACACAC..., ...AGAGAG..., ...ACTGTCTG..., etc., correspond to the sequence $\ldots K_{AT}K_{CG}K_{AT}K_{CG}\ldots$), a fact which has no biochemical support. Both limitations disappear by considering the different intrachain fugacities (they are, within a nearest-neighbor picture, 10 in number, and will be denoted by J_{AA}, J_{AT}, J_{AC}, J_{AG}, J_{TT}, J_{TC}, J_{TG}, J_{CC}, J_{CG} and J_{GG}). In fact, our approach thus far corre

sponds to assign to these 10 constants the value infinity. It is intuitive that finite values for these fugacities will make it *more difficult* to attain the point of infinite polymeric growth. We have indicated in Figure 6 the expected critical line assuming say that all the J's are equal among them (and equal to J), and that $K_{AT} = K_{CG} \equiv K$; note that K approaches K* when J diverges. The fact that the actual J's are *finite and different from one another*, will make all the critical lines (of Figure 5) to shift towards *higher* values of K_{AT} and K_{CG}. This shift is in general *different for differing sequences of nucleotides, even if they preserve the same sequence of* K_{AT}'s and K_{CG}'s. The result is indicated in Figure 7, by arbitrarily choosing $J_{CG} = J_{CG} = J_{GG} < J_{AA} = J_{AT} = J_{TT} < J_{AG} = J_{TC} < J_{AC} = J_{TG} < \infty$. We have not carried on actual RG calculations corresponding to finite J's. They are in principle tractable, though burdensome because of the large number of RG parameters. It is clear, in any case, that this is a realistic path for overcoming the two limitations mentioned previously.

Before closing the RG section let us mention a point that will turn out to be relevant later on. The present approach is approximative since we are renormalizing growing fragments of *finite* size b into growing fragments of size b' <b. The exact answer is hopefully achieved only in the $b \to \infty$ limit, which we have not attempted to perform. But we see that, within procedure II for instance (in fact the same facts occur within procedure I), the RG using (b,b')=(3,2) yields K* ≃ 0.18 whereas the RG using (b,b')=(9,5) yields K* ≃ 0.52 (see Figs. 3 and 5). This trend can be considered as an indication that K* can keep increasing while b increases. It could even happen that $\lim_{b\to\infty} K^* = \infty$, thus collapsing the entire set of critical lines onto infinity ($K_{AT}=K_{CG}=\infty$). Later on we come back onto this important point.

III - COMPUTER SIMULATION APPROACH

Let us now describe how we simulate the primordial soup in the computer. We consider two square lattices of size $L_x \times L_y$ (in crystal units), the *top* lattice and the *bottom* lattice, each with periodic boundary conditions in both directions. On the sites of the lattices we place A, T, C and G nucleotides (at most one nucleotide per site). We respectively denote by N_A^{tot}, N_T^{tot}, N_C^{tot} and N_G^{tot} the total number of A, T, C and G nucleotides placed in each lattice (the same quantities are assumed for both lattices). Covalent bondings are possible (to form dimers, trimers, etc.) only within each lattice, and only along the x-direction. That means that in *both* lattices *all* the chains will be parallel to the x-axis. The hydrogen bridges can appear only between the top and the bottom lattice. That means that if on a given site of the top lattice there is, say, an A nucleotide, and on the same site of the bottom lattice a T nucleotide, then they form a complementary pair and may form a bonding. We denote respectively by p_{AT} and p_{CG} the probabilities of chemical bondings of the A-T and C-G pairs between the two lattices. The variables p_{AT} and p_{CG} are convenient for computer use. They are directly related to K_{AT} and K_{CG} respectively: when K monotonously increases from zero to infinity, p monotonously increases from zero to one (the relation is similar to $p = K/(1+K)$).

The starting configuration is obtained by randomly placing nucleotides in form of monomers and in form of dimers on both lattices (we randomly place first the dimers and then the monomers). We can place, on each lattice, N_A A-monomers, N_T T-monomers, N_C C-monomers, N_G G-monomers, N_{AA} AA-dimers, N_{AT} AT-dimers, etc. The following relations are obviously satisfied:

$$N_A^{tot} = N_A + 2N_{AA} + N_{AT} + N_{AC} + N_{AG} \qquad (15.a)$$

$$N_T^{tot} = N_T + N_{AT} + 2N_{TT} + N_{TC} + N_{TG} \qquad (15.b)$$

$$N_C^{tot} = N_C + N_{AC} + N_{TC} + 2N_{CC} + N_{CG} \qquad (15.c)$$

$$N_G^{tot} = N_G + N_{AG} + N_{TC} + N_{TG} + 2N_{CG} \qquad (15.d)$$

and

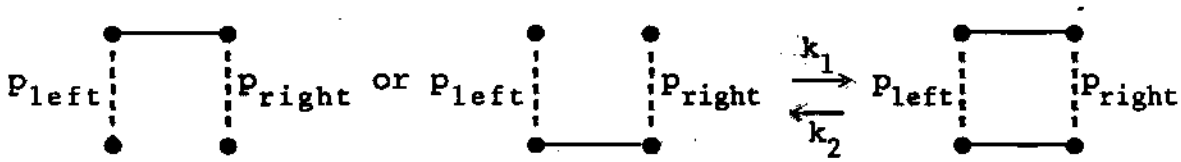
$$N_A^{tot} + N_T^{tot} + N_C^{tot} + N_G^{tot} \leq L_x L_y \qquad (16)$$

We shall work with $N_A^{tot} = N_T^{tot}$ and $N_C^{tot} = N_G^{tot}$; the ratio N_A^{tot}/N_C^{tot} will be used as a parameter of the model but we usually consider it to be unity. Also shall we assume that the AA, AT, TA and TT dimers appear equally often, i.e., $N_{AA} = N_{TT} = N_{AT}/2$; analogously $N_{CC} = N_{GG} = N_{CG}/2$. Finally, we assume $N_{AC} = N_{TC} = N_{AG} = N_{TG}$. Summarizing, we control the concentrations corresponding to each initial configuration by fixing 5 parameters, namely N_A^{tot}, N_C^{tot}, N_{AA}, N_{CC} and N_{AC}.

After the starting configuration is set, the computer simulates its time evolution. Two main processes occur: diffusion and growth. Diffusion is simulated in the following way. We randomly choose a chain (or monomer) of the top lattice. We then check whether we can move it one step along the positive x-direction (the first-neighbouring site along the +x end of the chain must be empty): if we can we do, if we cannot we abandon it. We then randomly choose another chain of the top lattice and check whether we can move it one step along the positive y-direction (*all* first-neighbouring sites on the +y side of the chain must be empty): if we can we do,

if we cannot we abandon it. We repeat the operation along the negative x-direction, and finally along the negative y-direction. Under these conditions a shift in y-direction is less likely to occur than in x-direction since it is less probable that *all* the sites along the side of the chain are empty than that only the site at the end of the chain is empty. To better control this anisotropy in the diffusion, which essentially seems physical to us, we have generalized the above procedure in the sense that we choose the x-direction m_x times and the y-direction m_y times, and allow for $m_x/m_y \neq 1$. We found, however, that our final result did not depend on m_x/m_y. The procedure used for the bottom lattice is exactly the same. So we define one *diffusion step* by m_x (and m_y) attempts of diffusion in the positive x (and y)-direction and m_x (and m_y) attempts in the negative x (and y)-direction and this for each of the two lattices. For most of our data we have chosen $m_x = m_y = 1$.

The second essential process is growth. Since we want to reach thermal equilibrium our growth mechanism must fulfill detailed balance. The autocatalytic growth (and breaking) process we shall consider is the following one:



(17)

where both p_{left} and p_{right} can take the values p_{AT} and p_{CG}, and where the transition rates k_1 and k_2 must satisfy

$$\frac{k_1}{k_2} = \frac{2p_{left}p_{right}}{p_{left}(1-p_{right}) + p_{right}(1-p_{left}) + (1-p_{left})(1-p_{right})} = \frac{2p_{left}p_{right}}{1-p_{left}p_{right}} \tag{18}$$

where we have assumed independence between the "left" and "right" hydrogen-like bonds linking complementary nucleotides. At the left hand side of Eq. (17) we have three chains of lengths ℓ_1, ℓ_2 and ℓ_3 respectively; at the right hand side we have two chains, whose lengths respectively are $\ell_1+\ell_2$ and ℓ_3, the third chain having acted as a catalyser for the (covalent) junction of the other two. A *growth step* is defined in the following way: Randomly any site is chosen for which a complementary pair is present on the two lattices. If the neighbouring site (conventionally in negative x-direction) does not have a complementary pair too, the growth attempt is abandoned. Otherwise there are three possibilities: (1) between the two sites there is no covalent bond in either of the lattices, (2) there is a covalent bond in only one of the lattices, and (3) there are covalent bonds in both lattices. In the first case the growth step is abandoned. In the second case the open bond will be closed by a covalent bond with probability $p_{left}p_{right}$, i.e., with probability p_{AT}^2, $p_{AT}p_{CG}$ or p_{CG}^2 depending if the two complementary pairs were both of the type A-T, one of type A-T and one of type C-G, or both of type C-G. Finally, in the third case, one of the two (randomly chosen) covalent bonds will be destroyed with probability $(1-p_{left}p_{right})/2$, i.e. $(1-p_{AT}^2)/2$, $(1-p_{AT}p_{CG})/2$ or $(1-p_{CG}^2)/2$ depending on if the complementary pairs were both of type A-T, one of type A-T and one of type C-G, or both of type C-G. Through this rule a thermodynamical equilibrium is well defined since a detailed balance condition is fulfilled. The variables p_{AT} and p_{CG} determine the "willingness" of a complementary pair to contribute to the creation of a covalent bond. Since the probability of having two neighbouring complementary pairs is low, only a small fraction of growth steps is successful, typically only about e-

very hundredth for the parameters that we are going to use.

In the time development, both processes, diffusion and growth, must be mixed. So we define one *time step* as consisting of first m_d diffusion steps and then m_g growth steps. In most cases we have chosen $m_d=m_g=10$. After the application of some time steps, a certain amount of the originally placed monomers and dimers has grown into longer chains (trimers, tetramers,...,polymers). In Fig. 8 we show a typical configuration at two different times (since the chains are aligned in the x-direction we have chosen L_x larger than L_y). Finally, if enough time steps have been applied the chain-length distribution n_ℓ will come into some equilibrium, where n_ℓ is defined as the number of chains of length ℓ. In Fig. 9 we show a typical equilibrium distribution of chain lengths. This distribution decreases monotonically like the cluster-size distribution of percolation and opposed to what is found in many irreversible cluster aggregation models [34-36]. We monitored two quantities that are related to n_ℓ: the mean chain length $\xi \equiv \langle \ell \rangle = \sum_{\ell=1}^{\infty} (\ell\, n_\ell) / \sum_{\ell=1}^{\infty} n_\ell$ and its fluctuation $\Delta \equiv [\langle \ell^2 - \langle \ell \rangle^2 \rangle]^{1/2} = [\sum_{\ell=1}^{\infty} (\ell^2 n_\ell) / \sum_{\ell=1}^{\infty} n_\ell - \langle \ell \rangle^2]^{1/2}$.

In order to control statistical fluctuations we used the following procedure: First m_{eq} time steps were discarded to assure that the system be in equilibrium. Then each m_s time steps we extract the data from our system and at the end we take the average value over all these data. m_s is usually chosen to be 200. In addition we also average over several starting configurations, but it turns out from our data that this second averaging is not really necessary; it seems that our system is ergodic. In addition to ξ and Δ we also monitor the length ℓ_{max} of the longest chain found at a given time step.

One of our central aims is to see if some selection of patterns

(or codes) naturally occurs in our model under an appropriate choice of parameters. It is not easy to quantify selection. In addition, useful quantities like the entropy [37] are very hard to obtain accurately via numerical simulations. We have opted, therefore, for a very simplified approach: We look at how often a certain type of covalent bond appears. So we calculate the fraction $f_{A,T}$ of covalent bonds that are of type (A,T), i.e., A-A, A-T, T-A or T-T; the fraction $f_{C,G}$ of covalent bonds that are of type (C,G), i.e., C-C, C-G, G-C or G-C, and the fraction $f_{A,C}$ of covalent bonds that are of type (A,C), i.e. A-C, C-A, A-G, G-A, T-C, C-T, T-G or G-T. The fraction is taken with respect to all covalent bonds on both lattices. At the initial time of any configuration we have

$$f_{A,T} = \frac{N_{AA} + N_{AT} + N_{TT}}{N_{AA} + N_{AT} + N_{TT} + N_{CC} + N_{CG} + N_{GG} + N_{AC} + N_{AG} + N_{TC} + N_{TG}} \qquad (19.a)$$

$$f_{C,G} = \frac{N_{CC} + N_{CG} + N_{GG}}{N_{AA} + N_{AT} + N_{TT} + N_{CC} + N_{CG} + N_{GG} + N_{AC} + N_{AG} + N_{TC} + N_{TG}} \qquad (19.b)$$

$$f_{A,C} = \frac{N_{AC} + N_{AG} + N_{TC} + N_{TG}}{N_{AA} + N_{AT} + N_{TT} + N_{CC} + N_{CG} + N_{GG} + N_{AC} + N_{AG} + N_{TC} + N_{TG}} \qquad (19.c)$$

At any time $f_{A,T} + f_{C,G} + f_{A,C} = 1$.

In addition to fractions $f_{A,T}$, $f_{C,G}$ and $f_{A,C}$, we also monitor the distribution of these three types of covalent bonds according to the chain lengths, i.e., we determine $n_\ell^{A,T}$, $n_\ell^{C,G}$ and $n_\ell^{A,C}$ which tell how many covalent bonds of type (A,T), (C,G) and (A,C) there are in a chain of length ℓ. Through a convenient normalization, we may impose $n_\ell^{A,T} + n_\ell^{C,G} + n_\ell^{A,C} = 1$. To illustrate the effect of selection we show, in Fig. 10, $n_\ell^{A,T}$ and $n_\ell^{C,G}$ obtained for a specific choice of parameters. Clearly, the two quantities be-

have differently; $n_{\ell}^{A,T}$ has larger values and fluctuates more for intermediate chains lengths ($5 < \ell < 20$). This disparity between covalent bonds of type (A,T) and type (C,G) is for us evidence that not all patterns appear equally often but that some of them have been selected to appear preferentially.

Let us further focus the results. How long and how large do we need to simulate in order to get reasonable results? Let us start by looking at a couple of histograms: the development of ξ in time as shown in Fig. 11 for both isotropic ($m_x=m_y=1$) and anisotropic ($m_x=1$, $m_y=9$) diffusion. At about 20000 time steps this 160 x 10 system seems to come so close to equilibrium that the statistical fluctuations seem to be of the same order of magnitude of the systematic deviation from equilibrium, so $m_{eq}=20000$ is reasonable. Furthermore, we see in Fig. 11 that the fluctuations in ξ increase with increasing ξ which comes from the fact that the change in length of a chain at one breaking is proportional to the length itself. Another interesting effect is that, instead of the most common exponential approach to equilibrium, we have here a nearly linear increase. Specifically, at very early times the slope is smaller than at later times. This uncommon behaviour presumably comes, among others, from the fact that the monomers must enter into a chain before they can be active as a complementary pair to a growth process. Comparison between Figs. 11(a) and 11(b) exhibits the statistical irrelevance of anisotropy in the diffusion steps. The time needed to perform a simulation such as those presented in Fig. 5 is about 10 minutes on an IBM 360/158.

To estimate the system size needed for a given choice of parameters it is good to see how long the longest chain was that ever appeared during a given simulation. For the simulation of Fig. 11(a),

for instance, it was 67, so that L_x=160 seems enough.

An important result concerns the existence of critical points, namely a choice of parameters at which the equilibrium characteristic length ξ of the chain diverges. Such a point is of deep interest in Biogenesis since, if the picture we are here developing for this prebiotic stage is correct, only approaching such a critical point one can generate arbitrarily long DNA chains. The most relevant parameters of the model are p_{AT} and p_{CG} since, as we said before, they contain in some way the external variables of the physical situation like temperature, pressure, humidity, various salt concentrations, light intensity, etc. We will therefore explore first the (p_{AT},p_{CG}) plane. In Fig. 12 we see how the equilibrium value of ξ changes along three different paths in the (p_{AT},p_{CG}) plane. Clearly ξ seems to diverge only at $p_{AT}=p_{CG}=1$ which is therefore the only critical point of our problem. This is a consequence of the one-dimensional nature of our growth model. We shall discuss this point later on. If one approaches $(p_{AT},p_{CG})=(1,0.9)$, for instance, the value of ξ appears to converge towards a finite value. The divergence of ξ at the critical point is consistent with a power law,

$$\xi \propto (1-p_{AT})^{-x} \qquad (20)$$

because the points in Fig. 12 lie quite well on straight lines. The exponent x does not depend on the path used to approach the critical point. We find $x=0.49 \pm 0.05$, which suggests that Eq.(20) could well be a square-root behaviour: $x=1/2$. The proportionality constant of Eq.(20) does, however, change if the path in the (p_{AT},p_{CG}) plane is changed. A formula that fits reasonably well all of our

data is

$$\xi \simeq \frac{A}{[(1-p_{AT})^2+(1-p_{CG})^2]^{1/4}} \qquad (21)$$

where A is of order unity.

The average longest chain ℓ_{max} and the fluctuations Δ of the chain length also diverge with a powelaw like that of Eq.(20) with the same exponent x within the statistical error bars. This is seen in Fig. 13 for one of the paths of Fig. 12 that approaches the critical point.

Let us now come back to selection and consider the ratio $r \equiv f_{A,T}/f_{C,G}$. If one approaches the critical point along the line $p_{AT}=p_{CG}$, one always has r=1 within the statistical fluctuations. On the other hand we show, in Fig. 14, what happens if one approaches the critical point along the line $1-p_{CG}=4(1-p_{AT})$. We see that $r > 1$ on the whole line, except possibly at the critical point itself. Approaching the critical point r-1 vanishes (i.e., $f_{A,T}-f_{C,G}$ tends to vanish) in a way that seems to be described by a powerlaw since the data in the log-log plot of Fig. 14 lie more or less on a straight line. We conclude from this that for all values $p_{AT} \neq p_{CG}$ there will be some selection in equilibrium; but since for the critical point itself p_{AT} equals p_{CG}, we cannot get selection for infinitely long chains in the present conditions (namely, $N_A^{tot}=N_C^{tot}$ and $N_{AA}=N_{CC}=N_{AC}/2$). Thus a possible scenario for selection in Biogenesis could have been that nature approached the critical point along a line $p_{AT} \neq p_{CG}$ so that $f_{A,T} \neq f_{C,G}$, and that before the critical point was reached the system started, for some reason, to defy thermal equilibrium by undergoing various non-equilibrium processes which would ultimately lead to a partial isolation of the system from the exterior by means

of membranes. In any event it is very unlikely that nature ever reached the critical point also because infinitely long chains do not exist.

Another selection mechanism can be related to asymmetries in the starting configuration, i.e., if we do not demand any more $N_A^{tot}=N_C^{tot}$ and $N_{AA}=N_{CC}=N_{AC}/2$. It is of course probable that the prebiotic soup was not that symmetric and for that reason we also investigated other starting configurations. First we looked at the case where $N_A^{tot}=N_C^{tot}$ but the dimers have different concentrations; so we chose $N_{AA}=18$, $N_{CC}=2$ and $N_{AC}=12$ for $p_{AT}=p_{CG}$ in an 80x10 system. We found that the equilibrium times were somewhat longer than in the symmetric case but, except for that, all equilibrium properties were essentially identical. So we conclude that disparities in the original concentrations of dimers do disappear in the course of equilibration. One should, however, remark that all types of dimers must be represented at the beginning by at least one sample because otherwise, due to our complementarity mechanism, this particular type of covalent bond will never be created during the growth.

The other, more radical, asymmetry that can be introduced in the starting configuration is $N_A^{tot} \neq N_C^{tot}$. Specifically we looked at $N_A^{tot}=272$, $N_C^{tot}=144$, $N_{AA}=18$, $N_{CC}=2$ and $N_{AC}=12$ at $p_{AT}=p_{CG}$ in an 80x10 system. We found, while approaching the critical point, $r=5\pm1$, i.e., a very strong preference for covalent bonds of type (A,T). This is therefore another possible selection mechanism which works even for $p_{AT}=p_{CG}$ and which subsists even at the critical point. The appearance of this type of selection is very easy to understand: since the total number of A and T nucleotides is larger than the total number of C and G nucleotides, it is more likely to find covalent bonds of the type (A,T) than covalent bonds of the type (C,G) in

the patterns.

Summarizing we have simulated the phenomenon in computer trying to make the model as realistic as possible without complicating it beyond our computational capacities. So we introduced: (i) straight DNA-like chains that are aligned along a unique direction and that perform a two-dimensional diffusion (instead of possibly very folded chains that can quite freely rotate and that perform a three-dimensional diffusion); (ii) discrete (instead of continuous) time and space; (iii) sequential diffusion and growth processes (whereas it seems more realistic that they have quite frequently occurred simultaneously); (iv) bi-planar configuration for autocatalysis (whereas it seems more realistic to occur in all directions); (v) fugacity-variables only for the A-T and C-G hydrogen-like bondings, the fugacities of the 10 possible first-neighbour covalent bonds being held infinitely strong (which clearly is but a first approximation); (vi) an autocatalytic growth (and breaking) process which takes into account the possible presence of complementarity links only on the two nucleotides involved in the first-neighbour covalent bond (in fact it is not obvious that nucleotides others than those do not play an appreciable role); (vii) independent occurrence of the two first-neighbouring hydrogen-like bondings responsible for the autocatalysis of a covalent bond (whereas some correlation cannot be excluded). In spite of these many simplifying assumptions, it seems a priori quite plausible (and the results reinforce a posteriori this belief) that the main ingredients have been retained in our model. In particular, the fact that we have allowed, together with the growth process, its reverse (breaking) process enables the system to achieve thermodynamic equilibrium, for each choice of parameters.

Our main computer results have been that only for $p_{AT} = p_{CG} = 1$ our chains grow infinitely long, and the equilibrium average chain length, its fluctuation as well as the longest chain length diverge at this point like $1/\sqrt{1-p_{AT}}$. Selection, i.e. preference of some patterns over others, can occur if $p_{AT} \neq p_{CG}$ or if in the starting configuration the total numbers of the different species of nucleotides are chosen differently. The whole picture suggests, for the particular stage of Biogenesis under discussion, that the main ingredients of a Darwinian-like molecular evolution, namely diversity and selection, can be naturally incorporated in a scheme of thermodynamical quasi-equilibrium. The phenomenon essentially looks like a critical one, the most relevant external parameters being the fugacities of the hydrogen-like bridges between the nucleotides of complementary pairs.

Certainly the present simulation is not exhaustive. A detailed investigation of the influence of the starting configurations would be welcome. A better understanding of the peculiar increase of the chain length at early times, as well as of our square-root laws and Eq.(21) is also needed.

IV- CONCLUSION

We have discussed the polymerization of codified self-replicating DNA-like macromolecules using two different theoretical approaches, namely a renormalization group (RG) and a computer simulation (CS). Let us concentrate on the phase diagrams they yield: see Fig. 15. Why are they not one and the same? The correct one is that of Fig.5(b), i.e., the CS one. Indeed, the growing double-chain

polymer is *one-dimensional* from the topological point of view, therefore the critical point has to be at $p_{AT}=p_{CG}=1$ (i.e., $K_{AT}=K_{CG}\to\infty$) since[38] no infinite chains are possible at any *finite* temperature for *finite* couplings (i.e., for $p_{AT}<1$ or $p_{CG}<1$, or equivalently $K_{AT}<\infty$ or $K_{CG}<\infty$). Why then the RG approach has led to Fig. 15(a)? The hint has already been given in Section II: the RG calculations are only *approximate* since the renormalizations have been done among cells with *finite* size. The exact answer can only be achieved in principle in the limit of cells with *infinite* size. Therefore for RG cells which are increasingly larger we should expect the critical lines of Fig.15(a) to shrink onto the $p_{AT}=p_{CG}=1$ corner, thus reproducing Fig.15(b). In fact we recall that we have numerically observed this shrinking tendency while considering larger and larger oligomers to perform the renormalization.

Is it then the phase diagram of Fig.15(b) the ultimate form we can expect? The answer is *no*. Indeed, if we take into account the chemical cross-links which *do exist* between different parts of the real *folded* polymer (see, for example, [39]), the system may become a fractal object with fractal dimensionality d_f higher than 1. Consequently, the system not being strictly one-dimensional any more, the statments developed in [38] do not apply, and we are in right for expecting critical lines like those appearing in Fig. 15(a).

To say it in other words, in the CS approach we did one approximation, namely we neglected the cross-links of the polymer. In the RG approach we did that approximation plus another one, namely not considering infinite large RG cells. It is our belief that the real phenomenon might have occurred in a manner which is closer to that of Fig.15(a) rather than that of Fig.15(b). Consequently, we verify once more that in Science it might happen to be preferable

to do *two* mistakes rather than only *one*! We are presently working on a more realistic computer simulation, which allows for folding and cross-links, in order to test the above expectations.

To conclude let us list some of the relevant aspects which emerge within our chemical quasi-equilibrium critical phenomenon picture for the polymerization of DNA-like macromolecules based on autocatalytic growth using a Crick and Watson-like complementarity between nucleotides:

(i) A *single* base-pair ($K_{AT}=K_{CG}$) yields critical growth consistent with *diversity* but not with *selection*;

(ii) Two (or more) base-pairs ($K_{AT}\neq K_{CG}$) yield polymeric growth consistent with *both diversity and selection*; these fundamental ingredients naturally come into the theory, thus providing a microscopic basis for Darwinian evolution; life would have then appeared from a certain amount of self-replicating codes, and not from a single one (those different codes would have grown at different, though close, moments of the Earth evolution);

(iii) The role played by *finite* values for the interchain covalent fugacities (J_{AA},J_{AC}, etc.) is to make possible realistic values for the nucleotide ratio (A+T)/(C+G) (roughly between 1/2 and 2);

(iv) In the old querelle "which came first: nucleic acids or proteins?", our picture suggests a more primitive role for the nucleic acids;

As a final remark we might add that, if analogies with spin 1/2 magnetic systems are to be done[15] for the prebiotic stage under study, the roughest "reasonable" model seems to be that in which *the binary code* ($S=\pm 1/2$) *refers to the base-pairs* (A-T or C-G) *and not to the nucleotides* (purine or pirimidine).

Unnecessary to insist how welcome would be experiments confirming the picture herein exposed.

It is with great pleasure that I remind here that without the enriching collaboration of R. Ferreira and H.J. Herrmann this review would had never been possible. I have benefited from stimulating related discussions with many people. This is maybe not the place for mentioning them all: I cannot however resist the temptation of making explicit some of them, namely G. Bemski, R. Maynard, L. Peliti, S. Solla and G. Weisbusch.

CAPTION FOR FIGURES

Figure 1: Example of autocatalytic process which can increase the mean length of the macromolecules by using a mechanism based on a Crick and Watson-like complementarity.

Figure 2: Configurational analysis of the catalysing fragments corresponding to the growth of a *dimer* (*a*) and of a *trimer* (*b*).

Figure 3: Critical line (full line) in the (K_{AT},K_{CG}) fugacity space (procedure II), separating the finite growth (FG) phase from the infinite growth (IG) one of the sequence $...K_{AT}K_{CG}K_{AT}K_{CG}...$ Arrows and dashed lines indicate RG flow; the central dot indicates the single base-pair critical fixed point (responsible for the university class of the whole critical line).

Figure 4: Configurational analysis, within both procedures I and II, of the catalysing fragments corresponding to the growth of a *pentamer* (sequence $...K_{AT}K_{XX},K_{YY},K_{CG}K_{AT}K_{XX},K_{YY},K_{CG}...$).

Figure 5: Critical lines (in the (K_{AT},K_{CG}) fugacity space) corresponding to the growth of selected sequences (the dashed line is indicative); FG (IG) denotes the finite (infinite) growth phase. The point at $K_{AT} = K_{CG} = K^*$ reproduces the fixed point of Figure 3; the dotted line is a symmetry axis of some of the sequences (e.g., $...K_{AT}K_{CG}K_{AT}K_{CG}...$ and $...K_{AT}K_{AT}K_{CG}K_{CG}K_{AT}K_{AT}K_{CG}K_{CG}...$). The arrows indicate a plausible (slow) time evolution of K_{AT} and K_{CG}.

Figure 6: Indicative FI-IG critical line (one and the same for all sequence types) corresponding to $K_{AT}=K_{CG}\equiv K$ and all J's equal among them (and equal to J). K* refers to the single base-pair critical point of Figure 3.

Figure 7: Indicative FG-IG critical lines corresponding to the growth of different *nucleotide sequences* (not only different sequences of interchain links). The dashed line is a symmetry axis of some sequences (e.g., ...ACAC..., ...AGAG...); the dotted lines indicate the value K* of all previous figures. The arrows indicate a plausible (slow) time evolution of K_{AT} and K_{CG}.

Figure 8: Starting configuration (a) and configuration after 3000 time steps (b) for $p_{AT}=p_{CG}=0.975$ in one of the two lattices of a system of size $(L_x,L_y)=(20,10)$.

Figure 9: Equilibrium chain length distribution n_ℓ (in arbitrary units) obtained for $p_{AT}=0.995$ and $p_{CG}=0.98$ after 40000 time steps. The parameters are $L_x=160$ and $L_y=10$, $N_A^{tot}=N_C^{tot}=416$, $N_{AA}=N_{CC}=N_{AC}/2=16$, $m_x=m_y=1$.

Figure 10: Distributions $n_\ell^{A,T}$ and $n_\ell^{C,G}$ (in arbitrary units) of covalent bonds of type (A,T) and of type (C,G) according to the length ℓ of the chain. Same parameters as in Fig. 9.

Figure 11: Histogram of ξ obtained with: (a) the same parameters of Fig. 9; (b) the same parameters of Fig. 9 except for $(m_x,m_y)=(1,9)$.

Figure 12: Log-log plot of ξ against $1-p_{AT}$ at equilibrium along three different paths in the (p_{AT},p_{CG}) plane: $p_{CG}=p_{AT}$ (x); $1-p_{CG}=$

$4(1-p_{AT})$ (o) and $p_{CG}=0.9$ (▲). In the insert we show these three paths in the (p_{AT},p_{CG}) plane itself. For all runs the starting configuration has been chosen completely symmetric, i.e., $N_A^{tot} = N_C^{tot}$ and $N_{AA} = N_{CC} = N_{AC}/2$, and $N_A^{tot} = 0.26\ L_x L_y$.

Figure 13: Log-log plot of Δ(●) and ℓ_{max}(▽) against $1-p_{AT}$ along the path $1-p_{CG} = 4(1-p_{AT})$ and the same starting conditions as in Fig. 12.

Figure 14: Log-log plot of r-1 against $1-p_{AT}$ along the same path and starting conditions of Fig. 13. For two values of p_{AT} we have represented two different statistical samples to exhibit the magnitude of the fluctuations.

Figure 15: Phase diagrams within the RG (a) and CS (b) approaches. FG and IG respectively denote the *finite growth* and *infinite growth* phases. The various critical lines in (a) correspond to various typical codes along the chains. The dot ● in (b) indicates the critical point for all codes.

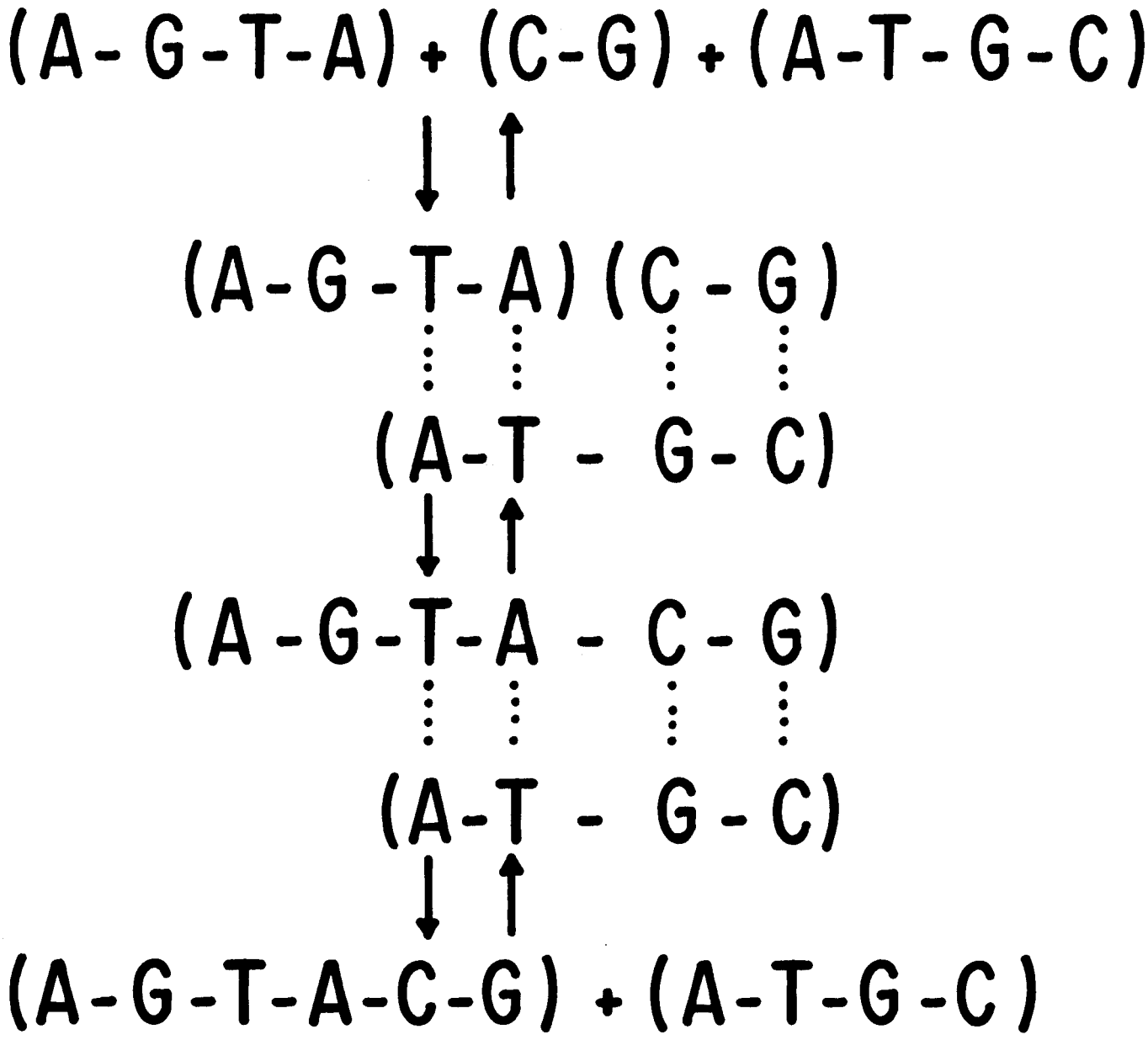


FIG.1

(a) A T (GROWING FRAGMENT)	WEIGHT	GROWTH EFFICIENCY I	GROWTH EFFICIENCY II	FUGACITY
A T	1/2	1	1	K
A T	1/2	1	1	K
A T A	1	1	1	K^2
T A T	1	1	1	K^2
T A T A	1/2	1	1	K^2
A T A T	1	1	2	K^2
T A T A	1/2	1	1	K^2

FIG. 2a

(b) A T A (GROWING FRAGMENT)	WEIGHT	GROWTH EFFICIENCY I	GROWTH EFFICIENCY II	FUGACITY
A T	1/2	1	1	K
T A	1/2	1	1	K
A T A	1	1	1	K^2
A T A	1	1	1	K^2
T A T A	1/2	1	1	K^2
A T A T	1	1	1	K^3
T A T A	1	1	1	K^3
A T A T	1/2	1	1	K^2
T A T A T	1	1	1	K^3
A T A T A	1	1	2	K^3
T A T A T	1	1	1	K^3
A T A T A T	1/2	1	1	K^3
T A T A T A	1	1	2	K^3
A T A T A T	1	1	2	K^3
T A T A T A	1/2	1	1	K^3

FIG.2b

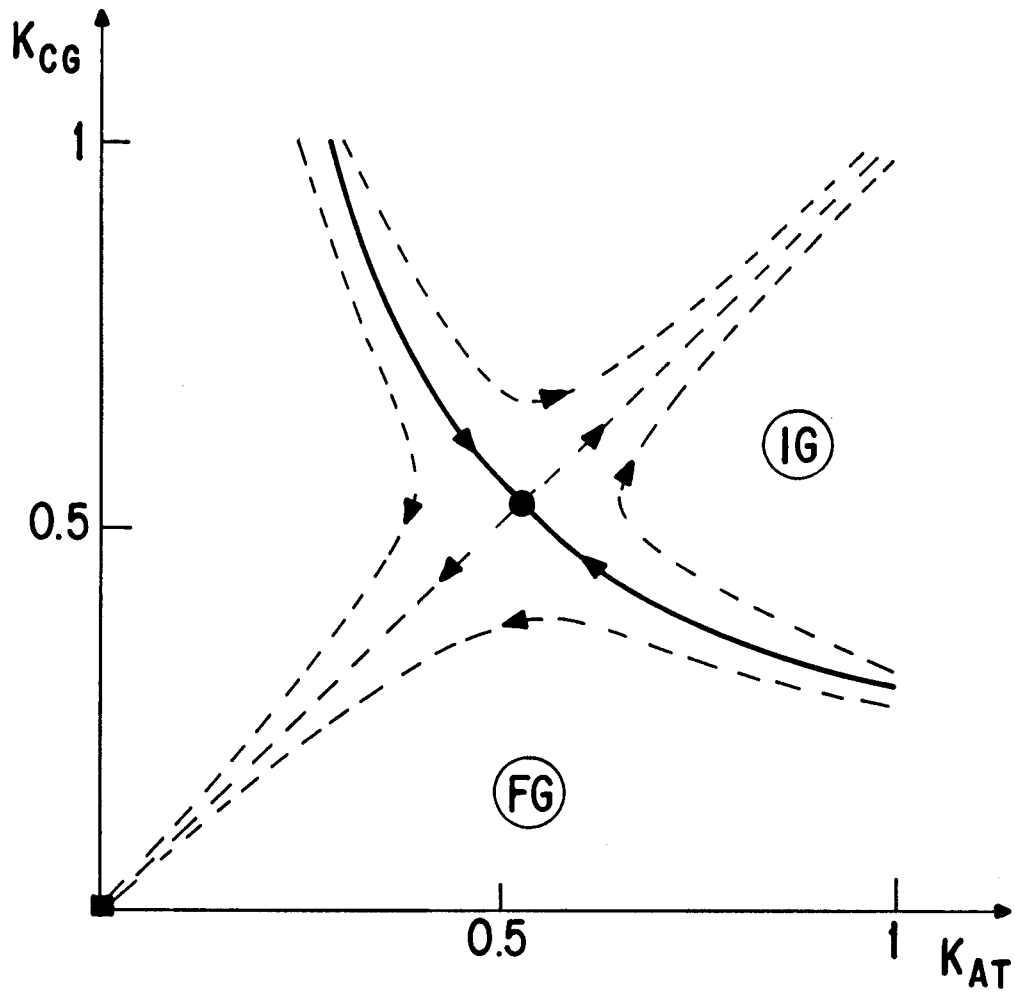


FIG. 3

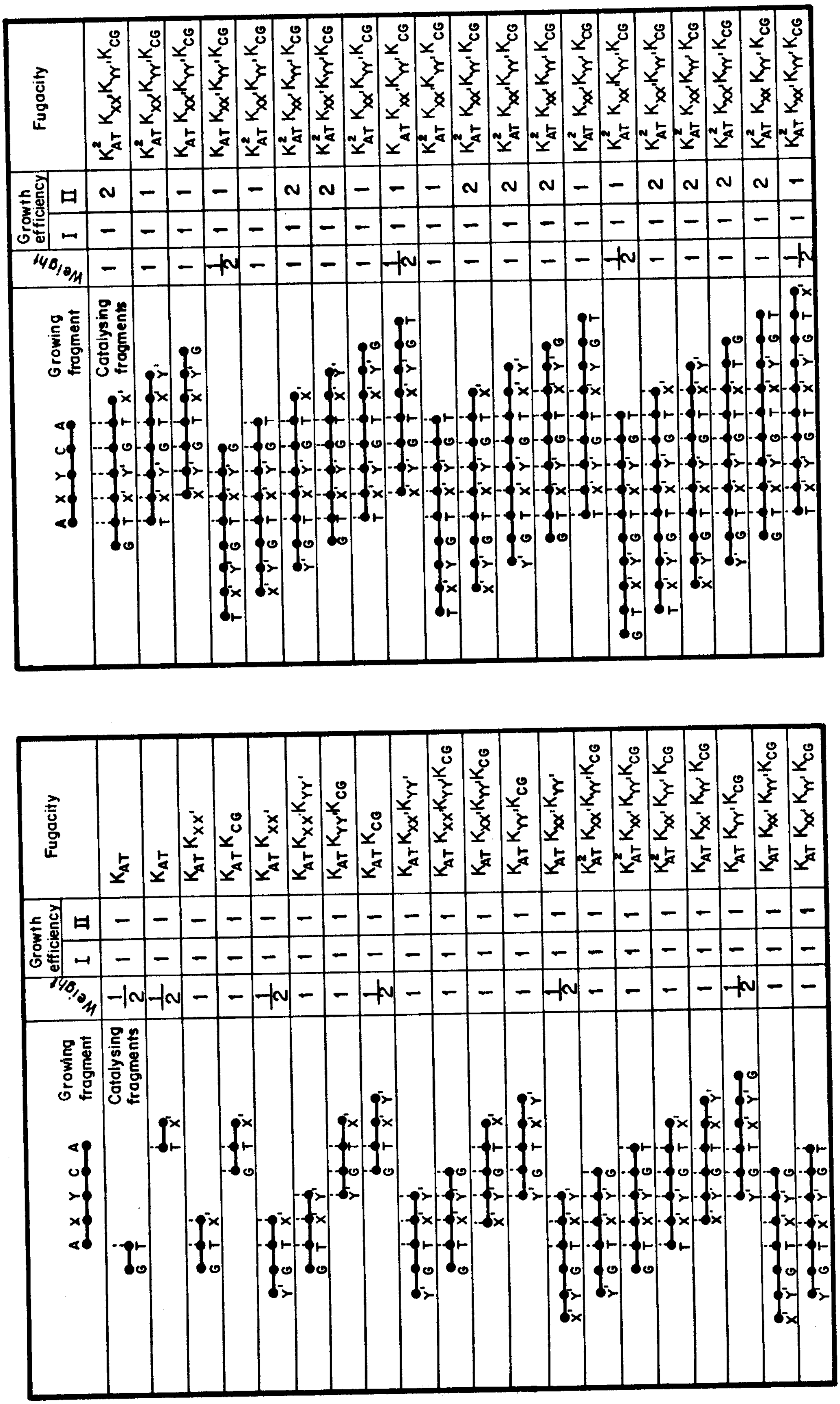

Growing fragment
Catalysing fragments
Weight
Growth efficiency
I
II
Fugacity

FIG. 4

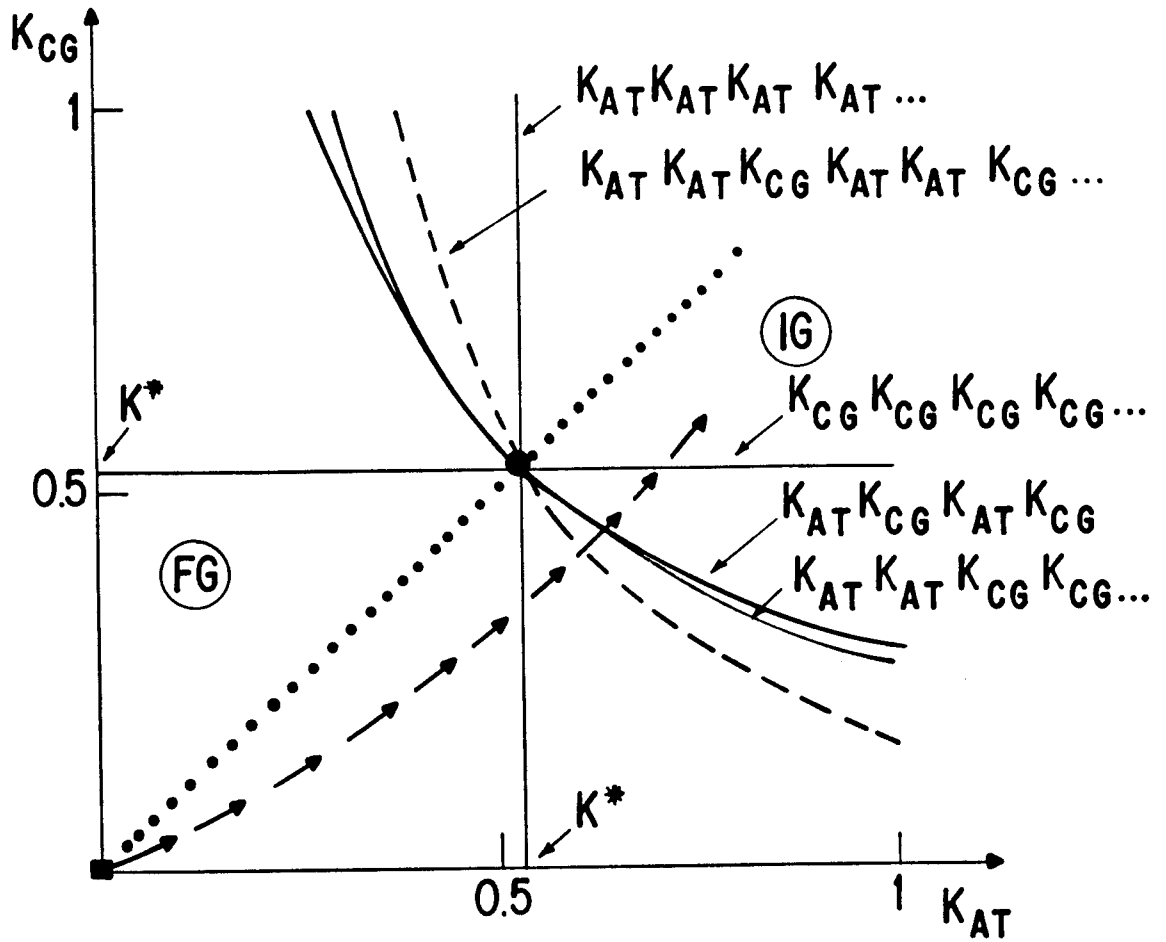

K_{CG}
1
0.5
K^*
FG
IG
$K_{AT}K_{AT}K_{AT}K_{AT}\ldots$
$K_{AT}K_{AT}K_{CG}K_{AT}K_{AT}K_{CG}\ldots$
$K_{CG}K_{CG}K_{CG}K_{CG}\ldots$
$K_{AT}K_{CG}K_{AT}K_{CG}$
$K_{AT}K_{AT}K_{CG}K_{CG}\ldots$
K^*
0.5
1
K_{AT}

FIG. 5

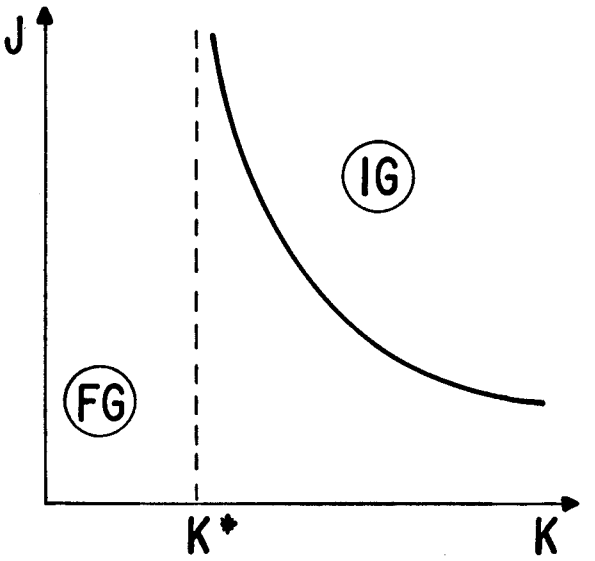


FIG. 6

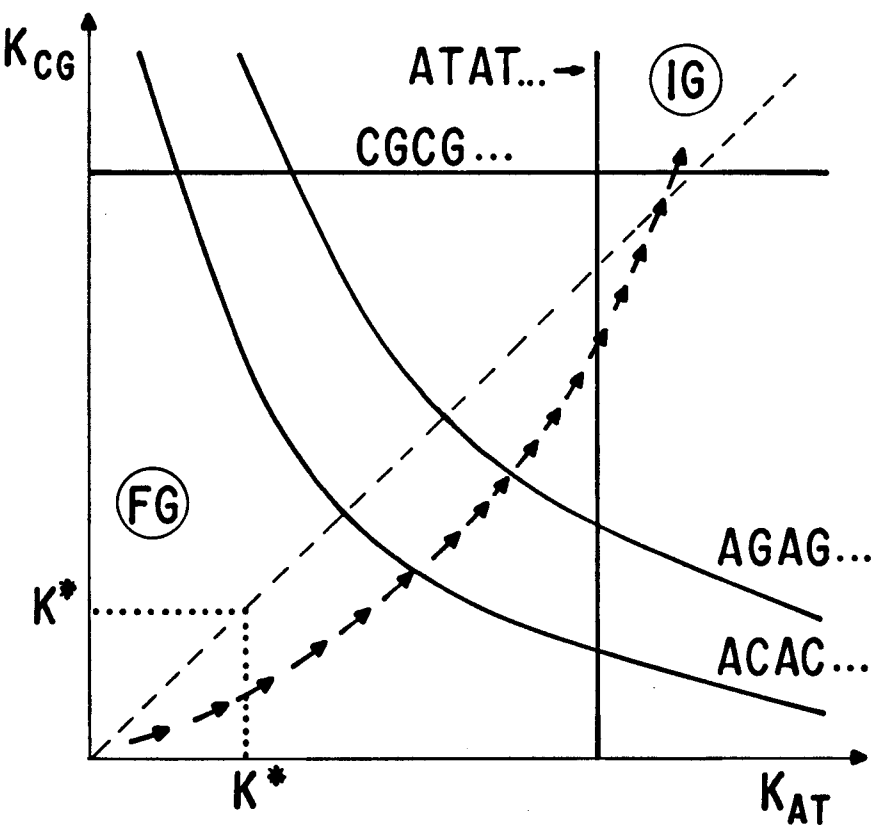
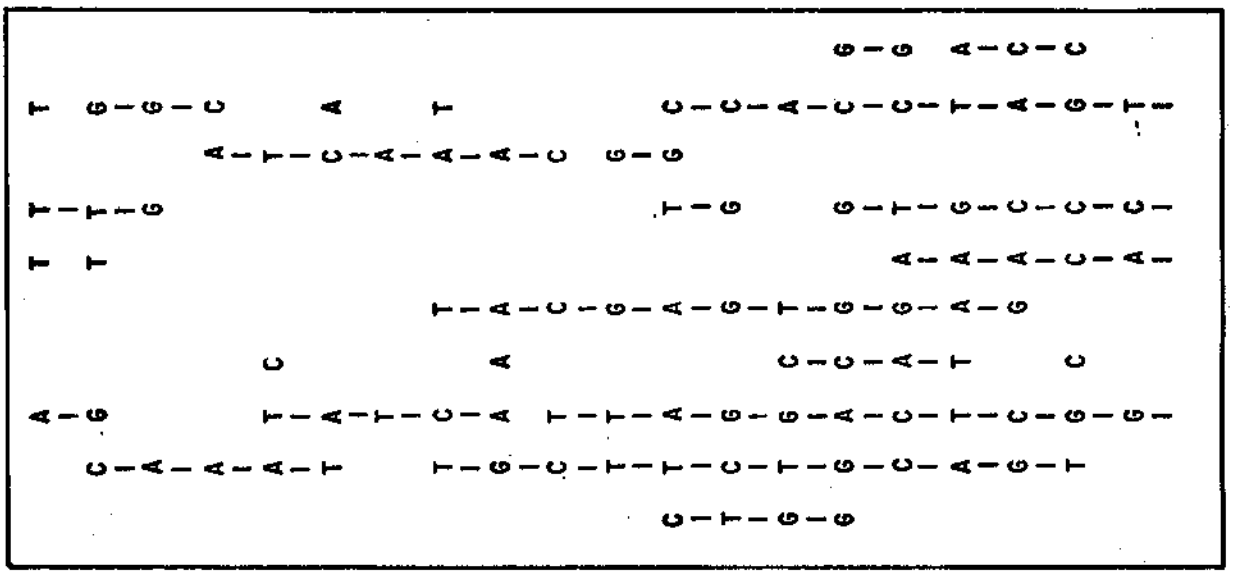
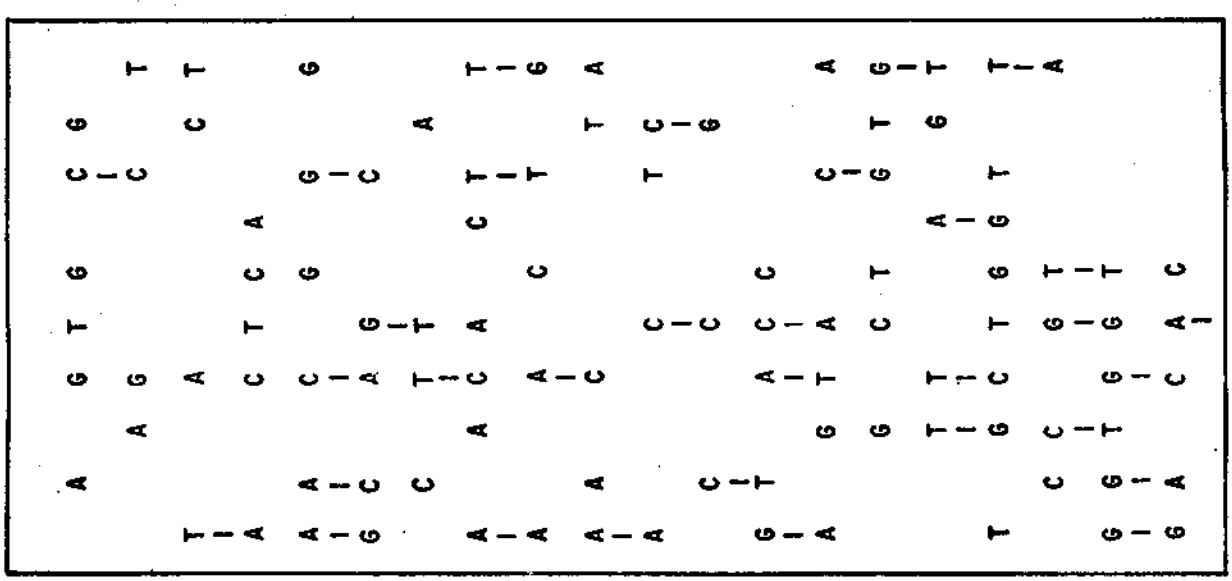


FIG. 7



(b)



(a)

FIG.8

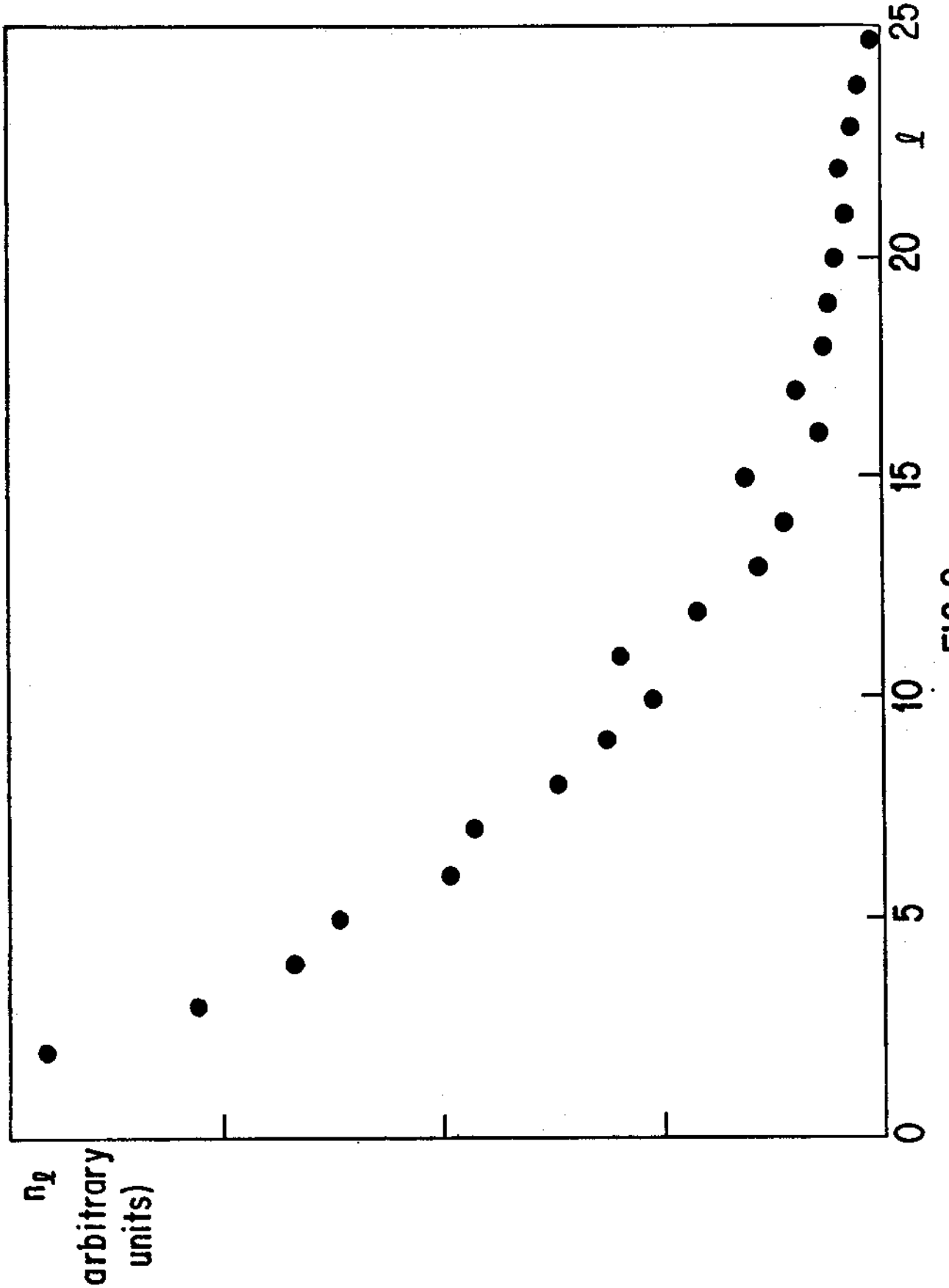


FIG.9

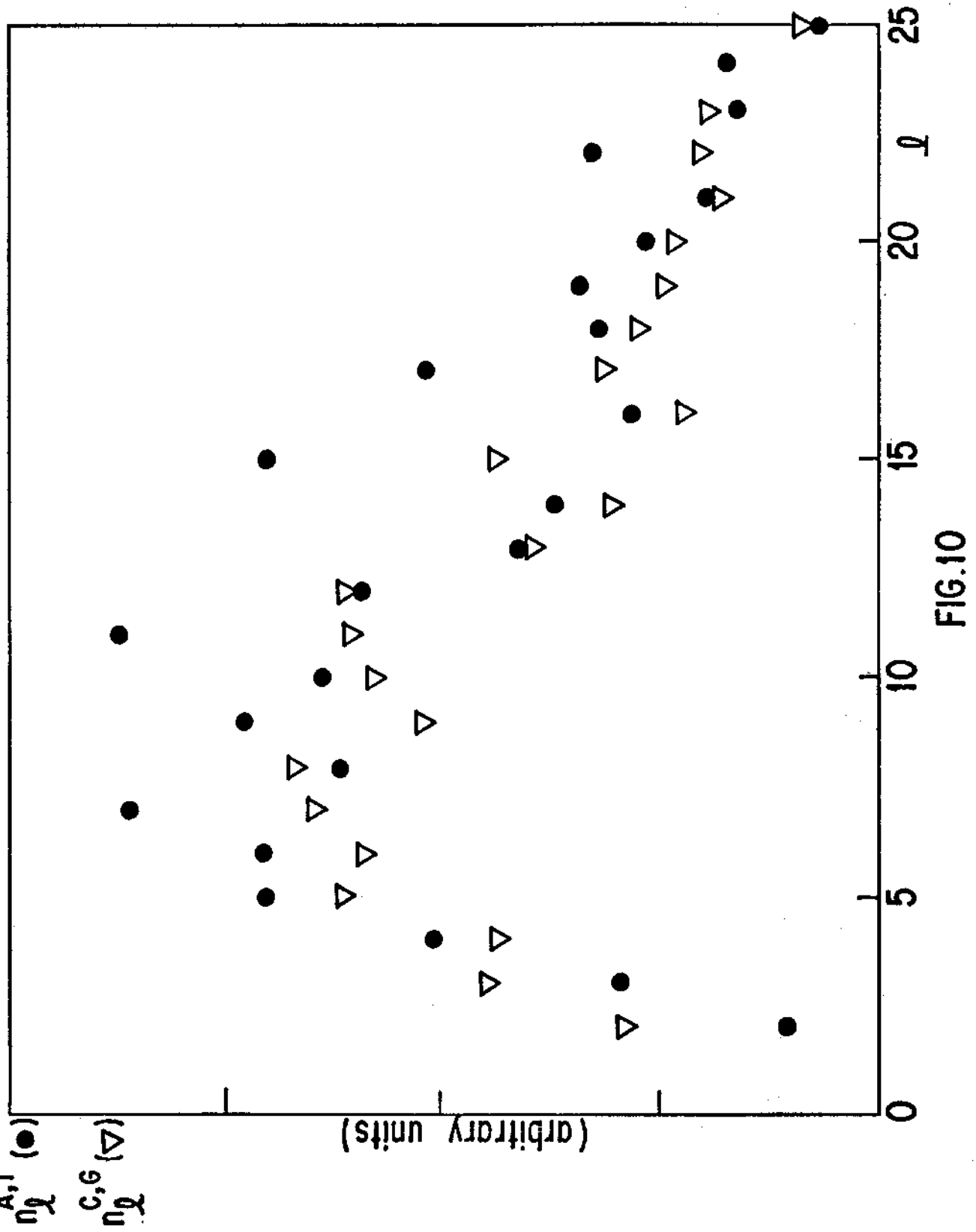


FIG.10

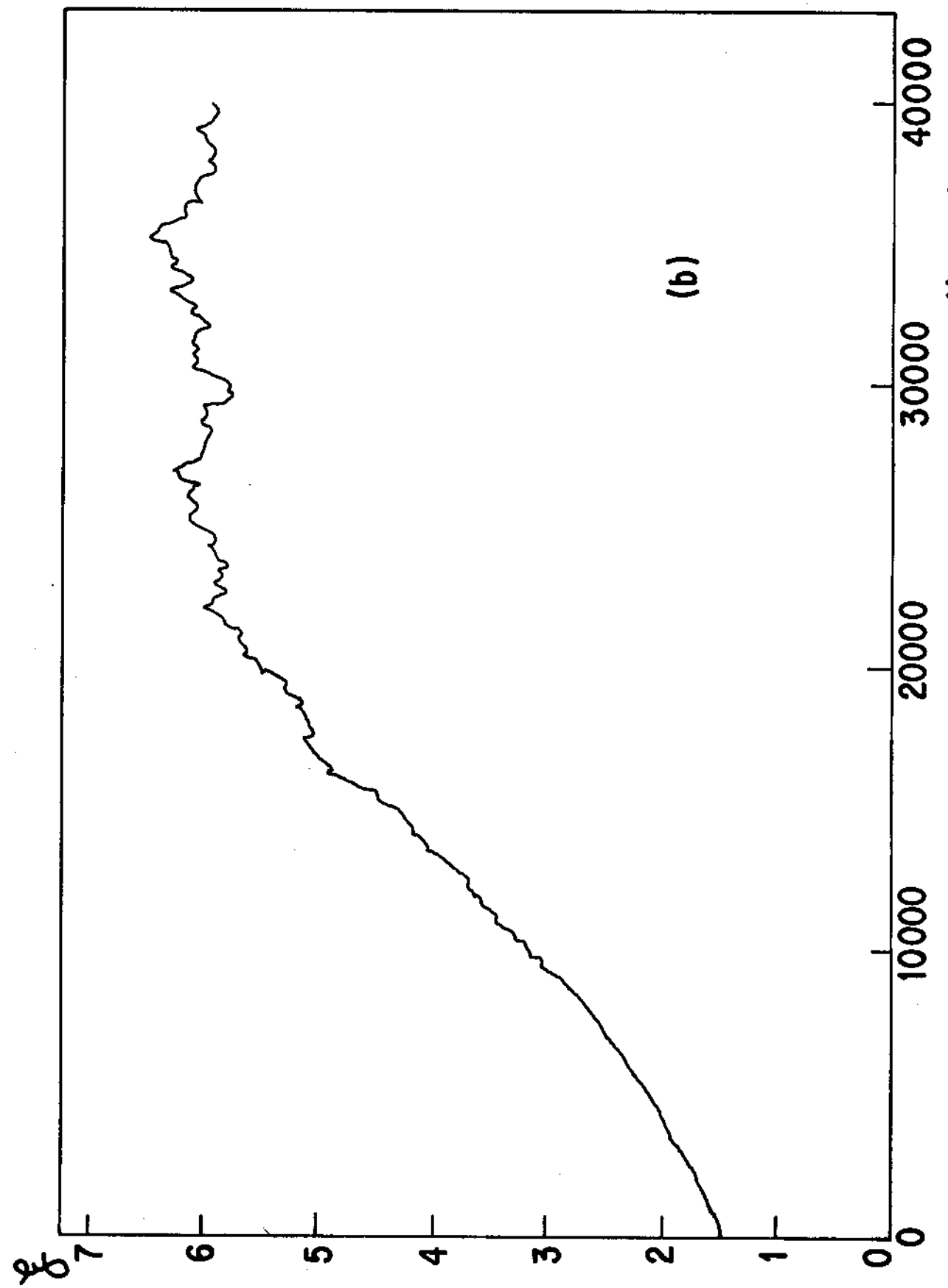


FIG.11

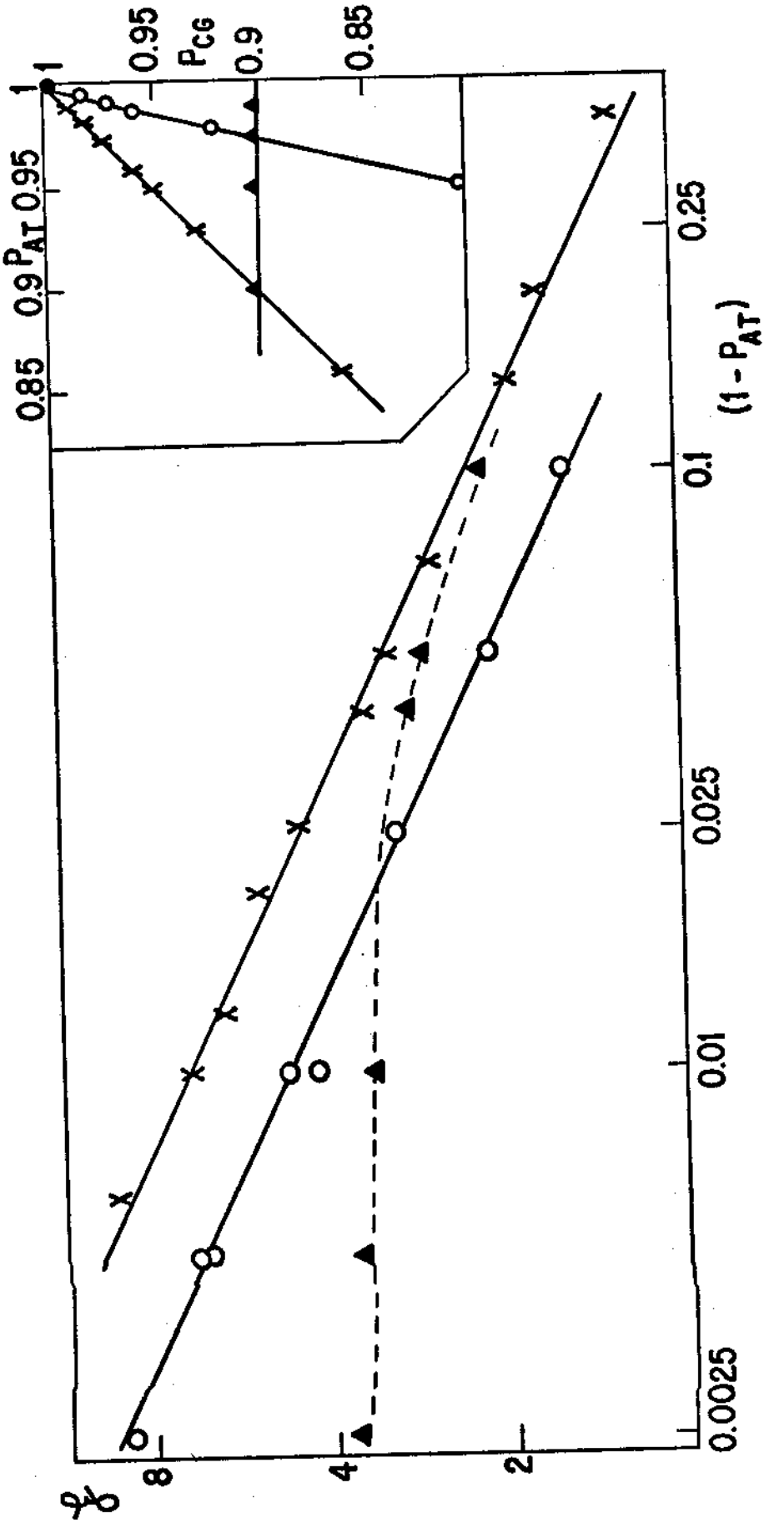


FIG.12

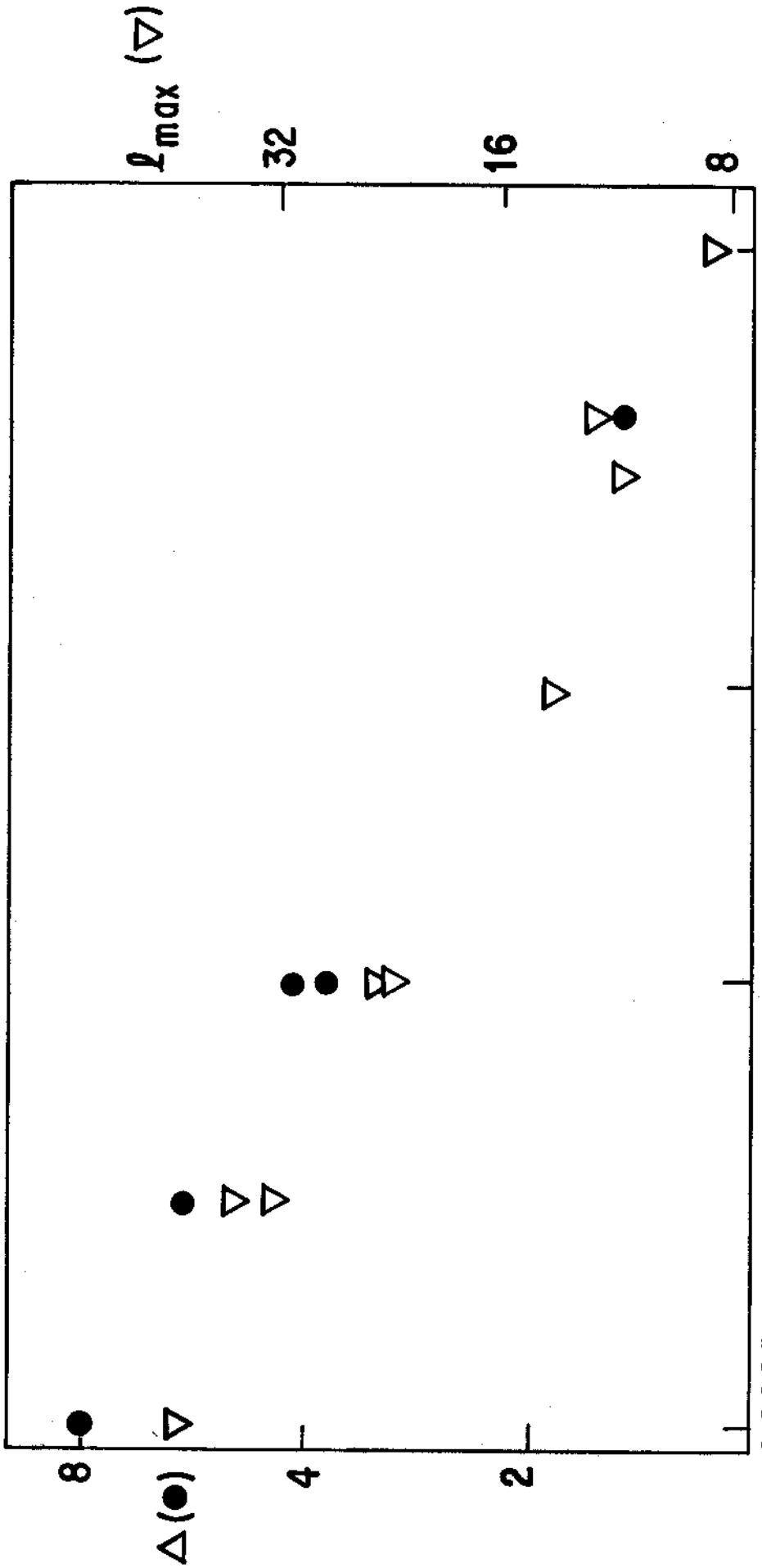


FIG.13

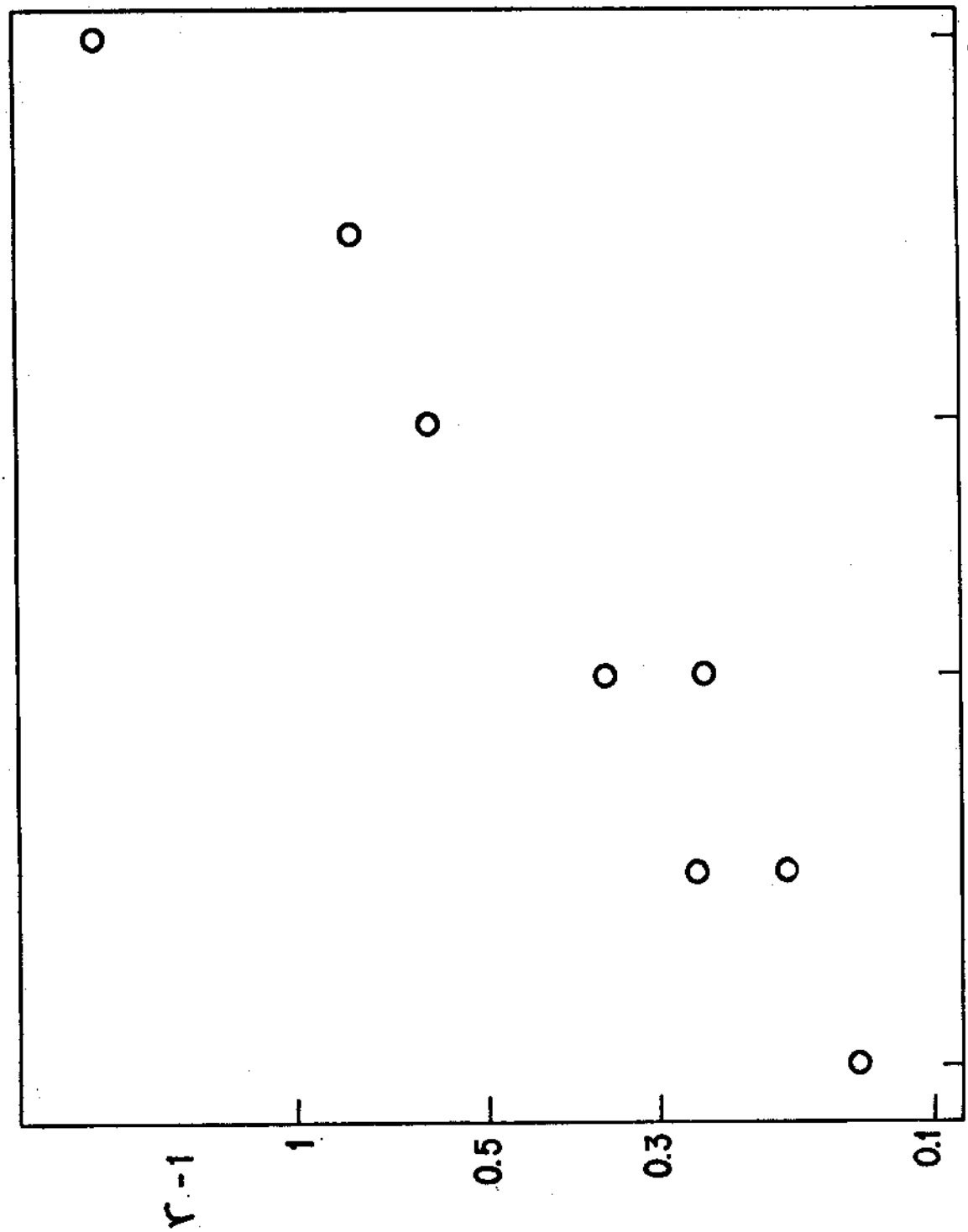


FIG.14

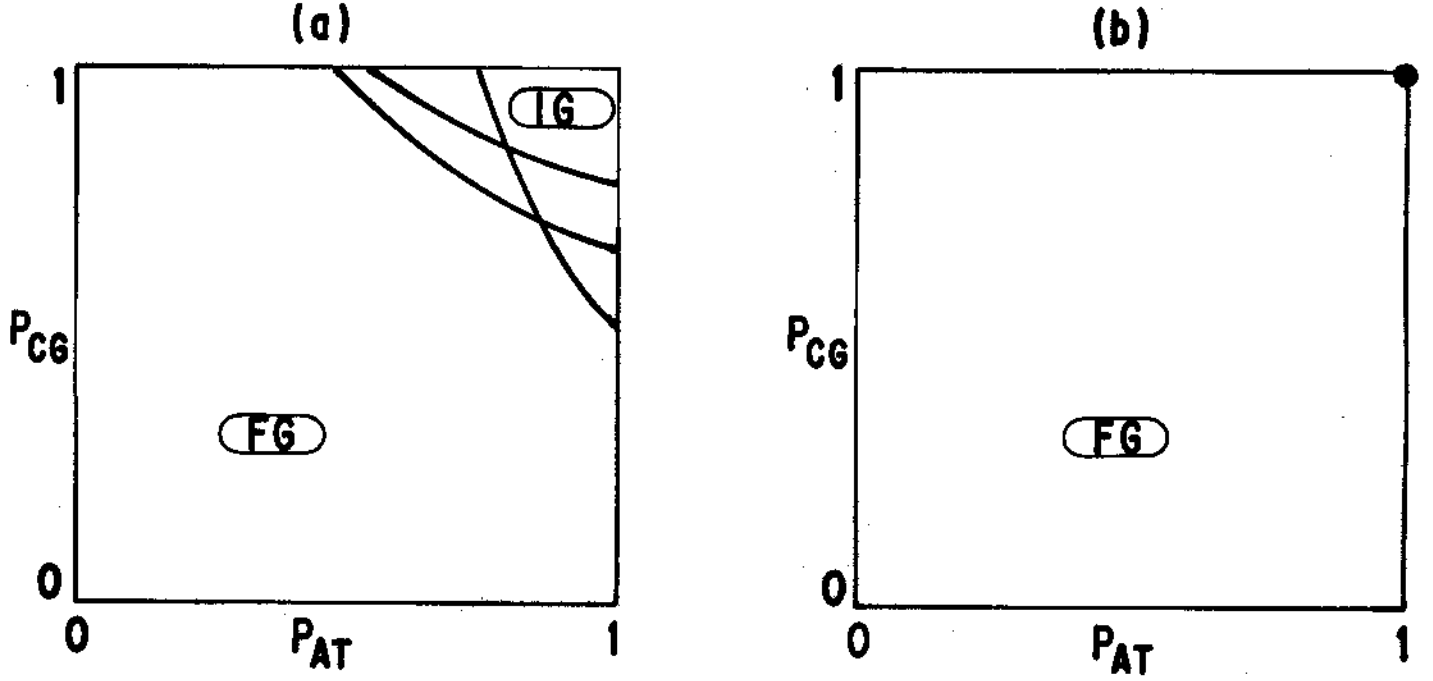


FIG. 15

REFERENCES

[1] R. Ferreira, Ciência e Cultura 37, 1745 (1985) [in portuguese].

[2] C. Tsallis and R. Ferreira, Phys. Lett. 99 A, 461 (1983).

[3] R. Ferreira and C. Tsallis, J. Theor. Biol. 117, 303 (1985).

[4] C. Tsallis and R. Ferreira, in "Statistical Physics" ed. H. E. Stanley (North-Holland, 1986) [Physica 140 A, 336 (1986)].

[5] H.J. Hermann and C. Tsallis, "Biogenesis and the growth of DNA-like polymer chains: a computer simulation", to be published (1987).

[6] C. Tsallis, "Biogenesis: diversity, selection and fractality", to appear in Quimica Nova (1988).

[7] J. Ninio and L.E. Orgel, J. Mol. Evol. 12, 91 (1978).

[8] J. Ninio, "Molecular Approaches to Evolution", Princeton University Press (1983).

[9] A.E. Oparin, "The Chemical Origin of Life", (Thomas, Springfield, 1964).

[10] S. Miller and L.E. Orgel, "Origin of Life on Earth" (Prentice Hall, Englewood Cliffs, 1974).

[11] G. Nicolis and I. Prigogine, "Self-organization in non-equilibrium systems" (Wiley, New York, 1977).

[12] M. Eigen and P. Schuster, Naturwiss 64, 541 (1977).

[13] M. Eigen and P. Schuster, Naturwiss 65, 341 (1978).

[14] F. Dyson, J. Mol. Evol. 18, 344 (1982).

[15] P.W. Anderson, Proc. Natn. Acad. Sci. USA 80, 3386 (1983).

[16] L. Demetrius, Proc. Natn. Acad. Sci. USA 81, 6068 (1984).

[17] P.W. Anderson and D. Stein, Proc. Natn. Acad. Sci. USA 81, 1751 (1984).

[18] P.W. Anderson and D. Stein, "Self-organizing System: The Emergence of Order", ed. F.E. Yates (Plenum, New York, 1985).

[19] P.W. Anderson "Computer modeling of prebiotic evolution: general theoretical ideas on the origin of biological information" (1985), to appear in Comments on Molecular and Cellular Biophysics (Gordon and Breach Science Publishers).

[20] J.D. Watson, "Molecular Biology of the Gene", 2nd. edn (Benjamin, New York, 1970).

[21] M. Eigen, Naturwiss. 58, 465 (1971).

[22] H. Kuhn, Angew. Chem. 84, 838 (1972).

[23] R. Lohrmann and L.E. Orgel, J. Mol. Biol. 12, 237 (1979).

[24] H. Kuhn and J. Waser, Angew. Chem. 93, 495 (1981).

[25] H. Kuhn and J. Waser, in "Biophysics", Hoppe, Lohrmann, Markel and Ziegler, pp. 830-847 (Springer, Berlin, 1983).

[26] A.G. Cairns-Smith, J. Theor. Biol. 10, 53 (1966).

[27] S.W. Fox and K. Dose, "Molecular Evolution and the Origin of Life", (M. Dekker, New York, 1977).

[28] R.S. Root - Bernstein, J. Theor. Biol. 94, 895 (1982).

[29] K.G. Wilson and J. Kogut, Phys. Rep. 12 C, 75 (1974).

[30] G.Toulouse and P. Pfeuty, "Introduction au Groupe de Renormalisation et à ses Applications" (Presses Universitaires, Grenoble, 1975).

[31] S.L.A. de Queiroz and C.M. Chaves, Z. Phys. B 40, 99 (1980).

[32] H. Gould, F. Family and H.E. Stanley, Phys. Rev. Lett. 50, 686 (1982).

[33] H.E. Stanley, F. Family and H. Gould, J. Poly. Sci.: Polymer Symposium 73 (1985).

[34] H.J. Herrmann, Phys. Rep. 136, 153 (1986).

[35] M. Kolb, Phys. Rev. Lett. 53, 1653 (1984).

[36] P. Meakin, T. Vicsek and F. Family, Phys. Rev. B 31, 564 (1985).

[37] S. Wolfram, Physica 10 D, 1 (1984).

[38] L. Landau and E. Lifchitz, "Physique Statistique" (Mir, Moscow, 1967), page 576.

[39] J.S. Helman, A. Coniglio and C. Tsallis, Phys. Rev. Lett. 53, 1195 (1984) and 54 , 1735 (1985).