ISSN 0029 - 3865

CBPF-NF-022/87

BIOGENESIS AND THE GROWTH OF DNA-LIKE POLYMER CHAINS: A COMPUTER SIMULATION

by

Hans J. HERRMANN* and Constantino TSALLIS

Centro Brasileiro de Pesquisas Físicas - CBPF/CNPq
Rua Dr. Xavier Sigaud, 150
22290 - Rio de Janeiro, Brazil

*Permanent address: Service de Physique Théorique, CEN Saclay, F-91191, Gif-sur-Yvette, France.

ABSTRACT

We study, through computer simulation, a crucial step of Biogenesis, namely the growth of self-replicating codified DNA-like polymers starting from a mixture of oligomers. We have adopted the growth scheme that has been recently proposed by Ferreira and Tsallis which incorporates usual ideas of autocatalysis through complementary pairs and within which a central role is played by the hydrogen-like links (characterized by the probabilities p_{AT} and p_{CG} of chemical bonding of the A-T and C-G pairs respectively) between the two chains of the growing polymer. We find that the average equilibrium polymeric length ξ diverges, for any fixed ratio $(1-p_{AT})/(1-p_{CG})$, as $\xi \propto 1/\sqrt{1-p_{AT}}$. Selection of patterns may happen at all stages and in particular at chemical equilibrium. Selection occurs via two different mechanisms: (i) away from the critical point $p_{AT} = p_{CG} = 1$ if $p_{AT} \neq p_{CG}$; (ii) both on and away from the critical point if the initial concentrations of nucleotides (A,T, C and G or their precursors) are different.

Key-words: Biogenesis; Growth; DNA polymer; Computer simulation.

1 INTRODUCTION

How did life appear on Earth? This question is so fundamental that it is worthwhile investigating every possible mechanism that can lead to an answer [1-4]. A crucial step of Biogenesis is the growth of self-replicating codified DNA-like polymers starting from a mixture of oligomers (monomers, dimers, trimers, etc.) [5-12]. By "DNA -like" we refer to double-chains of nucleotides like adenine (A), thymine (T), cytosine (C) and guanine (G). For the present purposes, the relevant links between the two chains are assumed to respect Crick and Watson-like complementarity [13], i.e., they occur between complementary pairs (A-T and C-G); the bond within each of these pairs is assumed to be a hydrogen-bridge (or a similar linking). Let us right now make it clear that, when we use the notation DNA, A, T, C and G, we do not necessarily refer to the well known macromolecule and its nucleotides; the notation might as well refer to their *precursors* (e.g., if the basic molecule happens to be RNA-like [2,14-18] rather than DNA, the notation A-T would then refer to the A-U pair).

Various mechanisms (clay [19], protein [20] and aminoacid pairing [21]) have been proposed for the growth of the polymers we are interested in. However, autocatalysis through A-T and C-G pairings occuring in a prebiotic or *primordial* soup seems the most appealing to us [2,8,10-12,14-18]; we therefore adopt it in the present work. More precisely, we formulate our (computational) approach within the scheme recently introduced by Ferreira and Tsallis [22]. Within this scheme, the prebiotic polymeric growth would have occured as a critical phenomenon controlled by the A-T and C-G bond fugacities and essentially occuring at thermodynamical quasi-equilibrium. The covalent (or covalent-like) bonds *along* the chains of the DNA-like molecule are considered quite rigid,

and only become chemically active in auto-catalytic complementary configurations (see Fig. 1). This growth mechanism being very close to that which is responsible for the replication of DNA double-chains seems biologically very sound. Ferreira and Tsallis have shown that different values for the A-T and C-G fugacities generate, in what concerns the sequences of nucleotides or "codes" of the macromolecules, *selection within the diversity*, basic condition for life.

The present work is devoted to the computer simulation of such a system, in order to exhibit and better understand its various ingredients. We exactly define our model and algorithm in Section 2; the body of the results is presented in Section 3.

2 MODEL AND ALGORITHM

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 probabi-

lity of chemical bonding of the A-T and C-G pairs between the two lattices.

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:

$$N_A^{tot} = N_A + 2N_{AA} + N_{AT} + N_{AC} + N_{AG} \qquad (1.a)$$

$$N_T^{tot} = N_T + N_{AT} + 2N_{TT} + N_{TC} + N_{TG} \qquad (1.b)$$

$$N_C^{tot} = N_C + N_{AC} + N_{TC} + 2N_{CC} + N_{CG} \qquad (1.c)$$

$$N_G^{tot} = N_G + N_{AG} + N_{TC} + N_{TG} + 2N_{GG} \qquad (1.d)$$

and

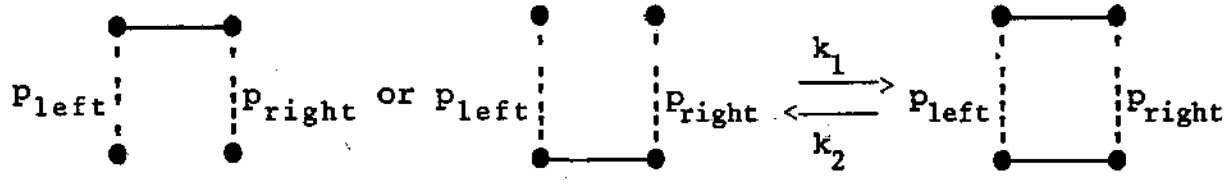
$$N_A^{tot} + N_T^{tot} + N_C^{tot} + N_G^{tot} \leq L_x L_y \qquad (2)$$

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:



(3)

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

$$\frac{k_1}{k_2} = \frac{2p_{left}p_{right}}{p_{left}(1-p_{right}) + p_{right}(1-p_{left}) + (1-p_{left})(1-p_{right})} = \frac{2p_{left}p_{right}}{1-p_{left}p_{right}} \tag{4}$$

where we have assumed independence between the "left" and "right" hydrogen-like bonds linking complementary nucleotides. At the left hand side of Eq. (3) 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 every 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. 2 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. 3 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 [23,24]. 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 [25] 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

$$f_{A,T} = \frac{N_{AA} + N_{AT} + N_{TT}}{N_{AA} + N_{AT} + N_{TT} + N_{CC} + N_{CG} + N_{GG} + N_{AC} + N_{AG} + N_{TC} + N_{TG}} \qquad (5.a)$$

$$f_{C,G} = \frac{N_{CC} + N_{CG} + N_{GG}}{N_{AA} + N_{AT} + N_{TT} + N_{CC} + N_{CG} + N_{GG} + N_{AC} + N_{AG} + N_{TC} + N_{TG}} \qquad (5.b)$$

$$f_{A,C} = \frac{N_{AC} + N_{AG} + N_{TC} + N_{TG}}{N_{AA} + N_{AT} + N_{TT} + N_{CC} + N_{CG} + N_{GG} + N_{AC} + N_{AG} + N_{TC} + N_{TG}} \qquad (5.c)$$

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. 4, $n_\ell^{A,T}$ and $n_\ell^{C,G}$ obtained for a specific choice of parameters. Clearly, the two quantities behave 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.

3 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. 5 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. 5 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. 5(a) and 5(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. 5(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 Ferreira and Tsallis picture for this prebiotic stage is correct, only approaching such a critical point one can generate arbitrarily long DNA chains. The most relevant parameters of our model are p_{AT} and p_{CG} since 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. 6 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 can speculate that, if our polymer model had an effective topological dimensionality higher than one (due, for instance, to cross-links in a folded double-chain), we would have critical *lines* in the (p_{AT},p_{CG}) plane (like those appearing in Refs. [22]) instead of a unique critical point $p_{AT}=p_{CG}=1$. But coming back to our one-dimensional model, 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,

$$\xi \propto (1-p_{AT})^{-x} \tag{6}$$

because the points in Fig. 6 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. (6) could well be a square-root behaviour: $x=1/2$. The proportionality constant of Eq. (4) 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

$$\xi = \frac{A}{[(1-p_{AT})^2+(1-p_{CG})^2]^{1/4}} \tag{7}$$

where A is of order unity.

The average longest chain ℓ_{max} and the fluctuations Δ of the chain length also diverge with a powerlaw like that of Eq. (6) with the same exponent x within the statistical error bars. This is seen in Fig. 7 for one of the paths of Fig. 6 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. 8, 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. 8 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 selec

tion 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±1, 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.

4 CONCLUSION

Following along the lines of the Ferreira and Tsallis scheme for the growth of self-replicating codified DNA-like polymers in a prebiotic soup, we have simulated the phenomenon by a 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 occured 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 ac-

count 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 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 number of the different species of nucleotides is 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 our work 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. (7) is also

needed. Finally, would it be interesting to include some more realistic features into our model.

Fruitful discussions with R. Ferreira and A.O. Caride at the early stages of this work are gratefully acknowledged. One of us (H.J.H.) has benefited from a Guggenheim Fellowship.

FIGURE CAPTIONS

Fig. 1 - Example for the formation of a covalent bond (———) between two chains through a third chain (G-A) which is attached by hydrogen-like bondings (....); $p_{AT}(p_{CG})$ denote the probability of the chemical bond A-T (C-G) to occur.

Fig. 2 - 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 $(L_x,L_y)=(20,10)$.

Fig. 3 - 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 $L_x=160$, $L_y=10$, $N_A^{tot}=N_C^{tot}=416$, $N_{AA}=N_{CC}=N_{AC}/2=16$, $m_x=m_y=1$.

Fig. 4 - 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. 3.

Fig. 5 - Histogram of ξ obtained with: (a) the same parameters of Fig. 3; (b) the same parameters of Fig. 3 except for $(m_x,m_y)=(1,9)$.

Fig. 6 - 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 $N_A^{tot}=0.26\ L_xL_y$.

Fig. 7 - 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. 6.

Fig. 8 - Log-log plot of r-1 against $1-p_{AT}$ along the same path and starting conditions of Fig. 7. For two values of p_{AT} we have represented two different statistical samples to exhibit the magnitude of the fluctuations.

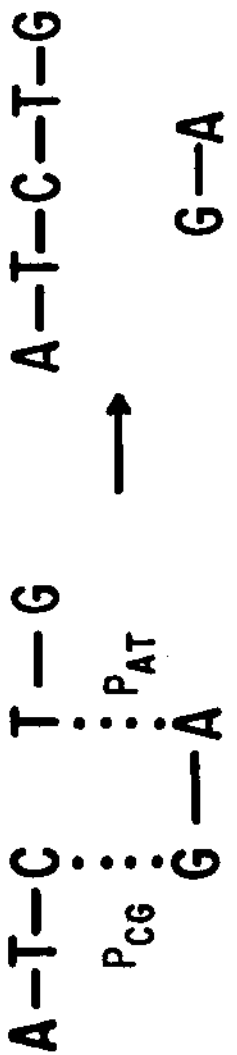
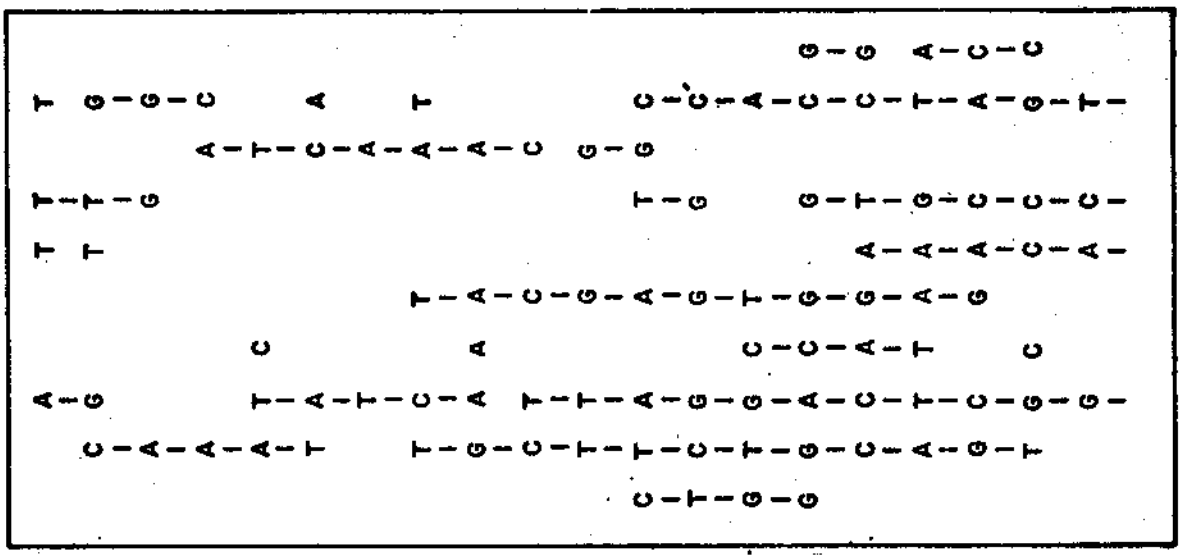
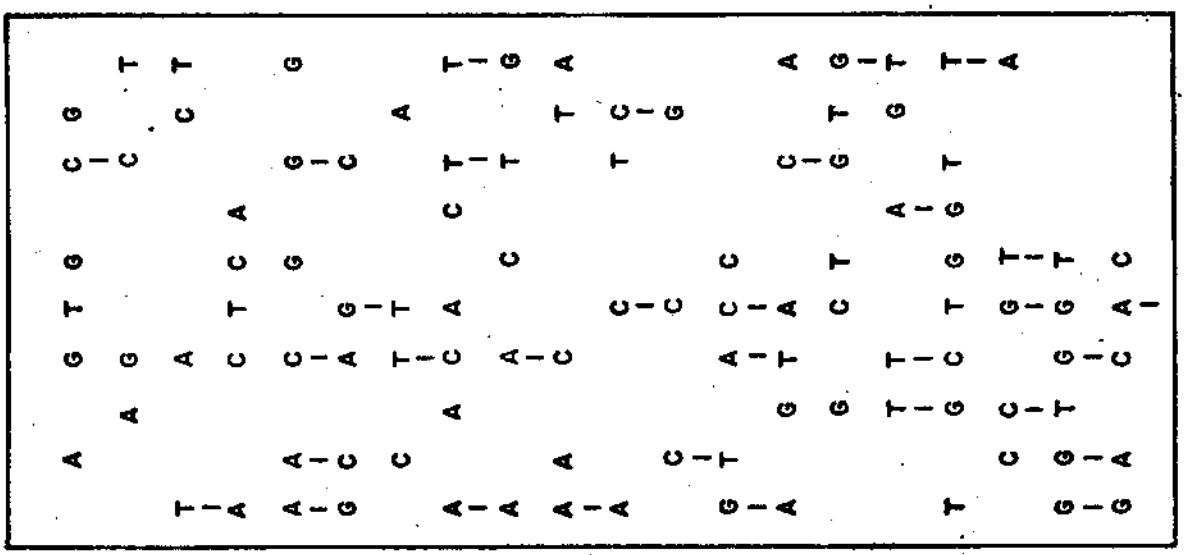


FIG.1



(b)



(a)

FIG. 2

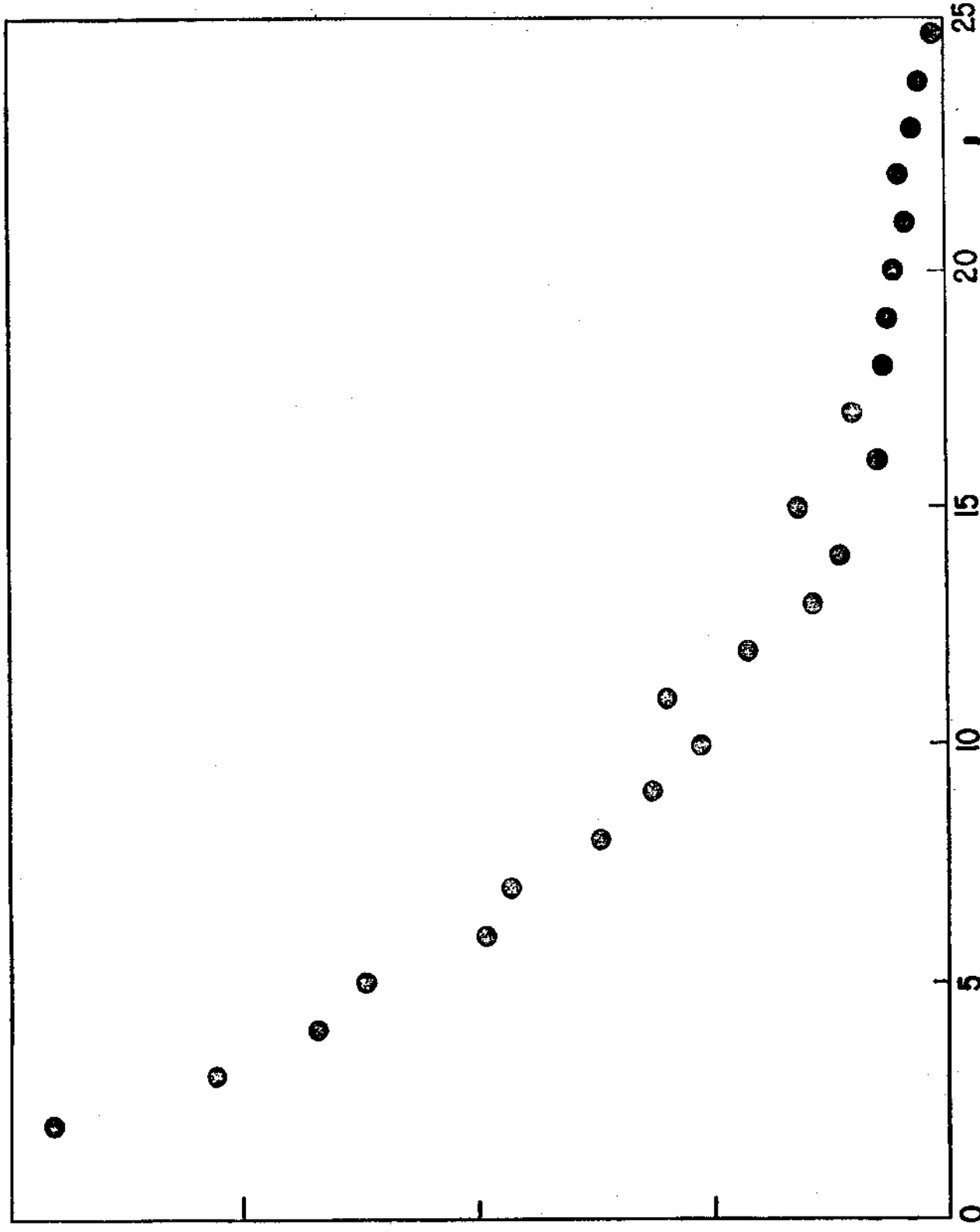


FIG. 3

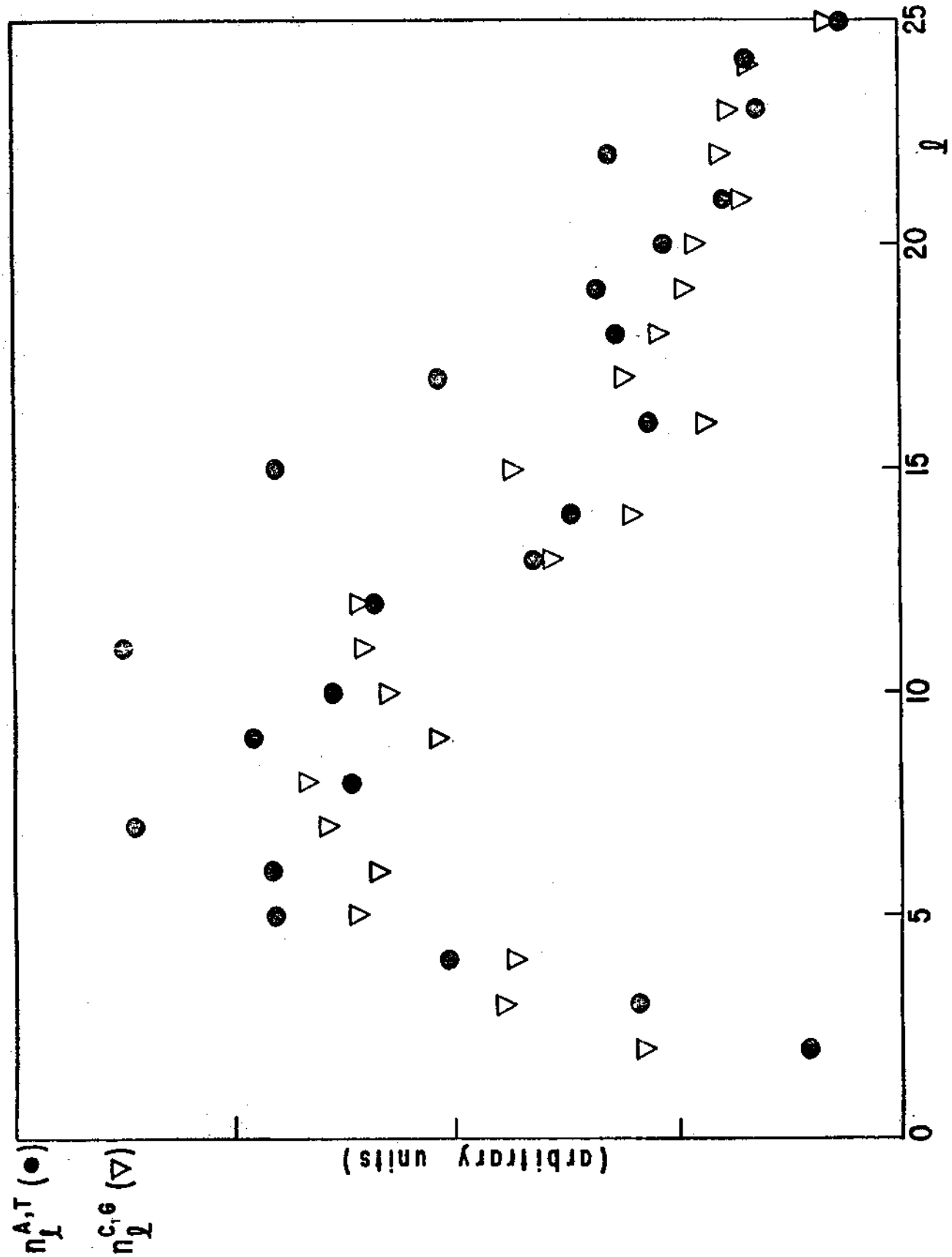


FIG. 4

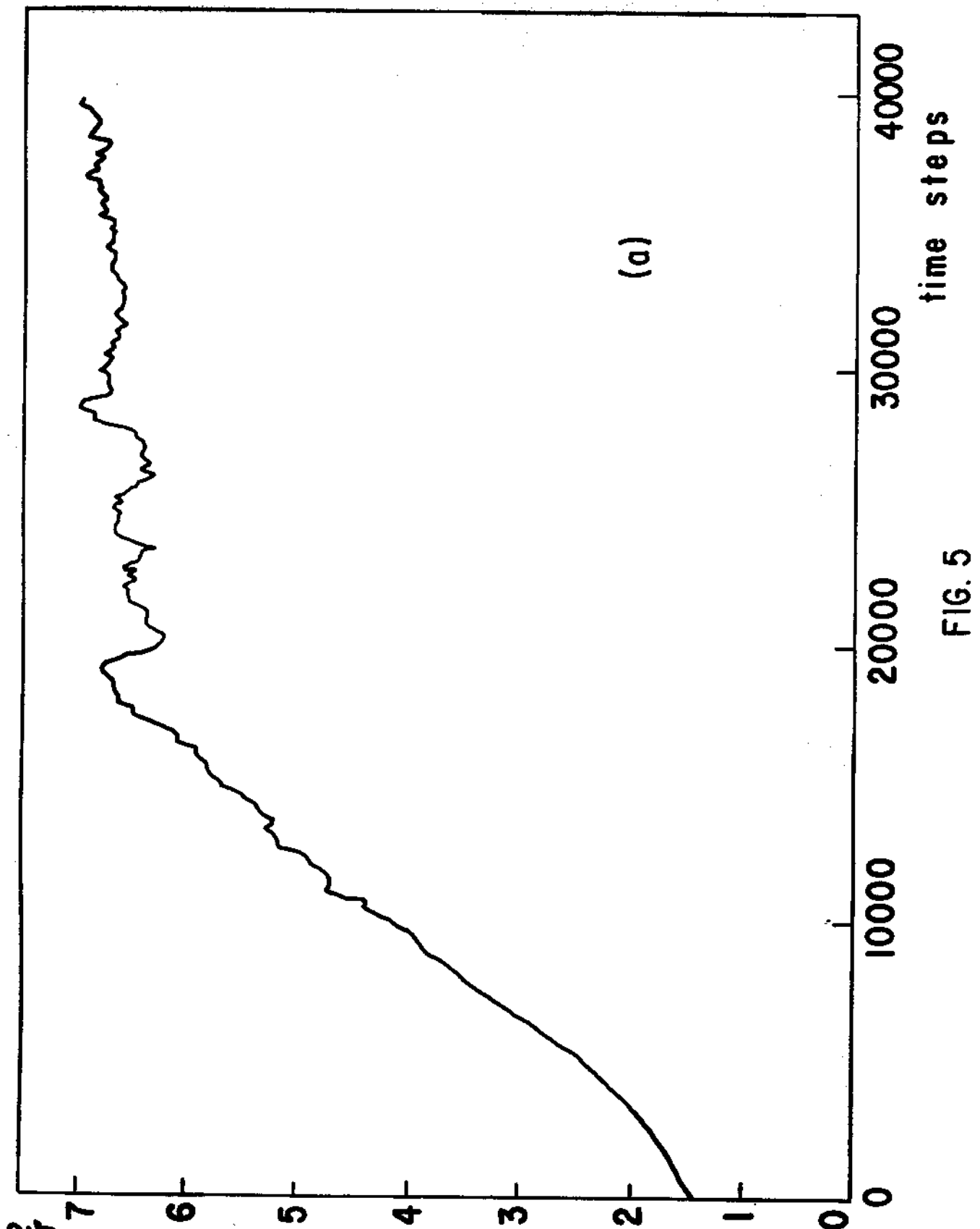


FIG. 5

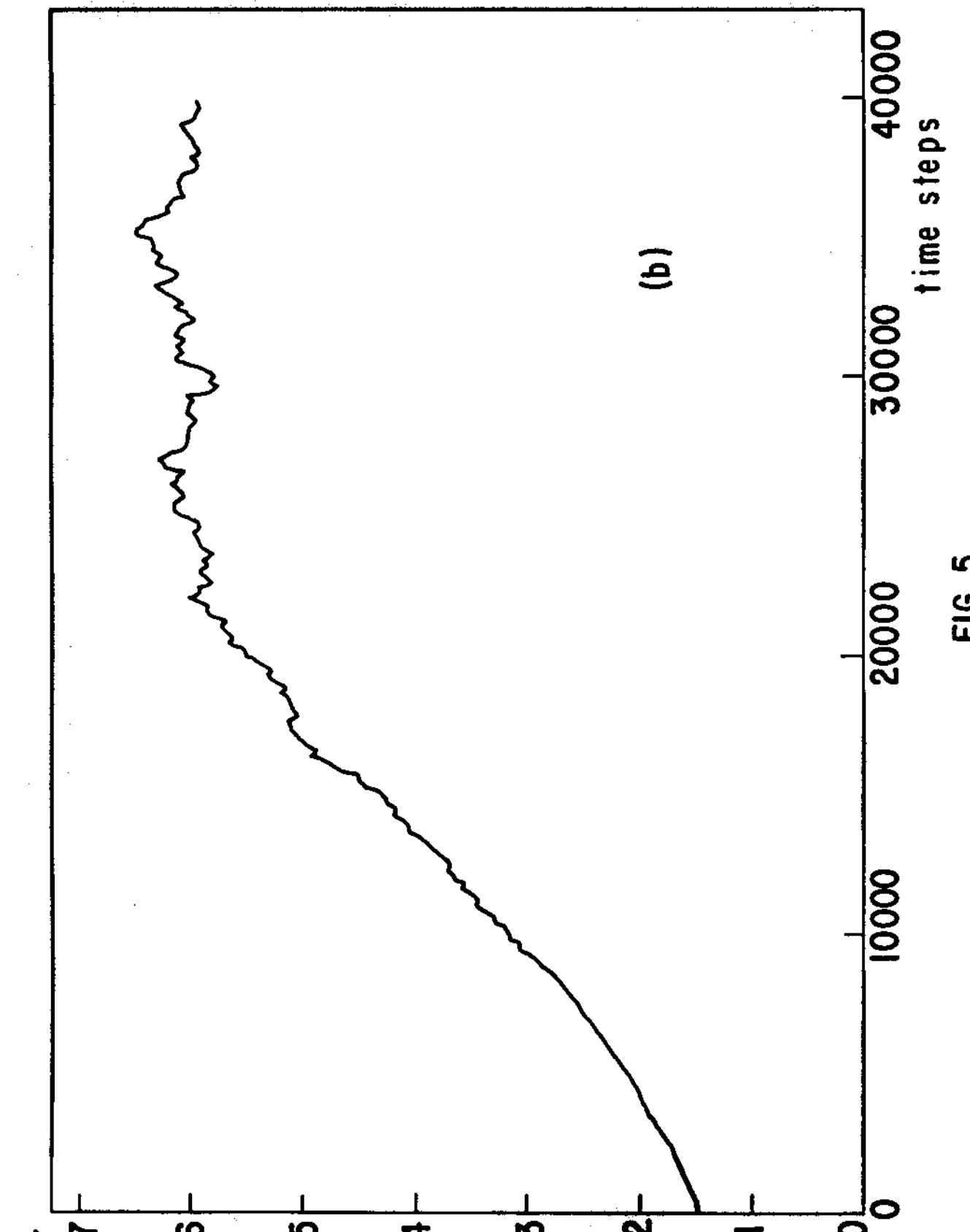


FIG. 5

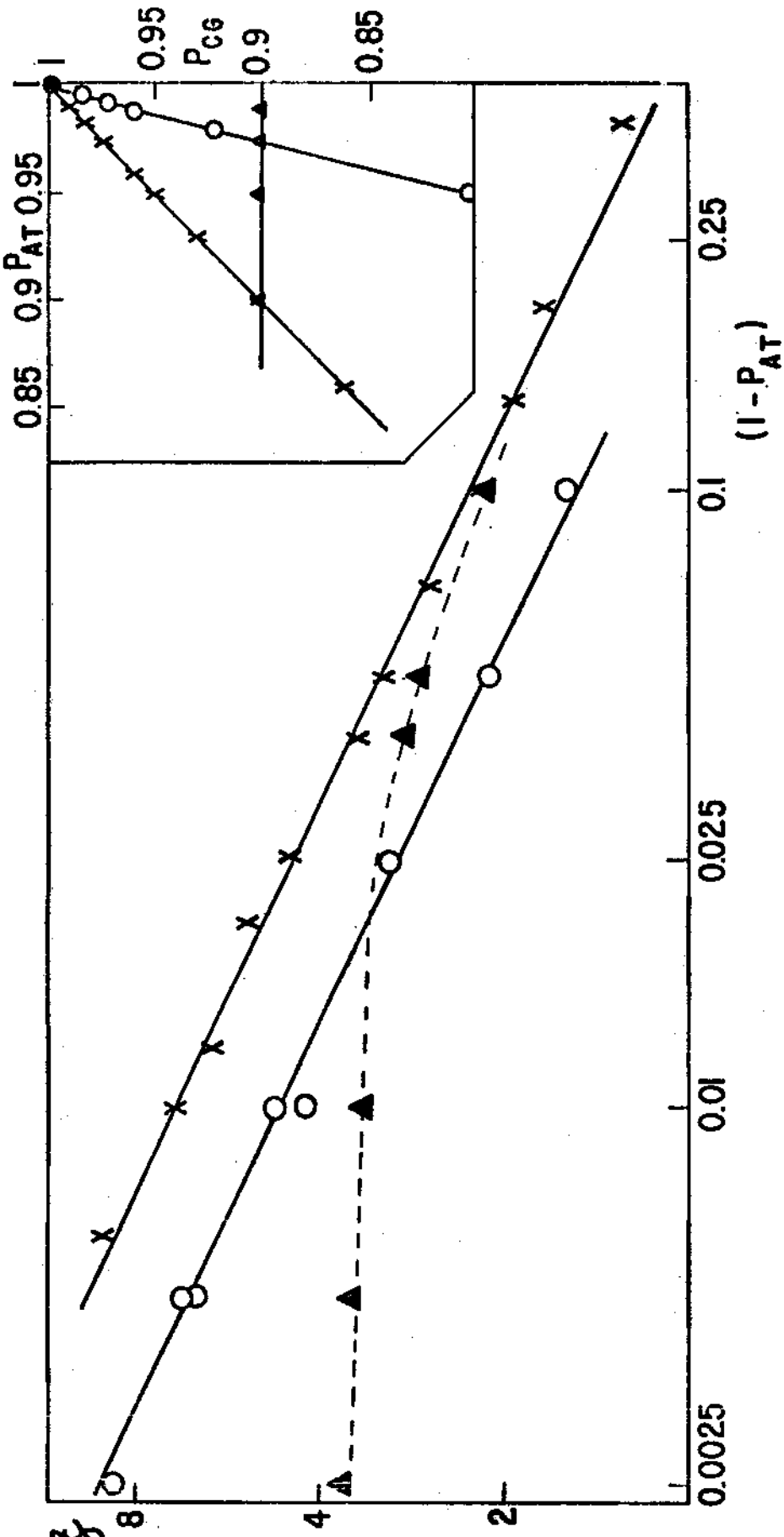


FIG.6

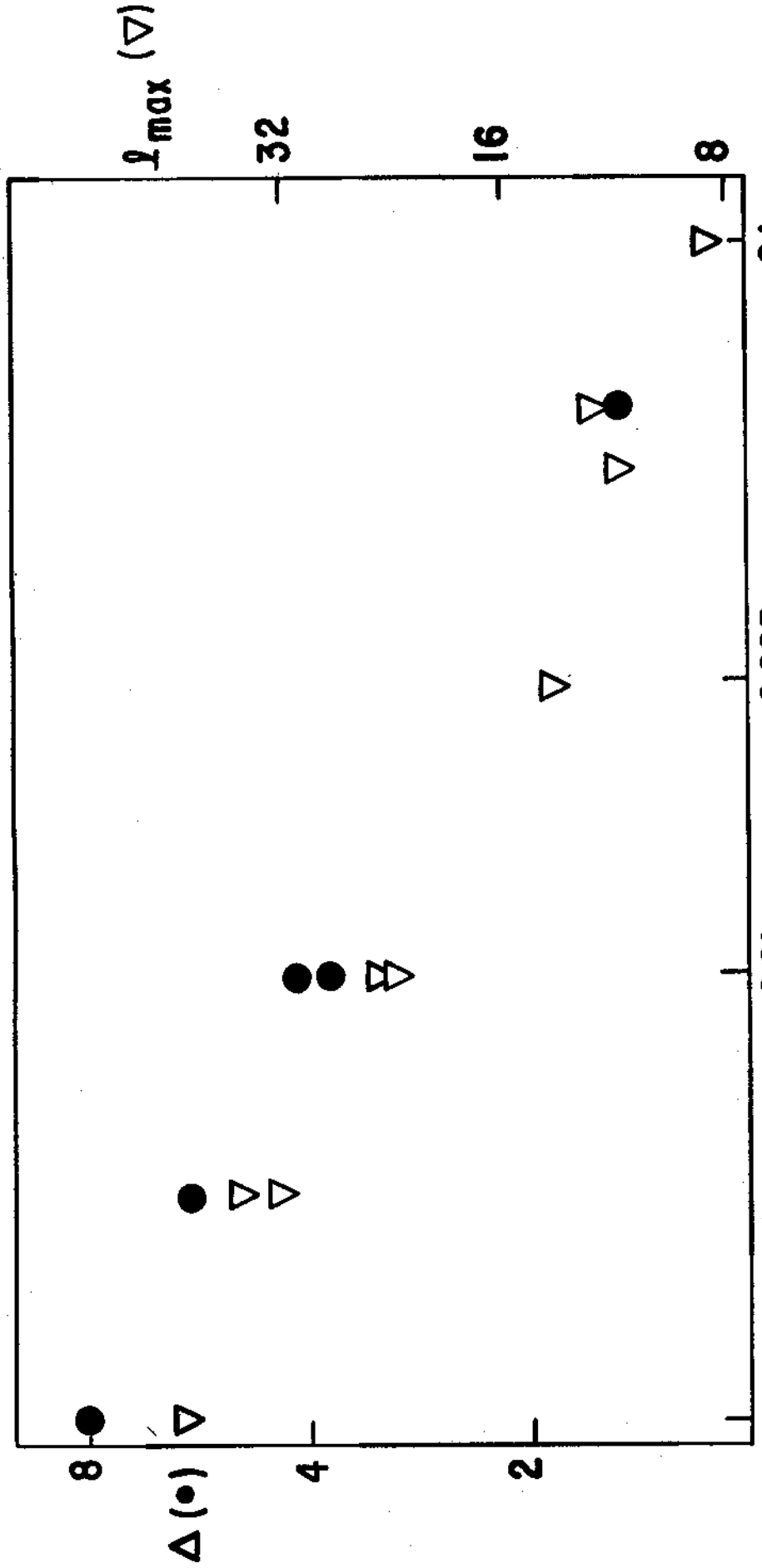


FIG. 7

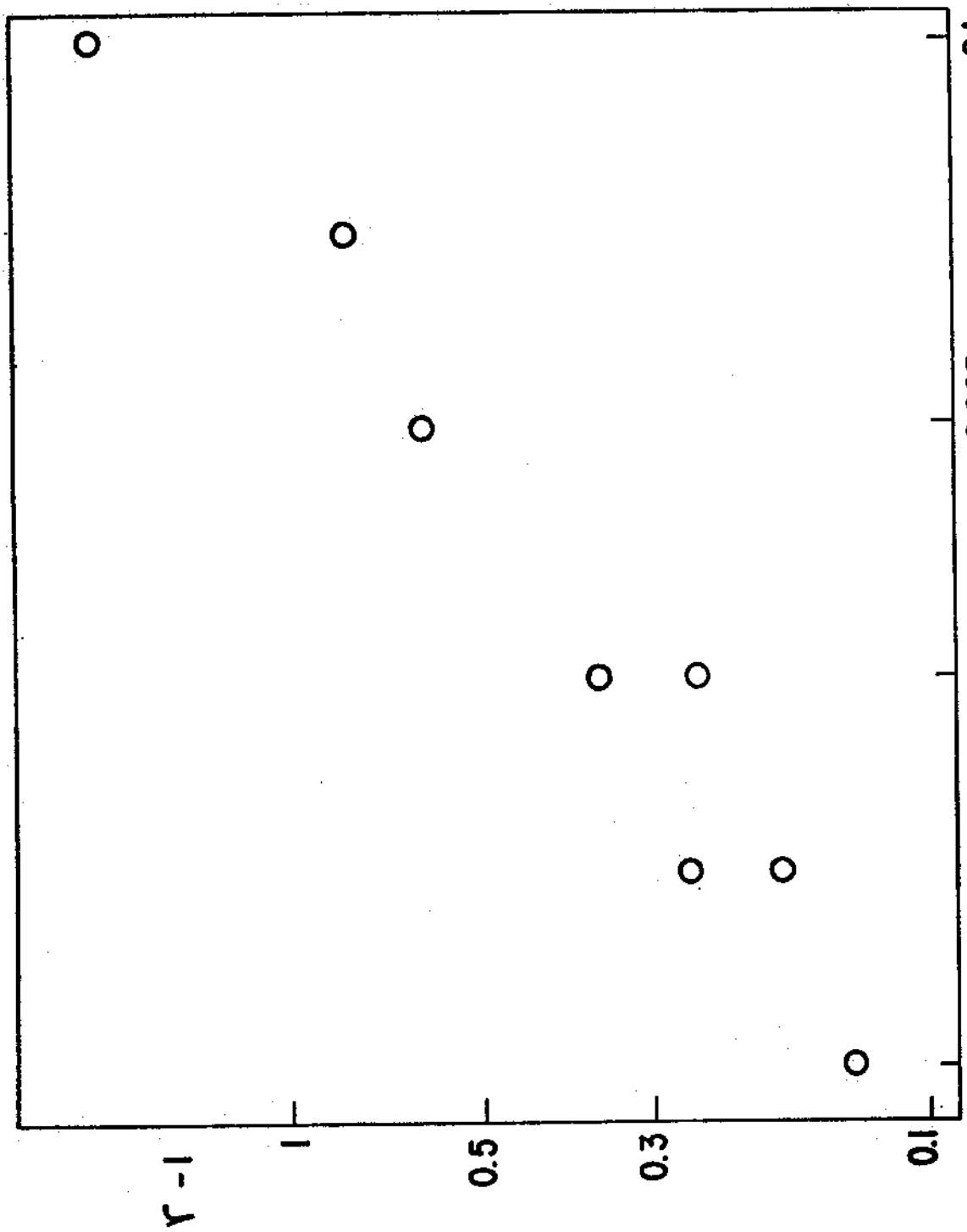


FIG.8

REFERENCES

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

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

[3] G. Nicolis and I. Prigogine, "Self-organization in Non-equilibrium Systems" (Wiley, New York, 1977).

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

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

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

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

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

[9] L. Demetrius, Proc. Nat. Acad. Sci. USA 81, 6068 (1984).

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

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

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

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

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

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

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

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

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

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

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

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

[22] C. Tsallis and R. Ferreira, Phys. Lett. 99A, 461 (1983); R. Ferreira and C. Tsallis, J. Theor. Biol. 117, 303 (1985); C. Tsallis and R. Ferreira, "A Critical Phenomenon Approach of Biogenesis", in "Statistical Physics", H.E. Stanley, ed. (North-Holland, Amsterdam, 1986).

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

[24] M. Kolb, Phys. Rev. Lett. 53, 1653 (1984); P. Meakin, T. Vicsek and F. Family, Phys. Rev. B 31, 564 (1985).

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