J Mol Evol (1982) 18:207–216

Journal of Molecular Evolution

A Theory for the Origin of a Self-replicating Chemical System

II: Computer Simulation of the Autogen

David H. White and Michael S. Raab

Chemistry Department, University of Santa Clara, Santa Clara, California 95053, USA

Summary. In order to better understand the feasibility and limitations of the autogen (White 1980), a computer simulation based on the fluctuating clay environment was used to test whether autocatalytic growth would occur under various conditions. The results suggest that overall accuracies of replication and translation in the range of 90% and 10%, and protoenzyme turnover numbers of 10–120 monomers/protoenzyme/day are adequate for exponential growth. Nucleation of the components of the autogen from random background oligomers would be extremely rapid if oligomers lengths 2–6 were adequately functional, whereas oligomer lengths much greater than 10 are prohibited. The autogen would most likely nucleate and grow to dominance either rapidly (10–100 cycles of roughly 1 day each) or not at all.

Key words: Prebiotic – Origin of Life – Autogen – Autocatalysis – Computer simulation

Introduction

The form of the first self-replicating chemical system on the primitive earth was proposed in the first paper of this series (White 1980) to be the autogen, consisting of two protogenes coding for two protoenzymes, one of which catalyzes replication, the other translation (See Fig. 1).

The autogen was found (White 1980) to be organized in a closed autocatalytic loop similar to a hypercycle (Eigen 1971; Eigen and Schuster 1977, 1978a, b). However, it differs from the hypercycle in that it involves shorter oligomers operating without specificity toward each other, depending instead on proximity of the components. The autogen is proposed to have nucleated from essentially random oligomers; in this regard it is similar to the proposal of Hoffmann (1975) for the stochastic nucleation of a system including adaptors for protein synthesis. Although the autogen theory is not dependent on a particular mechanism of protein synthesis, it was proposed that the most likely mechanism for the first self-replicating system would involve a direct singlet translation apparatus which would later evolve to a triplet adaptor system. (The transition will be discussed elsewhere.) This means that the nucleation of the components of the autogen would not have the critical reliance on accuracy in order to prevent error catastrophe which Hoffmann required. This point is discussed further below.

The ultimate verification or rejection of this theory must come from experimental evidence. However, in order to test the logic of the theory and to better understand its limitations and boundary conditions, a com-

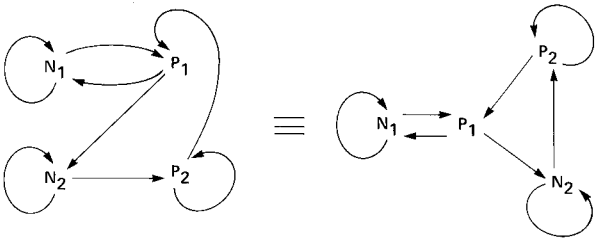

Fig. 1. Functional organization of the autogen. *Arrows* represent participation of one oligomer in the synthesis of another by templating or catalysis. Complementary oligonucleotides and their (non-functional) translation products are omitted. N_1 is the protogene for P_1, the replication protoenzyme, and N_2 the protogene for P_2, the translation protoenzyme. The organization is related to a hypercycle

Offprint requests to: D.H. White

0022–2844/82/0018/0207/$02.00

puter model was devised to simulate the autocatalytic growth or lack of growth of the autogen under certain conditions.

The computer simulation was based on the fluctuating clay environment (Lahav and Chang 1976; Lahav et al. 1978; Lahav and White 1980; White and Erickson 1980, 1981). The theory is applicable to a wide variety of environments besides clay surfaces, but these were modeled because of the availability of experimental data for comparison. Experimental measurements from the fluctuating clay environment were used to estimate parameters whenever possible, except that the turnover rates of the protoenzymes were taken as variables, and increased to determine whether exponential growth occurred. For convenience, turnover rates in the program were defined as molecules of oligomer per catalyst molecule per reaction cycle, whereas experimentally measured turnover numbers are expressed as monomer units incorporated into oligomer per catalyst molecule per day (White and Erickson 1980). This difference in units is reconciled in the discussion. Furthermore, most of the calculations were based on oligomers of decameric length. This is a conservative assumption, since shorter oligomers are easier to nucleate into a cooperative system (due to their greater abundance) and require lower accuracies and catalytic rates, as will be shown.

Assumption Used in the Computer Simulation

The basic procedure of the cycling experiment on clay was used to model an experiment which could conceivably produce a self-replicating system (White 1980). Although oligonucleotide formation has only been observed in this system in the presence of condensing agents (Odom et al. 1979), it was assumed to occur at 1/10 the rate of peptide formation. Two simplifications were made for convenience: oligonucleotides were assumed to produce direct copies without explicity calculating the amounts of their complements, and new products in a cycle were assumed not to participate in catalytic or template processes until the next cycle. These two assumptions cancelled each other to some degree and greatly simplified the mathematics, allowing each cycle to be treated in an episodic manner. This in turn allowed use of experimental measurements based on cycling to be used to estimate relevant parameters. It was assumed that monomers for both peptides and oligonucleotides were always in large excess. All association constants were assumed to be low enough that kinetics were linear with respect to each component (template and protoenzyme). In most calculations, all oligomers were assumed to be decamers (or decameric regions in longer chains) and to be produced at rates independent of chain length or identity of the oligomer. Protoenzymes recognized all oligonucleotide templates equally. Total equilibration outside of localized zones was assumed.

The calculation assumed two chambers in the experimental setup, one the reaction chamber containing clay which could be dried and heated, the other an "ocean" containing the monomers in excess, from which an aliquot could be added to the clay and partially withdrawn again to simulate incoming and outgoing tides. The sequence for each cycle was as follows:

1. The chamber was dried out and products were synthesized due to three types of reaction outlined by the theory (White 1980): background (essentially random) synthesis, uncatalyzed template reaction of oligonucleotide, and template reaction catalyzed by the appropriate protoenzyme. Specific products arose not only from catalyzed and uncatalyzed reactions involving the particular protogene coding for that product also by error from other oligonucleotides.
2. A certain fraction of products were degraded due to heating on the clay as well as by hydrolysis of those dissolved in the "ocean".
3. A fraction of the "ocean" was added to the clay reservoir, allowing a fraction of the oligomers to stay bound to the clay.
4. A fraction of the supernatant was returned to the ocean, bearing dissolved oligomer products. The remainder of the supernatant provided monomers and some dissolved oligomers to return to the clay surface in the next cycle.
5. Three different assumptions were made to describe the interactions of oligomers remaining localized near each other on the clay surface. These are discussed in the next section.

The format of the computer calculations is shown in the Appendix.

Results of the Computer Simulation

The most critical input to the simulation was the nature of localization of genes and their products (discussed by White 1980). When no localization was assumed, there was no exponential growth since all oligonucleotides were affected equally by any functional catalyst and the parent protogene had no selective value. A partial localization was next attempted. It was assumed that progeny were able to stick near the parent with 90% efficiency and react with it in the next cycle about 50% of the time. (These numbers could be varied, provided that the turnover rates were also varied in approximately an inverse fashion.) The result was not a closed hypercycle, but demonstrated many features of the model. Macroscopically observable exponential growth was not achieved until a turnover rate of 30,000 oligomer products/protoenzyme/cycle was assumed for both catalysts, although microscopic increases were seen with

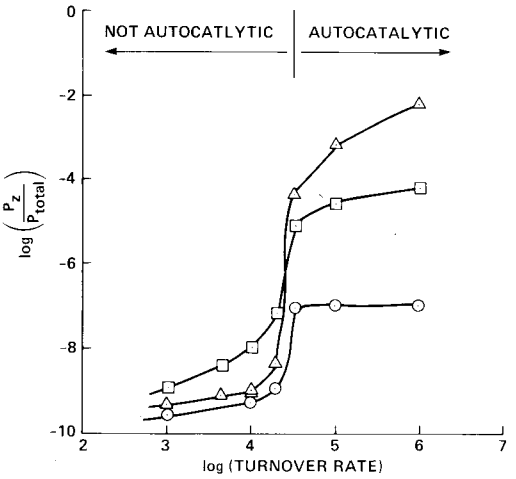

Fig. 2. Yield at 100 cycles vs. turnover rate for nonhypercyclic, partially localized system. Both protoenzymes were assigned the same value of turnover rate in a particular calculation. △ = P_1, replication protoenzyme; □ = P_2, translation protoenzyme; 0 = P_0, nonfunctional product from translation of N_0

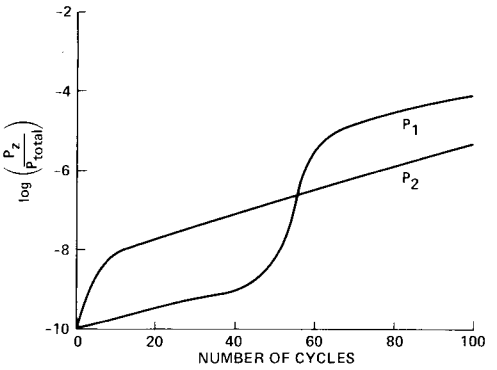

Fig. 3. Yield as a function of cycle number for a partially localized, non-hypercyclic system with protoenzyme turnover rates = 30,000 oligomers/protoenzyme/cycle

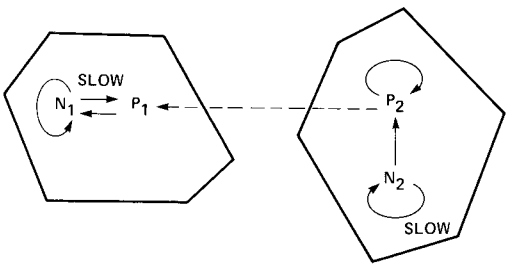

Fig. 4. Result of Partial Localization: Two separate localized zones, mildly linked by diffusion of P_2 through solution

turnover rates as low as 100. The resulting peptide yields after 100 cycles as the turnover numbers were varied are shown in a log–log plot (Fig. 2). At low turnover numbers, all peptide sequences were close to the assumed random background level of 10^{-10}. Genetically coded but nonfunctional peptides (P_0) showed a moderate increase at the transition region to reflect the fact that (essentially random) background peptides were no longer contributing significantly to the yield, and that all oligonucleotides, whether functional or not, were being translated by the high levels of translation protoenzyme. P_1 and P_2 increased dramatically as turnover rates exceeded 30,000/cycle. P_2 increased to 0.01% of all peptide and P_1 to about 1% when turnover rates were increased to 10^6/cycle. A time-dependent plot of P_1 and P_2 (Fig. 3) at the turnover rate of 30,000/cycle reveals what was happening. P_2 began to grow rapidly at first, then at a slower (but still exponential) rate thereafter. The reason for the slowdown is apparent in Fig. 4. P_2 was catalyzing its own formation in the vicinity of N_2 (producing the rapid initial rise) but was limited by the moderately slow uncatalyzed growth of N_2. Meanwhile, P_1 was growing slowly due to an N_1–P_1 autocatalytic cycle in other locales, but this cycle was limited by the very slow uncatalyzed translation of P_1. Once enough P_2 was built up to diffuse to the other locale sufficiently, P_1 became translated by P_2 catalytically and reached a new growth rate which was dependent on the supply of P_2. However, P_2 did not made a similar transition because N_2 gained no selective advantage over other random oligonucleotides.

Clearly, the above situation was not hypercyclic, since the cycle was not closed, but represented two independent, mildly autocatalytic cycles from Stage IV (Fig. 1 of White 1980) with a one-directional link growing in between them due to diffusion of P_2 as the oligomer concentrations increased. The extremely high turnover numbers required for autocatalysis in this situation are unlikely to be accessible either in the laboratory or prebiotically on the primitive Earth. For all practical purposes, autocatalytic growth of independent protogenes is not feasible.

However, the independence of the two autocatalytic groups is unlikely to last for a long time. As soon as one locale contains both protogenes as well as their protoenzyme products, which becomes more likely as all of the oligomers increase in concentration, the true autogen would develop. The likelihood of nucleation of the components of the autogen as a function of oligomer length is discussed more fully in the next section. The computer model was redesigned so that a single locale (such as the surface of a single clay particle) had one copy of each protogene and one copy of each protoenzyme (actually, only the translation protoenzyme was required, produced by uncatalyzed translation or by the catalytic cycles of the partially localized stage) as well as ten nonfunctional oligonucleotides. The same rate equations were used as before, except that the protoenzymes which were in the localized region encountered only the oligonucleotides of the same localized region, and made total numbers of products equal to their turnover numbers, acting on each protogene in proportion to its frequency in the region. As products increased, it was

assumed that the locale expanded to colonize nearby regions as the populations saturated the original region, so no restraints were placed on exponential growth of the localized products. During each cycle 10% of the localized oligomers were assumed to escape into the environment, where the same processes as previously assumed by the partially localized model could occur, but these were not significant until the product reached macroscopic amounts. The total pool of oligonucleotides was assumed to increase in proportion to the amounts of functional oligomers, due to natural selection of the components of the autogen from error copies (equations 29–30, App.). This assumption is discussed more fully in a later section.

Results of the fully localized model after 100 cycles as a function of catalytic turnover rate (assumed to be equal for both protoenzymes) are shown in Fig. 5. Turnover rates of 12 oligomers/protoenzyme/cycle were adequate to produce yields of functional product approaching 1% of the total oligomer at 100 cycles. Each separate calculation showed patterns similar to each other, with background levels (10^{-10}) of each peptide sequence occurring throughout the early cycles. However, the localized products were building up exponentially from very low absolute numbers until they matched the background levels (roughly 10^5 copies of each sequence). Then exponential growth of both protoenzymes was visible in the bulk environment. At turnover rates less than 12/cycle, the amounts of P_1 and P_2 were found to be equal as were N_1 and N_2, due to the assumption of equal interaction with each oligonucleotide by a given protoenzyme. As the products in the bulk environment began to make significant contributions, the two products were differentiated. This can be seen with turnover numbers greater than 12, where the localized products have "colonized" most of the environment and the $N_2 \rightarrow P_2$ autocatalytic cycle has begun to operate independently of the hypercycle.

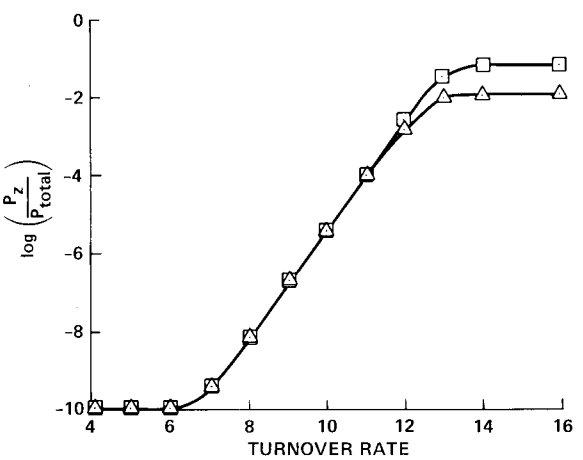

Fig. 5. Yield at 100 cycles vs. turnover rate for hypercyclic, fully-localized system. △ = P_1, replication protoenzyme; □ = P_2, translation protoenzyme

At turnover rates as low as 6/cycle, a steady exponential growth of the localized hypercycle was observed which was just breaking through the background levels at 100 cycles. Several hundred cycles would be required to reach macroscopic product levels. Even at a turnover rate of 5, very slight growth of localized hypercycle occurred, but it would not reach significant product levels except with an extremely large number of cycles. Finally, some work with differing turnover numbers for the two protoenzymes showed a much greater sensitivity to the turnover rate of the translation protoenzyme P_2, whereas the turnover rate of P_1 could be as low as 1/cycle and still give vigorous hypercyclic exponential growth.

The boundary conditions of turnover numbers and accuracies of translation and replication were studied to see how they interact with each other. Changes in the accuracy of uncatalyzed translation had no effect on the number of cycles in which the system attained exponential growth, and changes in the accuracy of uncatalyzed replication had a minor effect due to its slight contribution to product yields. The accuracies of catalyzed translation and replication were far more critical. Changes in either of the catalyzed accuracy rates could be matched by roughly inversely proportional changes in turnover numbers to reestablish the threshhold for exponential growth, but the replication accuracy must be high enough to prevent too rapid accumulation of error copies (discussed below).

The initiation of the localized region required only the translation protoenzyme and the two protogenes, since the replication protoenzyme is rapidly produced. Without the translation protoenzyme, turnover numbers had to be increased to 200 to observe the threshhold. Changes in oligomer length affected the ease of autocatalytic growth insofar as the accuracies were changed. If accuracies and background levels were altered to be appropriate for a trimer rather than decamer (assuming constant selectivity per digit), exponential growth was observed with a turnover rate of 2/cycle. In addition, the initiation would be expected to be much more favorable for trimers due to the greatly increased probability of production of the background oligomers (see next section), but this was not explicitly modeled in the computer simulation.

It is clear from the computer simulation that low catalytic turnover rates of protoenzymes are adequate for the hypercycle to grow exponentially. The calculated result of roughly 12 oligomers/protoenzyme/cycle, which produced vigorous growth to high levels within 100 cycles, was based on assumed decameric product. The corresponding turnover rate based on amino acids incorporated by the translation protoenzyme would be 120 amino acids incorporated/protoenzyme/cycle. The maximum turnover rate observed in the laboratory for histidyl-histidine in clay cycling experiments (White and Erickson 1980) was 18 amino acids incorporated/

catalyst/cycle, about one order of magnitude lower. Furthermore, the histidyl-histidine result was for simple peptide bond formation rather than translation from an oligonucleotide. It's activity as a translation protoenzyme is currently being studied. However, the turnover rate of histidyl-histidine has not been optimized. Increasing turnover rates were found as the absolute amount of histidyl-histidine was decreased, with no indication of a maximum. Histidyl-histidine is only a dipeptide, and only a few dipeptides have been tested. The best hexa- or decapeptide protoenzyme might well have a larger turnover rate. Finally, although decamers were assumed in the calculations, if lower oligomers were equally efficient as protoenzymes, then considerably lower turnover rates of the protoenzymes (about 10/cycle for a trimer) will suffice. If short peptides are as efficient catalysts of translation as histidyl-histidine is of non-translated oligoglycine synthesis, then it is likely that some peptides will be capable of achieving the threshhold turnover rate which the calculations show is required for the autogen to grow to macroscopic amounts.

Frequency of Nucleation of the Autogen

The critical event in the origin of the autogen is the nucleation of the necessary components in a single localized region (White 1980). The cycling reaction of glycine on clay may be used as a model environmnent to make order-of-magnitude estimates of the likelihood of nucleation. Kaolinite is assumed to have particles of 0.2μ approximate dimension or 2×10^7 Å^2 surface area per particle, and 20 m^2/g total surface area. From this an approximate particle weight of 10^{-14} g is calculated and it will thus be assumed that 10^{14} localized regions are available for distribution of oligomers in a typical experiment. The glycine oligomer yields decrease by approximately a factor of five for each increase of oligomer size by one (Lahav et al. 1978). Thus, approximately 1% yield of dimer in 11 cycles from 100 μmol of glycine on 1g kaolinite corresponds to 10^{17} molecules of dimer per cycle, extrapolated to 10^{14} molecules of hexamer and 10^{11} molecules of decamer. Nucleotide oligomerization is not known in this environment except with condensing agents, but will be assumed to be a factor of 10 less than peptide yields.

Three different situations will be calculated: a) the frequency of nucleation of two specific oligonucleotide sequences (the two protogenes) by chance in the same localized region; b) the frequency of nucleation of two protogenes plus a particular peptide sequence (the translation protoenzyme) by chance in the same localized region; and c) the frequency of nucleation of two protogenes in the same region plus the crude translation of one protogene to produce the translation protoenzyme. The parameters for crude translation are the same as those employed in the computer simulation, with roughly a twofold greater rate, or selectivity, of the preferred amino acid over other amino acids. Once the two protogenes and the translation protoenzyme are nucleated, the replication protoenzyme is quickly produced and the autogen grows exponentially.

The following parameters are defined as a function of oligomer length x:

R = Localized regions = 10^{14}

P_B = Total background peptide yield per cycle = $10^{(18-0.7x)}$

N_B = Total background oligonucleotide yield per cycle = $10^{(17-0.7x)}$

P_p = Probability of a specific peptide sequence of length $\underline{x}$ from random reaction of ten amino acids = 10^{-x}

N_p = Probability of a specific oligonucleotide sequence of length $\underline{x}$ from random reaction of four nucleotides = $4^{-x} \simeq 10^{-0.6x}$

P_{ut} = Probability of producing the master copy peptide from uncatalyzed (crude) translation of a particular oligonucleotide sequence (at a rate of 0.1 products per oligonucleotide in one cycle and an accuracy of $10^{-0.6x}$) = $10^{(-1-0.6x)}$

Frequency of nucleation is calculated for 2 protogenes as follows:

$$(N_B N_p)^2/R = 10^{(20-2.6x)} \tag{39}$$

For 2 protogenes plus 1 protoenzyme by chance:

$$(N_B N_p)^2(P_B P_p)/R^2 = 10^{(24-4.3x)} \tag{40}$$

For 2 protogenes plus 1 protoenzyme by crude translation:

$$(N_B N_p)^2 P_{ut}/R = 10^{(19-3.2x)} \tag{41}$$

The resulting calculations, shown in Table 1, reveal a surprisingly steep dependence of nucleation frequency on oligomer length. A combination of specific dimer sequences takes place with 10^{15} copies, hence an average of more than one on each clay particle in a single cycle, since there are more copies of a particular dimer than there are clay particles. On the other hand, a combination of decamers has a 10^{-19} probability of getting even a single region containing the components of the autogen (two protogenes and one protoenzyme) by chance combinations. At higher oligomer lengths, the crude translation process is much more likely to produce a functioning autogen than is the totally random combination process, since the frequency of a particular peptide sequence from random synthesis drops off very

Table 1. Localized zone nucleation frequency per cycle as a function of oligomer length[a]

Localization assumptions	*Oligomer length:* 2	4	6	8	10	Maximum oligomer length (frequency ≥ 1)
2 protogenes	10^{15}	10^{10}	10^{4}	10^{-1}	10^{-6}	7
2 protogenes plus 1 protoenzyme by chance	10^{15}	10^{6}	10^{-2}	10^{-10}	10^{-19}	5
2 protogenes plus 1 protoenzyme by crude translation	10^{13}	10^{6}	10^{0}	10^{-6}	10^{-13}	6

[a]Number of localized regions nucleated per cycle, calculated by equations in the text

rapidly with increasing peptide length. The maximum oligomer length that can spontaneously produce one or more copies of the autogen per cycle, and hence be highly probable of developing within a few cycles, is the hexamer. Longer oligomer lengths are allowed if reaction is continued for many cycles due to increasing total yields of oligomers as well as shuffling of a portion of the oligomers into new combinations with each cycle. Allowed oligomer length is approximately 2 longer at 100 cycles and a total of 4 longer at 10,000 cycles, assuming increases in yield proportional to the number of cycles at each oligomer length (which is in fact not feasible past a certain point.)

These calculations are so crude as to be merely suggestive. The restrictions noted above may be relaxed somewhat due to a variety of conditions: higher relative yields of the higher oligomers as the reaction progresses; higher probability of a particular sequence if the number of recognizably different monomers is less than ten amino acids or four nucleotides or if crude translation is more accurate; higher than background amounts of translation protoenzymes due to $N_2 \rightarrow P_2$ autocatalytic cycles if the turnover number is high enough; or the existence of many sequences which have a particular catalytic ability rather than a unique sequence for each protoenzyme. If one protogene is allowed to be shorter than the other, the results may also be adjusted accordingly. For instance, if the translation protoenzyme is catalytic as a tetrapeptide, and the corresponding translation protogene is also a tetramer, then the replication protogene may be a nonamer to produce similar probabilities as when all three are hexamers.

The general conclusion can be drawn that larger oligomer size very rapidly reduces the probability of nucleation to the vanishing point. In practical terms, if oligomer lengths much larger than ten are required for catalytic function, it is doubtful that the autogen could get started even in millions of years. The most likely oligomer lengths are in the range of roughly 2–8 for nucleation to occur spontaneously. Once autocatalytic growth is started, oligomer length is less critical to the efficiency of the autogen, and evolution might occur to somewhat longer chain lengths if these were more efficient catalysts.

Error Catastrophe

A further problem of the nucleation process is the possibility of an error catastrophe. The autogen is not susceptible to the same kind of error catastrophe suggested by Orgel (1963). He proposed that errors in a complex protein synthesis machinery by some component (such as an amino-acyl tRNA synthetase) might lead to less accurate synthesis of all of the components, eventually leading to an exponential increase of errors and total loss of function. In the autogen there is not the same kind of complex machinery as in the modern cell, in which a single component could affect the product resulting from the cooperative interaction of many components, nor is very much precision required.

However, another kind of error catastrophe is possible, due to pollution of the localized environment with error copies and other nonfunctional oligonucleotides. Since the protoenzymes were postulated not to recognize their parent protogenes, they would produce less net functional product if time were wasted translating and replicating nonfunctional oligonucleotides. As more error copies accumulated, the ratio of functional to nonfunctional oligonucleotides would decrease until the rate of synthesis could no longer compensate for degradation rates. Since it is the information content of the oligonucleotide pool which is important, a buildup of nonfunctional peptides would probably be of little consequence. (Some peptides might conceivably act as inhibitors, but short oligomers are assumed here to bind weakly and reversibly.)

The most critical parameter is the net replication accuracy. If this number is low, such as 0.1, then the ratio of functional protogenes to total oligonucleotides in the localized environment will be rapidly degraded. On the other hand, a higher replication accuracy such as 0.9 would lead to a more gradual reduction of the ratio of functional to nonfunctional oligonucleotides. Calculations based on a localized environment that is assumed to contain a single homogeneous pool of oligonucleotides (equations 27–28, App.) showed that an overall replication accuracy (C_1) of 0.9 produced growth above the background level only with turnover rates higher than 100 oligomers/catalyst/cycle. Fruthermore, growth leveled of to a plateau and eventually decreased as the ratio of polluting to functional oligomers continued to grow. When replication accuracy was 0.99, a turnover rate of 10 was adequate to produce exponential growth above background. The results were also sensitive to the

uncatalyzed replication accuracy, since it was assumed to be less accurate than catalyzed replication and made significant contribution to oligonucleotide production when catalytic turnover numbers were low. This suggests that too high a rate of uncatalyzed replication would be detrimental to the growth of the autogen unless it were as accurate as catalyzed replication.

These results make it clear that a homogeneous pool would be at the mercy of the polluting error copies unless their rate of formation were severely limited. However, a homogeneous model for the localized environment is unrealistically simple. Any tendency for the protoenzymes to remain near the parent protogene would increase the proportion of functional products compared to products from other oligonucleotides. Alternatively, weak recognition of the parent protogene by the progeny would produce a similar preference. More importantly, statistical fluctuations within the localized region would produce some areas with higher densities of the components of the autogen, and as products colonized other nearby localized regions where there were few background oligomers, some would have higher ratios of autogen components to polluting oligomers than the original, more highly populated region due to statistical fluctuations among low populations. Thus new regions would be formed, either as subregions of the original or as new colonies, which would have a greater proportion of functional molecules, and which would reproduce more efficiently until they too began to build up an accumulation of error copies. Other regions which had few or no functional oligomers by statistical fluctuation would be less efficient and would reproduce more slowly or not at all. Thus, natural selction due to autocatalysis would allow the continual emergence of the self-reproducing autogen from among the error copies if the replication accuracy were not too low.

Another feature which may serve to attenuate the problem of polluting oligonucleotides is that the turnover rates of the protoenzymes may be roughly proportional to oligonucleotide concentration rather than constant as was assumed in the calculations. Since the protoenzymes are assumed not to bind strongly to the template, they would not easily be saturated as is the case with modern enzymes, and they might continue to reproduce the components of the autogen at approximately the same rate regardless of the presence of polluting oligomers. This question may eventually be answered by experiments with model protoenzymes.

In order to model a situation in which natural selection continually "purified" the components of the autogen from the background oligomers, or in which the presence of pollutants was irrelevant due to inability to saturate the protoenzyme rates, equations 29–30 were used to calculate the total amounts of oligomers present. It was arbitrarily assumed that ten polluting oligomers were present to start, and that a steady state ratio was obtained, matching the ratio of total catalyzed products to total functional oligomers produced by the appropriate protoenzyme in a single cycle. This made it somewhat harder for autocatalytic growth to start but allowed it to proceed exponentially once the steady state was reached. This model was used for all of the calculations reported in previous sections; thus exponential growth above background levels was detected with turnover rates of 6 or more. The two models were virtually identical when uncatalyzed replication was eliminated and catalyzed replication had an accuracy of 0.99. Under less favorable circumstances where the two simple models used here show significant divergence, a more realistic model would probably give results somewhere between them. As a result, turnover rates calculated earlier may be somewhat optimistic, but it is suspected that the steady-state model used for most of the calculations is more realistic than a homogeneous pool in which no selection can occur. An overall catalyzed replication accuracy of 0.9 will probably permit exponential growth by these considerations, but may be considered a rough approximation of the practical lower limit.

Conclusion

The formation and autocatalytic growth of the autogen are possible if certain boundary conditions are satisfied. These conditions interact with each other and are dependent on the experimental environment as well, but a typical set of conditions for the clay cycling environment modeled here are oligomer lengths no greater than decameric, overall catalyzed replication and translation accuracies of 90% and 10%, and protoenzyme turnover numbers of 120 monomers/protoenzyme/cycle. Turnover numbers could be considerably less for oligomers shorter than the decamer. Oligomer length is a critical parameter for nucleation of the autogen, with oligomers of length 2–6 capable of nucleating rapidly, whereas lengths much greater than 10 monomer units prohibit the kind of spontaneous assemblage from essentially random components that is postulated for the autogen.

The histidyl-histidine experimental results are encouraging from the standpoint of the catalytic activity available from a short peptide, although it has not yet been demonstrated that it or any peptide is capable of catalysis in a translation system. Rates comparable to those already observed for histidyl-histidine would be adequate for protoenzymes if the oligomers are in the range 2–3 monomers in length. Roughly an order of magnitude greater rate would be required than has been observed so far if the components of the autogen were decamers.

The first self-replicating system would have been even simpler than the autogen if one protoenzyme were able to catalyze both replication and translation with the required accuracy (White 1980), leading to a two-component second order hypercycle. This is conceivable but

may not be realistic due to the unlikelihood of finding the required enhancement of transition state selectivity of both processes by the same protoenzyme. Furthermore, it would allow only a marginal improvement in ease of nucleation (allowing oligomers of length 7–8 instead of 6) or might allow small reductions in efficiency of protoenzyme catalysis. Thus the one-protoenzyme model is not necessary and is most likely not sufficient for the spontaneous formation of of a self-replicating system.

This study was limited to demonstrating the self-consistency of the proposed model, and to determining approximate boundary conditions for autocatalytic growth of the autogen. It was not intended to be a complete mathematical analysis, due to the many parameters which could be varied. The computer model was applied to the clay cycling environment because some of the required parameters were known or could be estimated in this system. However, this is not the only nor even necessarily the optimum environment in which the autogen is potenially capable of developing. The theory can be applied to any environment which provides appropriate background oligomer synthesis, localization mechanisms, protoenzyme turnover rates, and accuracies of catalyzed translation and replication.

An important result of this study is that the autogen is much more efficient as a functional, whole system than are its isolated components. The system produced autocatalytic growth with turnover numbers four orders of magnitude lower than those required for isolated protogenes. This provides a basis for a strong and continual selection of the functional autogen from error copies and from skewed distributions. Both continual selection and the hypercyclic organization of the autogen would help to prevent dominance by any single component of the autogen. Finally, the fact that the autogen can be nucleated without dependence on prior "evolutionary" development of either polynucleotides or peptides alone places this theory firmly in the school of thought that the first self-organized, genetic systems consisted of both classes of polymers (Eigen 1971; Hoffmann 1975).

The rapid nucleation and exponential growth of the autogen resembles an all-or-nothing phenomenon. Either oligomers are functional at short enough lengths and turnover numbers are large enough for growth to occur to macroscopic amounts in 10–100 cycles, or slightly longer oligomers or lower turnover rates give essentially no observable autocatalytic growth at all. Thus, if the autogen is capable of nucleation and growth, it will most likely grow to dominance in a relatively short period of time. The implications of this conclusion for the origin of life on the Earth, as well as for the design of experiments to test the autogen theory, are discussed in the previous paper (White 1980).

Acknowledgements. Discussions with Noam Lahav of the Hebrew University of Jerusalem were instrumental in designing the experimental model on which these calculations were based. Computer time was donated by Robert MacElroy of NASA-Ames Research Center and by the University of Santa Clara. Financial support of Nasa–Ames University Consortium Interchange Nos. NCA2-OR685-702, and -806 is gratefully acknowledged.

References

Eigen M, (1971) Naturwissenschaften 58:465–523
Eigen M, Schuster P (1977) Naturwissenschaften 64:541–565
Eigen M, Schuster P (1978a) Naturwissenschaften 65:7–41
Eigen M, Schuster P (1978b) Naturwissenschaften 65:341–369
Hoffmann GW (1975) Annu Rev Phys Chem 26:123–144
Lahav N, Chang S (1976) J Mol Evol 8:357–380
Lahav N, White D, Chang S (1978) Science 201:67–69
Lahav N, White DH (1980) J Mol Evol 16:11–21
Odom DG, Lahav N, Chang S (1979) J Mol Evol 12:259–264
Orgel LE (1963) Proc Natl Acad Sci USA 49:517–521
Seidl F, Folsome CE, Lawless JG (1978) Abstract, Pacific Conference on Chemistry and Spectroscopy, San Francisco, pp XX
White DH (1980) J Mol Evol 16:121–147
White DH, Erickson JC (1980) J Mol Evol 16:279–290
White DH, Erickson JC (1981) J Mol Evol 17:19–26

Received July 8, 1980/Revised April 27, 1981

Appendix – Definitions and Equations Used in Calculations

Definition of Oligomer Populations:

P_z^n = Number of copies of peptide sequence z present in reaction cycle n

N_z^n = Number of copies of oligonucleotide sequence z present in reaction cycle n
($z = 1$: Protoenzyme and Protogene for Replication)
($z = 2$: Protoenzyme and Protogene for Translation)
($z = 0$: Functionless Peptide and Oligonucleotide of a particular sequence, representing one member of the class of functionless sequences and used for comparison purposes)

P_T^n = Total number of peptides of all sequences present in reaction cycle n

N_T^n = Total number of oligonucleotides of all sequences present in reaction cycle n

O_z^n = Number of copies of peptide sequence z in ocean reservoir

Q_z^n = Number of copies of oligonucleotide sequence z in ocean reservoir

S_z^n = Number of copies of peptide sequence z in supernatant in clay reaction chamber during wetting phase

T_z^n = Number of copies of oligonucleotide sequence z in supernatant in clay reaction chamber during wetting phase

P_{zL}^n = Number of copies of peptide sequence z in localized environment during cycle n

N_{zL}^n = Number of copies of oligonucleotide sequence z in localized environment during cycle n

P_{TL}^n = Total number of peptides present in localized environment during cycle n

N_{TL}^n = Total number of oligonucleotides present in localized environment during cycle n

Definition of Constant Parameters:

B_1 = Turnover rate for Replication Protoenzyme (number of product oligonucleotides per protoenzyme molecule per reaction cycle), value varied in each trial

B_2 = Turnover rate for Translation Protoenzyme (number of product peptides per protoenzyme molecule per reaction cycle), value varied in each trial

C_1 = Accuracy of catalyzed replication (fraction of oligonucleotide products with sequence identical to original oligonucleotide, ignoring complementary oligonucleotide step) = 0.9 normally, varied 0.99 to 0.001

C_2 = Accuracy of catalyzed translation (fraction of peptide products with "correct" sequence of amino acids corresponding to coding relationship with corresponding protogene) = 0.1 normally, varied 0.9 to 0.001

C_3 = Accuracy of uncatalyzed template replication = 0.5 normally, varied to 0

C_4 = Accuracy of uncatalyzed template translation = 0.000001 normally, varied to 0

D = Number of uncatalyzed replication products per oligonucleotide per cycle = 0.1 normally, varied from 0 to 1

E = Number of uncatalyzed translation products per oligonucleotide per cycle = 0.1 normally, varied from 0 to 1. (Measured enhancement factors of glycine oligomers due to polynucleotides on clay are in this range; White and Erickson 1981

P_B = Number of peptide molecules synthesized by background reaction (neither catalyzed nor templated) per cycle = 10^{14}

N_B = Number of oligonucleotide molecules synthesized by background reactions per cycle = 10^{13}

P_P = Probability of producing a particular peptide sequence by chance during background reaction = 10^{-10} normally for decamers, varied from 10^{-3} to 10^{-20} depending on assumed oligomer length

N_P = Probability that protoenzyme *z*, if bound to clay sequence by chance during background reaction = 10^{-6} normally for decamers, varied from 10^{-2} to 10^{-12} depending on assumed oligomer length

F_z = Probability that peptide *z* will remain bound to clay when supernatant is added = 0.9 for all *z* (measured 0.97 for histidyl-histidine; White and Erickson 1980)

H_z = Probability that oligonucleotide *z* will remain bound to clay when supernatant is added = 0.9 for all *z* (most polynucleotides bind 0.95; Seidl et al. 1978)

G_z = Probability that protoenzyme *z*, if bound to clay nea near parent protogene, will find the parent protogene and react with it = 0.5 normally for all *z*, varied from 0.01 to 1

K_z = Probability that peptide *z* will degrade in aqueous solution during one cycle = 0 for all *z*

L_z = Probability that oligonucleotide *z* will degrade in aqueous solution during one cycle = 0.01 for all *z*

M_z = Probability that peptide *z* will degrade on the clay surface during one cycle = 0.1 for all *z*

R_z = Probability that oligonucleotide *z* will degrade on the clay surface during one cycle = 0.1 for all *z*

J_1 = Fraction of "Ocean" reservoir added to clay reaction chamber during wetting phase of each cycle = 0.1

J_2 = Fraction of supernatant in clay reaction chamber returned to "ocean" reservoir before drying phase of each cycle = 0.5

J_3 = Fraction of localized peptides released to environment each cycle = 0.1

J_4 = Fraction of localized oligonucleotides released to environment each cycle = 0.1

Equations:

New Product Synthesis (for each specific product sequence *z* during reaction cycle *n*):

$$P_z^n = \frac{N_z^{n-1}}{N_T^{n-1}} P_2^{n-1} B_2 C_2 + N_z^{n-1} C_4 E + P_p P_B + P_p P_2^{n-1} B_2 + P_p E N_T^{n-1} + P_z^{n-1} \quad (1)$$

$$N_z^n = \frac{N_z^{n-1}}{N_T^{n-1}} P_1^{n-1} B_1 C_1 + N_z^{n-1} D C_3 + N_p P_1^{n-1} B_1 + N_p D N_T^{n-1} + N_p N_B + N_z^{n-1} \quad (2)$$

$$P_T^n = P_2^{n-1} B_2 + E N_T^{n-1} + P_B + P_T^{n-1} \quad (3)$$

$$N_T^n = P_1^{n-1} B_1 + D N_T^{n-1} + N_B + N_T^{n-1} \quad (4)$$

Product Degradation during Cycle (assumed to occur after synthesis):

$$O_z^n = (1 - K_z) O_z^n \quad (5)$$

$$Q_z^n = (1 - L_z) Q_z^n \quad (6)$$

$$P_z^n = (1 - M_z) P_z^n \quad (7)$$

$$N_z^n = (1 - R_z) N_z^n \quad (8)$$

Addition of Water from Ocean:

$$S_z^n = J_1 O_z^n \quad (9)$$

$$T_z^n = J_1 Q_z^n \quad (10)$$

$$O_z^n = (1 - J_1) O_z^n \quad (11)$$

$$Q_z^n = (1 - J_1) Q_z^n \quad (12)$$

Transfer of Products from Clay to Supernatant:

$$S_z^n = (1 - F_z) P_z^n + S_z^n \quad (13)$$

$$T_z^n = (1 - H_z) N_z^n + T_z^n \quad (14)$$

$$P_z^n = F_z P_z^n \quad (16)$$

$$N_z^n = H_z N_z^n \quad (16)$$

Return of Part of Supernatant to Ocean, Binding of Remaining Solutes in Supernatant to Clay during Drying:

$$P_z^n = (1 - J_2)\, S_z^n + P_z^n \tag{17}$$

$$N_z^n = (1 - J_2)\, T_z^n + N_z^n \tag{18}$$

$$O_z^n = J_2\, S_z^n + O_z^n \tag{19}$$

$$Q_z^n = J_2\, T_z^n + Q_z^n \tag{20}$$

Partial Localization Assumption-Probability of Binding Products to Clay near Parent Protogene and Formation of Additional Products:

$$P_2^n = G_2 B_2 C_2 (P_2^n - P_2^{n-1})\, H_2 F_2 + P_2^n \tag{21}$$

$$N_1^n = G_1 B_1 C_1 (P_1^n - P_1^{n-1})\, H_1 F_1 + N_1^n \tag{22}$$

$$P_T^n = G_2 B_2 (P_2^n - P_2^{n-1})\, H_2 F_2 + P_T^n \tag{23}$$

$$N_T^n = G_1 B_1 (P_1^n - P_1^{n-1})\, H_1 F_1 + N_T^n \tag{24}$$

Full Localized Assumption – Protogenes and Protoenzymes Localized on a Single Clay Particle:

$$P_{zL}^n = N_{zL}^{n-1} C_4 E + \frac{N_{zL}^{n-1}}{N_{TL}^{n-1}} P_{2L}^{n-1} B_2 C_2 + P_{zL}^{n-1} \tag{25}$$

$$N_{zL}^n = N_{zL}^{n-1} C_3 D + \frac{N_{zL}^{n-1}}{N_{TL}^{n-1}} P_{1L}^{n-1} B_1 C_1 + N_{zL}^{n-1} \tag{26}$$

$$P_{TL}^n = P_{TL}^{n-1} + N_{TL}^{n-1} E + P_{2L}^{n-1} B_2 \tag{27}$$

$$N_{TL}^n = N_{TL}^{n-1} + N_{TL}^{n-1} D + P_{1L}^{n-1} B_1 \tag{28}$$

Altered Full Localization Assumption – Total Localized Peptides and Nucleotides Proportional to Components of the Autogen plus Ten Nonfunctional Oligomers (see text):

$$P_{TL}^n = \frac{1}{C_2} (P_{1L}^n + P_{2L}^n) + 10 \tag{29}$$

$$N_{TL}^n = \frac{1}{C_1} (N_{1L}^n + N_{2L}^n) + 10 \tag{30}$$

Release of Localized Products to Environment (Full Localization Assumption):

$$P_z^n = P_z^n + J_3 P_{zL}^n \tag{31}$$

$$N_z^n = N_z^n + J_4 N_{zL}^n \tag{32}$$

$$P_{zL}^n = (1 - J_3) P_{zL}^n \tag{33}$$

$$N_{zL}^n = (1 - J_4) N_{zL}^n \tag{34}$$

$$P_T^n = P_T^n + J_3 P_{TL}^n \tag{35}$$

$$N_T^n = N_T^n + J_4 N_{TL}^n \tag{36}$$

$$P_{TL}^n = (1 - J_3) P_{TL}^n \tag{37}$$

$$N_{TL}^n = (1 - J_4) N_{TL}^n \tag{38}$$