Templates, autocatalysis and molecular replication

Braja Gopal Bag¹ and Günter von Kiedrowski²

¹Department of Organic Chemistry, Indian Institute of Science, Bangalore, India
²Institute of Organic Chemistry and Biochemistry, Albert Ludwig University, Freiburg, Germany

Abstract: The studies of self-replication in chemical, viz. non-enzymatic model systems have gained significance in recent years for an improved understanding of the origin of life. This article briefly reviews the progress achieved, mainly during the past decade, from organic and bioorganic chemists point of view (1).

INTRODUCTION

Man has always been curious to know and find out the mystery behind the living systems. The theory of biological evolution and the principle of biological descendance within indicates that the origin of all forms of living cell was itself a "cell". However, spontaneous constitution of all the components of a cell (e.g., nucleic acids, proteins etc.) with defined functions seems unlikely based on random driven events. Hence, the "cell" must have been preceded by less complex precursors which, during the process of evolution, had led to the formation of it. The finding that RNA can act both as a carrier of genetic information and as a catalyst has made it as the most likely candidate for the first prebiotic self-replicating molecule (2). The demonstration of self-replication in the case of simple model systems has become a challenge for organic and bioorganic chemists for tracing back the original path of evolution.

A self-replicating molecule, by definition, is capable of acting autocatalytically for its own synthesis. Moreover, such autocatalytic molecule acts as a template to bind the precursors by noncovalent forces and organize them in such a way that the reactive groups come in close proximity. A schematic, minimal representation of self-replication is shown in Fig. 1. The ternary complex C is formed reversibly from the template T and the precursors A and B. The close proximity of the reactive ends facilitates the formation of a covalent bond between the two precursors. Reversible dissociation of the duplex D gives two template molecules, each of which can begin a new replication cycle. Self-replication has been achieved in the laboratory using both nucleotide and non-nucleotide model systems as discussed in the following two sections.
NUCLEIC ACID REPLIATORS

Early studies by Orgel and coworkers were concerned with the template directed synthesis of oligonucleotides from activated mononucleotides. 5'-phosphorimidazolides (1) were reacted in the presence of oligonucleotide templates. These studies revealed the influence of various parameters (e.g., the template, nature of leaving group, presence of metal ions, etc.) on the regioselectivity (2'-5' or 3'-5') of the condensation. They also demonstrated that pyrimidine-rich oligo- and polynucleotides could act as templates for the polycondensation whereas the purine rich could not (3). The latter findings appeared to be the major obstacles for achieving efficient template synthesis and non-enzymatic self-replication when starting from mononucleotide precursors. Inspite of the difficulties encountered with such systems, the Orgel group succeeded to demonstrate non-enzymatic template directed synthesis of fully complementary products: The pentamer pGGCGG, for example, was obtained in 18% yield from a mixture of nucleotides (Base = C, G) in the presence of CCGCC as a template (4). Later, templates as long as 14-mers were successfully transcribed (5).

Although the work of Orgel and coworkers showed that transcription with information transfer does occur in the absence of enzymes, a complete replication cycle could not be achieved when using mononucleotides as building material. It thus seemed worthwhile to study oligonucleotides as building material and to employ chemical ligations instead of template-directed polycondensations. Chemical ligations, viz. template-directed condensations of activated oligonucleotides, were known since 1966 (6) and had been studied extensively by the group of Shabarova in Moskow (7). The first example of a nonenzymatic self-replicating system was based on an autocatalytic chemical ligation reaction (8). A 5' terminally protected trideoxynucleotide 3' phosphate (2) d(Me-CCG-p) and a complementary 3' protected trideoxy- nucleotide (3) d(CGG-p') were reacted in the presence of a water soluble carbodiimide (EDC) to yield the self-complementary hexanucleotide (4) d(Me-CCGCGG-p') next to the 3'-3' linked pyrophosphate of 2. The sequences chosen were such that the product (4) could act as a template for its own production. The reaction proceeds via the termolecular complex C formed from 2, 3 and 4 (Fig. 2) in which the activated 3' phosphate of 2 is attacked by the adjacent 5'-OH group of 3 forming a 3'-5' internucleotide bond between the trimers. The resulting template-duplex then can dissociate to yield two free template molecules each of which can initiate a new replication cycle. Kinetic studies revealed that template-molecules are formed on an autocatalytic, viz. template-dependent, and a parallel non-
autocatalytic pathway. The rate of the autocatalytic synthesis was found to be proportional to the square root of the concentration of the template (square root law of autocatalysis). According to theory, a square-root law is expected if most of the template molecules remain in the unbroken double helical form (9). In other words, a square-root law reflects the influence of both, autocatalysis and product inhibition.

A similar type of autocatalytic behavior was reported by Zieliński and Orgel for the synthesis of tetranucleotide triphosphoramidate \( \mathcal{T} \) \((\text{G}N_{55}\text{C}N_{55}\text{G}N_{55}\text{C}N_{55})\) from the dinucleotide analogues \( \mathcal{S} \) \((\text{G}N_{55}\text{C}N_{55})\) and \( \mathcal{G} \) \((\text{G}N_{55}\text{C}N_{55})\) (10). This was the first demonstration of self-replication of nucleic acid like oligomers with a modified backbone (Fig. 3). Again the condensation was achieved in the presence of water-soluble carbodiimide EDC and the autocatalytic nature of template synthesis was obvious from the square root dependence of the initial reaction rate on the concentration of the template.

![Fig. 3 Self-replicating system by Zielinski and Orgel (10).](image)

Every true autocatalytic system should reflect a sigmoidal concentration-time profile (9). This was not evident in catalytic synthesis. The autocatalytic nature of these systems was established indirectly by an observation of the increase in the rate of reaction when seeding the reactions with the template. First direct evidence for a sigmoidal increase of template concentration was found in a modified hexadeoxynucleotide system in which the \(5'\)-hydroxylgroup of trimer \(3\) was replaced by a \(5'\)-aminogroup (11). Interestingly, the initial increase of template concentration did not resemble an exponential growth function but a parabola. Parabolic growth is the expected type of growth and a direct consequence of the square root law when the non-autocatalytic synthesis is negligible (Note a). From theoretical considerations it was concluded that the autocatalytic growth order - which is \(1/2\) in the case of parabolic growth - is solely determined by the thermodynamics of the coupled equilibria, and not by the energy of the transition state (9).

Independent evidence that the autocatalytic nature of hexamer synthesis is due to template-effects resulted from a series of experiments in which the sequence of trimer \(2\) was varied (12). Trimer \(3\) was introduced in its \(5'\)-phosphorylated form (p3), so that its carbodiimide-dependent condensation with \(2\) and the variants of \(2\) could yield hexamers bearing central \(3'-5'\)-pyrophosphate modifications. Also, in each experiment symmetrical self-condensation products bearing \(5'\)-\(5'\)- and \(3'\)-\(3'\)-pyrophosphate linkages could form. It was found that product distribution and the relative rates of product formation do indeed reflect the influence of template autocatalysis: The \(3'-5'\)-pyrophosphate from \(2\) and p3 phosphate was formed significantly faster as compared to the \(3'-5'\)-pyrophosphates from the sequence variants. Moreover, addition of template \(4\) being complementary to \(2\) and p3 stimulated the synthesis of the proper \(3'-5'\)-pyrophosphate while having a negligible influence on the variant sequences. These and other related studies in the \(3'-5'\)-phosphoramidate series showed that template autocatalysis can only occur if the sequences of both trimers match the sequence of the hexamer according to the Watson-Crick base-pairing rules (13). They also showed that the condensation reactions are predominantly controlled by the stacking of nucleic acid bases flanking the newly formed internucleotide link. Hexamers with a central G-C subsequence, for example, are formed one order of magnitude faster than hexamers with a C-G subsequence (13).

As a general result we found the following reactivity order: \(\text{G-G} > \text{G-C} > \text{C-G} > \text{C-C}\) (13), an order which was also established from nonautocatalytic chemical ligations (7). Further studies revealed a remarkable temperature-dependence of hexadeoxynucleotide self-replication (13). Each parabolic replicator has a rate optimum at a certain temperature, which was found to be close to the measurable melting temperature of the respective hexamer.

Note a: The equation \(\frac{dc}{dt} = a.c^p + b\) is a general form of the rate equation for template synthesis, where \(a.c^p\) is the autocatalytic term (\(c\) is the concentration of the template) and \(b\) is the non-autocatalytic term. For \(0 < p < 1\) the autocatalytic term describes subexponential growth, for \(p = 1\) parabolic growth and for \(p = 1\) exponential growth. Sufice it to say, parabolic growth is also a subexponential growth.
duplex (13). This is again in accordance with the theory of minimal replicators (9). Generally, the autocatalytic rate is given by $k [C]$, where $k$ is the rate constant for irreversible internucleotide bond formation and $[C]$ is the equilibrium concentration of the termolecular complex. As the rate constant $k$ increases with the temperature according to Arrhenius' law whereas $[C]$ decreases due to the melting of the termolecular complex, the rate as a function of temperature is expected to pass a maximum at the temperature $T_m$. $T_m$ depends on the concentrations of the template and its precursors as well as thermodynamic stabilities of the termolecular complex and the template duplex (9).

Recent studies in our lab were dedicated to the question, whether and how autocatalytic template effects could show up in more complex self replicating systems (14). The complexity of such systems depend on the number of products being formed in the course of the reaction. As an example, the trimer 3'-phosphate MTM-CCGp (MTM = 5'–methylthiomethyl), the 5'-aminodimer 3'-phosphate H$_2$N-CGp, and the 5'-amino-monomer H$_2$N-G were allowed to react in the presence of the carbodiimide EDC. Five products all bearing 3'-5'-phosphoramidate linkages (pn) were observed: the trimer H$_2$N-CGpnG, the tetramer H$_2$N-CGpnCGp, the tetramer MTM-CCGpnG, the pentamer MTM-CCGpnCGp, and the hexamer MTM-CCGpnCGpnG. The latter product is synthesized from the pentamer and the monomer, and, parallelly, from the precursor trimer and the trimeric product. The whole reaction system was then separated into less complex subsystems. For example, in order to analyze the formation of the pentamer separately, we employed a 5'-aminodimer which could not react at its 3'-phosphate. In a series of experiments, each subsystem was studied with respect to the effect of each reaction product. Standard oligodeoxynucleotides having the sequences of the reaction products were employed as model templates. These experiments allowed us to decipher the dynamic structure of the whole reaction system. We found that the hexamer and the pentamer are coupled autocatalysts: They behave both, as autocatalytic "egoists", and at the same time as mutually catalytic "altruists". Contrary, the tetramer MTM-CCGpnG, which is formed as the main product predominantly via its non-autocatalytic channel (efficient G-G stack), behaves as an isolated autocatalyst. In the whole reaction system, the coupled autocatalysts compete with the latter product for common precursors. Competition of replicators for common resources is the prerequisite for selection. Selection in the biological sense usually means the "takeover" of resources by a species which reproduces more efficiently than its competitors ("survival of the fittest"). However, as selection also depends on the population number of a species (its "concentration"), a less efficient species may win if it starts at a higher population level. In any case, the population number of a species directs the flow of resource consumption. This was indeed observed in the above reaction system: Upon "seeding" with hexamer molecules the production of the egoistic tetramer was substantially suppressed while the synthesis of both coupled autocatalysts increased. Although this result resembles selection, it is only a rudimentary form of selection: "True" (Darwinian) selection necessitates exponential and not parabolic growth.

Another set of experiments came from the objection that the natural prototype of nucleic acid replication utilizes complementary rather than self-complementary strands (15). Orgel's experiments had shown that only pyrimidine-rich oligonucleotides but not their purine-rich complements are efficient templates when starting from activated mononucleotides. It thus seemed worthwhile to test, whether or not a replication cycle could be achieved when using pairs of complementary templates instead of a single self-complementary sequence and oligomeric instead of monomeric precursors. We designed a system based on hexameric templates and trimeric precursors. The trimer sequences A (=CCG) and B (=CGG) were chosen such that each hexameric condensation product (AA, BB, AB, BA) would have a central GC subsequence. Each condensation lead to the formation of a 3'-5'-phosphoramidate bond. A minimal model of cross-catalytic replication of the complementary sequences AA and BB could thus be based on the following sequence of reactions:

$$A + A + BB \rightarrow A.A.BB \rightarrow A.AB.B \rightarrow AA + BB$$  \hspace{1cm} (1) \\
$$B + B + AA \rightarrow B.B.AA \rightarrow BB.AA \rightarrow BB + AA$$  \hspace{1cm} (2)

Irreversible ligation of template fragments takes place within the reversibly formed termolecular complexes (A.A.BB → AA.BB & B.B.AA → BB.AA). Reversible dissociation of the template duplex AA.BB gives the template molecules each of which can act cross-catalytically to continue the replication cycle. Under the experimental conditions the autocatalytic synthesis of self-complementary products AB & BA occurs simultaneously:
Kinetic studies revealed that both complementary sequences AA and BB are indeed formed with similar efficiency, although they differ in their pyrimidine-content. Moreover, complementary and self-complementary products exhibited the same time-course of synthesis, as long as all four products were synthesized simultaneously. Contrary, in experiments where only a single hexamer could form, the self-complementary products were synthesized much faster than the complementary ones. The difference is expected: A single self-complementary hexamer can autocatalytically accelerate its synthesis, whereas a single complementary hexamer cannot. Cross-catalysis necessitates the simultaneous formation of both complementary products and thus cannot take place in experiments in which only one product is synthesized. As to be expected, the cross-catalytic self-replication of complementary hexamers occurred with product inhibition leading to parabolic growth characteristics.

A novel scheme for the replication of complementary strands (Fig. 4) has been reported by Li and Nicolaou (16) based on the principles of triple-strand (17) (Watson-Crick-Hoogsteen) and double strand (Watson-crick) formation. At acidic pH (pH 6), the palindromic duplex DNA \( (S_1.S_2) \) serves as a template to attract complementary single-strand DNA fragments \( [F_1] \) and \( [F_2] \) (step (i)) and the two reactive ends come in proximity. N-cyanoimidazole \( (R) \) is used as the condensing agent (step (ii)). The newly formed triple helix \( (S_1.S_2.S_1^*) \) dissociates into the double strand \( (S_1.S_2) \) and a single strand \( (S_1^*) \) (step (iii)) when the pH is raised to 7. The newly formed single strand \( (S_1^*) \) serves as a template to attract complementary fragments \( [F_3^*] \) and \( [F_4] \) (step (iv)) and leads to the duplex \( (S_1^*.S_2*) \) in the presence of the condensing agent. The replication cycle is completed by readjusting the pH (pH 6) in which the original double helix is replicated at the expense of four half strands. This scheme of stepwise replication has the potential of overcoming the problem of product inhibition, although it is limited to palindromic sequences and to those in which purine and pyrimidine bases can not occur simultaneously in the same template strand.

![Fig. 4 Self-replication through triple-helix by Li and Nicolaou (16).](image-url)
REPLICATION IN SYNTHETIC MODEL SYSTEMS

Rebek and coworkers have demonstrated the self replication of a synthetic model system based on Kemp's triacid derivatives (Fig. 5) (18). The reactants 4 and 5 contain an activated carboxyl ester and a primary amine, respectively. When both are bound on the template 6, the reactive groups come into spatial proximity and the condensation leads to a new template 6. The source of autocatalysis observed in this system has been recently reinterpreted by Menger et al. (19). Evidence was given that simple amides catalyze the acylation reaction between 4 and 5 leading to 6. Since 6 also contains an amide moiety, it was argued that the rate enhancement is due to amide-autocatalysis and not due to template-autocatalysis. More recently, a pathway for complex-assisted amide autocatalysis has been proposed (19b). From the present stage of discussion it is however not clear, to what extent the observed autocatalysis in the above system originates from the presence of an amide moiety. We here suggest to replace the latter by an isosteric linkage such as a trans C=C or a trans C=N-moiet and to study the reaction between 4 and 5 in the presence of this model template. In any case it should be emphasized that the reaction kinetics of a later system, in which the template formed much stronger complexes, did reveal a square-root law supporting self-replication with product inhibition (20).

Another simpler self-replicating system has been developed (21) based on derivatives of (2-formylphenoxy) acetic acid 7 and 3-aminobenzamidine 8 (Fig. 6). Autocatalytic condensation of 7 and 8 (R1=R3=Me, R2=R4=NO2) takes place in DMSO as a solvent to form the anil 10. As expected, the condensation shows a square-root law for the autocatalytic contribution. When R1=Me, R2=NO2, R3= 'Bu and R4=H, however, the rate of the reaction depends linearly on the concentration of the template 9. Obviously, for whatever reasons, product inhibition is not operative here. Although the latter result originates from a catalytic rather than an autocatalytic effect, it suggests that exponential replication of non-nucleotide replicators might be achievable when using the proper system. A possible recipe is what one

Fig. 5 Artificial self-replicating system by Tjivikua et al. (18).

Fig. 6 Artificial self-replicating system of Terfort et al. (21).
may term the modulation of molecular recognition: to search for a template in which the linkage moiety affects the recognition properties of the binding sites, either sterically or electronically. Further exploration along these lines will tell the answer.

CONCLUSIONS AND OUTLOOK

For a minimal self-replicating system that operates as figured in Fig. 1, the rate of autocatalytic synthesis typically depends on the square-root of the template concentration and thus the template growth is of parabolic rather than of exponential nature. In a theoretical treatment Szathmáry and Gladkikh have shown that parabolic growth leads to the coexistence of self-replicating templates, which compete for common resources under stationary conditions (22). Coexistence means that the stronger replicator is not able to takeover the common resources completely. If two non-selfreplicating molecules, C1 and C2, compete for common precursors the selectivity which is a metric for coexistence is determined by reactivity: [C1]/[C2] = k1/k2. For parabolic replicators whose autocatalytic rate constants are k1 and k2, it follows from Szathmáry's treatment: [C1]/[C2] = (k1/k2)^2. Thus, small differences in reactivity lead to a higher selectivity as compared to non-selfreplicating molecules. This enhancement of selectivity which is partly implicit in the results reported, clearly needs a more elaborate experimental demonstration.

It was also stated by Szathmáry that the Darwinian kind of selection ("survival of the fittest") necessitates exponential growth of competing replicators. For equilibrated self replicating systems (meaning that the template-directed condensation is slow as compared to internal equilibration) the autocatalytic reaction order (which is 1 for the case of exponential growth and 1/2 for the case of parabolic growth) is determined by the population of the complexes involved (9). A minimal self-replicating system (Fig. 1) is expected to exhibit exponential growth if the termolecular complex is thermodynamically more stable than the template duplex. The usual situation, however, is just opposite due to entropic reasons. We see two possible ways to overcome the normal order of stabilities.

In the first approach, we plan to increase the thermodynamic stability of the termolecular complex. This approach employs what one may term a minimal replicase. A minimal replicase is conceived to act as a nucleophilic catalyst of the phosphoryl transfer step occurring in the termolecular complex (covalent catalysis). It is also conceived to possess a double-strand binding function. As long as the replicase is part of termolecular complex it stabilizes the latter, for example by wrapping itself around the complex. Stabilization occurs in an intramolecular sense. After phosphoryl-transfer the replicase can only bind to the resulting template duplex in an intermolecular sense. Thus, before and after the phosphoryl transfer, we are dealing with termolecular complexes here. A minimal replicase may be based on an oligonucleotide or a peptide possessing, for example, an imidazole moiety as the nucleophilic catalyst. To find a replicase molecule in a population of oligonucleotides we are currently employing the technique of directed molecular evolution (23). To find a peptide replicase, a strategy for combinatorial catalyst design might prove appropriate.

In the second approach, we are currently investigating methods to prevent the formation of duplexes between the template and its copy. This approach might ultimately lead to the non-enzymatic exponential replication of arbitrary sequences. Once exponential replication with information transfer has been achieved in the laboratory, synthetic molecular evolution mimicking Darwinian behavior is no longer a dream.

ACKNOWLEDGMENTS

Financial support by Deutsche Forschungsgemeinschaft, Fonds der chemischen Industrie, German-Israeli Foundation, NATO and Alexander von Humboldt foundation is gratefully acknowledged. G.v.K. would like to express thanks to a group of talented individuals whose contributions are cited in the references. Further thanks are devoted to those who are currently involved in related research: R. and T. Brandsch, J. Burmeister, M. Kindermann and A. Luther. A fruitful collaboration and bidirectional scientific exchange with the labs of Andy Ellington (Bloomington, USA) and Meir Lahav (Rehovot, Israel) is also acknowledged. B.G.B. thanks Alexander von Humboldt foundation for a fellowship.
REFERENCES


