NASA NCR- 97-

Progress Report and Request for Continued Support (2nd year)

for Cooperative Agreement NCC 2-534

Proposal Submitted to:	University Affairs Office NASA/Ames Research Center
Project Title:	Towards Self-Replicating Chemical Systems Based on Cytidylic and Guanylic Acids
Support Requested:	$ 99,840.00
Period:	January 15, 1998 through January 14, 1999
Submitted:	October 15, 1997
Submitted by:	The Regents of the University of California University of California, Santa Cruz
Institution Address:	Department of Chemistry and Biochemistry University of California, Santa Cruz Santa Cruz, CA 95064
Principal Investigator:	A Kanavarioti Dr. Anastassia Kanavarioti Associate Research Chemist (408) 459-2978
Co-Principal Investigator:	C. F. Bernasconi Dr. Claude F. Bernasconi Professor of Chemistry (408) 459-2035
Approved:	Karen Ann Fry Karen Ann Fry Contracts and Grants Officer (408) 459-4114 kafry@cats.ucsc.edu

Title: Towards Self-Replicating Chemical Systems Based on Cytidylic and Guanylic Acids

Progress Report for the Period January 15, 1997 to October 14, 1997

This project is aimed towards a better understanding of template-directed reactions and, based on this, towards the development of efficient non-enzymatic RNA replicating systems. These systems could serve as models for the prebiotic synthesis of an RNA world. The major objectives of this project were: (a) To elucidate the mechanistic aspects of template-directed (TD) chemistry, (b) to identify the conditions, environmental and other, that favor "organized chemistry" and stereoselective polymerization of nucleotides and (c) to search and, hopefully, find catalysts that will improve the efficiency of these reactions. Enhanced efficiency is expected to facilitate the road towards a self-replicating chemical system based on all four nucleic acid bases.

During the first nine months of the granting period from January 1997 to October 1997, we have made substantial progress towards the first two objectives. The research resulted in the publication of one paper (# 1; identified as in press in the final report of the last granting period), in the revision and submission of another manuscript (# 2; identified as submitted in the final report of the last granting period), in two manuscripts to be submitted shortly (# 3 and 4) and several manuscripts in preparation. A reprint of # 1 is enclosed and the abstracts of the manuscripts # 2-4 are included in the Appendix.

List of publications and manuscripts.

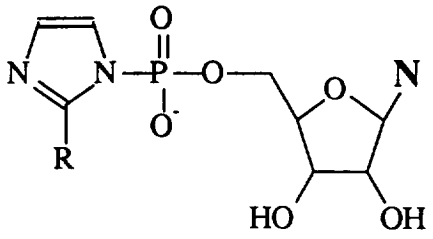
1. Kanavarioti, A., "Dimerization in Highly Concentrated Solutions of Phosphoimidazolide Activated Mononucleotides," *Origins Life Evol. Biosph.* **1997,** *27,* 357-376.
2. Kanavarioti, A., "Preference for Intenucleotide Linkages as a Function of the Number of Constituents in a Mixture," *J. Mol. Evol.*, submitted.
3. Kanavarioti, A., Bernasconi, C. F. and Baird, E. E., "Effects of Monomer and Template Concentration on the Kinetics of Non-enzymatic, Template-Directed Oligoguanylate Synthesis," manuscript in preparation.
4. Kanavarioti, A., Baird, E. E., Hurley, T. B. and Carruthers, J. A., "Poly(C)-Dependent Diguanylate Synthesis: Template-directed or Template-induced?" manuscript in preparation.

Summary of the research activities during the above period:

During this period our activities were directed towards (i) synthesizing activated nucleotides to be used as substrates, (ii) using these substrates in order to determine the effect of the leaving group (imidazole (Im), 2-methylimidazole (2-MeIm) and 2,4-dimethylimidazole (2,4-diMeIm)) in the product distribution, (iii) developing techniques for analysis of mixtures by LC/MS, (iv)

creating a protocol in order to obtain kinetic parameters of the dimerization reaction and (v) analyzing kinetic data obtained with the poly(C)/2-MeImpG system and writing the corresponding papers (# 3 and 4 in the above list). With the exception of item (v), the experimental work for the projects (i) - (iv) is still in progress. This work was primarily performed by Dr. Kanavarioti with the assistance of an undergraduate, Lynn F. Lee, and by consultation with Prof. Bernasconi of UCSC and Dr. Chang of NASA/Ames.

(i) We have synthesized, purified and characterized phosphoimidazolide activated nucleotides (see structure). All four 2-MeImpN and all four ImpN (with N=A, C, G, U for adenosine, cytidine, guanosine and uridine, respectively) have been prepared. We have also synthesized 2-MeImpdG, 2,4-diMeImpG and 2,4-diMeImpC. All the syntheses were done with the ribonucleotide, except in the case of 2-MeImpdG where the deoxyribonucleotide was used as starting material. The reasons for making the deoxy substrate, 2-MeImpdG, will become clear later. The 2,4-diMeImpG and 2,4-diMeImpC are new nucleotide derivatives, synthesized for the first time in order to introduce and test a new activating group that has the potential of being a better leaving group than 2-MeIm.



R = H, CH_3

N = Uracil, Cytosine, Adenine or Guanine

(ii) Phosphoimidazolide activated nucleotides, collectively abbreviated *pN, slowly hydrolyze in neutral aqueous solutions by cleavage of the P-N bond. Besides 5'NMP, other products include the three isomeric dimers, i.e. the pyrophosphate, N5'ppN, and the two internucleotide linked, pN2'pN and pN3'pN, the latter being the natural RNA dimer. We have established (# 1 in the publications list) that increasing initial concentration of *pN yields increasing yields of the dimerization products, as expected for a bimolecular reaction. We also established, as expected, that both 2-MeImpN and ImpN for a specific N yield the same products. However, we were surprised to find that the relative yield of the dimers varies dramatically with the leaving group, i.e. 2-MeIm or Im. Studies with 2,4-diMeIm are still at a preliminary stage. Specifically we find that 2-MeIm dramatically favors the formation of the internucleotide-linked dimers and disfavors the pyrophosphate dimer as compared with Im. Under the tested conditions, the observed ratio N5'ppN/ pN2'pN is approximately 0.8 with 2-MeIm and 7 with Im. This observation was shown to be valid with all four nucleobases. Considering that the ratio

pN2'pN/pN3'pN = 3 does not vary with leaving group or metal ion (Mg or Mn), it is inferred that Im is not as well suited for oligomerization as 2-MeIm. This information is important because most reactions aimed at testing minerals or metal ions as potential catalysts for nucleotide oligomerizations have been performed with Im and thus their full potential may not have been realized. In addition, our experiments confirm the observation made in template-directed reactions, namely that 2-MeImpN are better monomers for incorporation as compared with ImpN derivatives (see Orgel's work).

(iii) With a grant from the Keck foundation the UCSC Chemistry and Biochemistry Department has recently purchased a new electrospray LC/MS (liquid chromatography/ mass spectroscopy) facility. Consequently, a large part of our activities was devoted this year into breaking in the new equipment, running standards and developing protocols for its use with our reaction mixtures. In this respect, we had to modify the chromatography used so far and purchase a new column that is more appropriate for the LC/MS runs. We have been able to obtain molecular weight information by LC/MS for most of the substrates we made and for some selected reaction samples with a small number of constituents. Our goal is to be able to take any sample (a mixture of 10 to 15 reaction products) and obtain molecular weight information by LC/MS for every constituent of the sample that is more than a few percent by weight. We feel confident that next year we will be using the LC/MS routinely. Molecular weight information will facilitate greatly product identification and thus increase our productivity.

(iv) Until recently, kinetic studies of oligomerization reactions were scarce. We have pioneered this area in 1993 with the first kinetic study of the poly(C)-directed 2-MeImpG oligomerization. Since then, we and others have determined the kinetics of elongation in oligoguanylate and oligoadenylate synthesis (Ferris' work and Goebel's work). This year after establishing the dramatic differences in dimer distribution induced by the leaving group, we focused on the kinetics of the dimerization reaction. Although dimerization is, in principle, a less complex process than elongation, the determination of the dimerization kinetics present a real challenge. This is because dimerization is under all tested conditions a minor or side-reaction. For example, in the absence of a template, dimerization is outcompeted by hydrolysis and in the presence of the template or another appropriate catalyst, dimerization is outcompeted by elongation/oligomerization. In addition, dimerization is a second-order process, in contrast to hydrolysis and elongation which are or can be made to obey pseudo-first order conditions and therefore are easier to determine. The new methodology for obtaining dimerization rate constants was tested under a number of conditions with dimer formation in the reaction of morpholine-pG with 2-MeImpdG (dimer(s): morpholine-pG-pdG) and preliminary determinations were done in the self-condensation of 2-MeImpC (dimer(s): pCpC). The kinetics of the reaction of morpholine-pG

with 2-MeImpdG will be used as a control for 2-MeImpG dimerization in the presence/absence of poly(C). This is because in the presence of the template the former reaction does not lead to oligomerization products in contrast to the latter.

Research Plans for the 2ndYear:

The basic plan for this coming year is to continue the work in progress with the emphasis placed on the projects described below. The experiments will be performed by Dr. Kanavarioti and Lynn Lee using the facilities in Dr. Chang's laboratory at Ames and in Prof. Bernasconi's laboratory at UCSC.

(a) Kinetics of dimerization will be determined with a number of *pN derivatives, in order to determine the order of reaction, i.e. first vs. second order and to establish whether or not the predominance of stacking interactions in the concentrated solutions leads to dimer formation by a first-order process. In addition, it is of interest to establish the pH - rate profile in order to establish the nature of the nucleophile, i.e. anion vs. alcohol, etc. The kinetics will also be instrumental in determining whether or not 2,4-diMeImpN are "better" substrates, i.e. form more of the RNA dimer, than 2-MeImpN. Moreover, the kinetics will clearly define the best conditions for favoring RNA synthesis over the other possible pathways.

(b) In case 2,4-diMeIm moiety turns out to be a better leaving group than 2-MeIm we will attempt the synthesis of 2,4,5-triMeIm, the preparation and reaction of substrates activated by it.

(c) As described in detail in # 2 Abstract (see Appendix) we have determined product distribution in the reactions of activated nucleotides in up to 1 M concentrated aqueous solutions. We measured, among other products, the amount of RNA dimer(s) formed in the self-condensation, i.e. reaction with a single nucleotide, in reactions with binary and tertiary mixtures using 2-MeImpG, 2-MeImpU and 2-MeImpC as substrates. The surprising finding is a trend of approximately two-fold increase in the yield of internucleotide linked dimers, including the RNA one, as a function of the number of constituents in the mixture. In other words the sum of the percent yield found for RNA dimer(s) is the highest in the tertiary mixture (about 10 %) and the smallest in the self-condensations (about 4 %). Control experiments excluded the possibility that the effect is due to an added catalyst or a concentration or pH change. The observation of increased RNA synthesis with increasing number of constituents presents an example of chemical evolution and its validity should be tested with adenosine. Therefore, we plan on repeating the above experiments with 2-MeImpU, 2-MeImpC and 2-MeImpA instead of 2-MeImpG and on determining whether or not there is a similar trend with this combination of nucleotides. Only after such a trend is seen with adenosine, experiments with all four nucleotides can be initiated.

(d) In 1996 we initiated experiments with polyribonucleotides of mixed sequence, such as poly(C,G,A) and poly(C,U,A) acting as the template. We did that in order to get away from the reactions with homopolymers, such as poly(C), that are relatively easy to analyze and evaluate, but simplistic and, perhaps, of limited value. At this time, we are at advantage for performing such experiments because of the existence of the LC/MS facility which will facilitate the identification of products and because conditions have been found, i.e. the highly concentrated aqueous solutions, under which a substantial amount of activated dimers can be formed in situ and react in the presence of the template. It is plausible (see last year's proposal) that activated dimers will be better building blocks than monomers in the template-directed reactions. Preliminary experiments with poly(C,G,A) have shown that this polymer exhibits very poor templating capabilities, most likely because of its high content of G which creates substantial self-structure and diminishes the extent of single strand formation. The poly(C,U,A) will be tested next.

In conclusion, we would like to request the continuation of the contract NCC 2-534. We believe that the 2nd year will give us the opportunity to test one or two new leaving groups that have the potential to enhance oligomerization efficiency, to test the hypothesis that dimers are better building blocks than monomers and confirm or disprove the relationship between number of constituents and increased RNA yields which points out to a link between prebiotic chemistry and an RNA world. The latter would represent a substantial contribution to the field of nucleic acid chemistry and its implications for chemical evolution and the origin of life.

Participation in meetings:

As part of the activities during the above period of performance the PI (i) presented a seminar at the California State Polytechnic University, Pomona, on February 11, 1997 (ii) participated in the Astrobiology Workshop organized by both UCSC and NASA/Ames scientists which was held at UCSC on March 22, 1997, (iii) gave a talk entitled "Chemistry and the Origin of Life" on behalf of the UCSC Speakers Bureau to the Mid-County Exchange Club in Aptos, California, on June 4, 1997 and (iv) attended The RNA Stucture Symposium, held at UCSC on June 25-29, 1997.

In addition, the undergraduate student, Lynn F. Lee, who has been doing research under the supervision of the PI towards a senior thesis in Chemistry participated in the 9th Annual ACS Undergraduate Research Symposium in Chemistry on May 3, 1997, at UCSC and presented a paper entitled "Effect of Leaving Group on the Oligomerization of Phosphoimidazolide-activated Nucleotides."

Appendix

 # 2 Abstract : Phosphoimidazolide activated ribomononucleotides (*pN, see structure) are useful substrates for the non-enzymatic synthesis of oligonucleotides. In the presence of metal ions dilute neutral aqueous solutions of *pN (0.01 M) typically yield only small amounts of dimers and traces of oligomers; most of *pN hydrolyzes to yield nucleoside 5'-monophosphate (5'NMP). An earlier investigation of *pN reactions in highly concentrated aqueous solutions (up to 1.4 M) showed, as expected, that the percent yield of the condensation products increases and the yield of the hydrolysis product correspondingly decreases with *pN concentration (Kanavarioti, 1997). Here we report product distributions in reactions with one, two or three reactive components at the same total nucleotide concentration. *pN used as substrates were the nucleoside 5' phosphate 2-methylimidazolides, 2-MeImpN with N = cytidine (C), uridine (U) or guanosine (G). Reactions were conducted as self-condensations, i.e. one nucleotide only; with two components in the three binary U,C-, U,G- and C,G-mixtures and with three components in the ternary U,C,G-mixture. The products are 5'NMP, 5',5'-pyrophosphate-, 2',5'-, 3',5'-linked-dimers, cyclic dimers and a small percentage of longer oligomers. The surprising finding was that, under identical conditions, including the same total monomer concentration, the product distribution differs substantially from one reaction to another, most likely due to changing intermolecular interactions depending on the constituents. Even more unexpected was the observed trend according to which reactions of the U,C,G-mixture produce the highest yield of internucleotide-linked dimers, whereas the self-condensations produce the lowest yield and the reactions with the binary mixtures produce yields that fall in between. What is remarkable is that the approximately two-fold increase of the percent yield of internucleotide-linked dimers is not due to a concentration effect or a catalyst, but it is due to the increased complexity of the system from a single to two and three components. These observations, perhaps, provide an example of how increased complexity in relatively simple chemical systems leads to organization of the material and consequently to chemical evolution. A possible link between prebiotic chemistry and the postulated RNA world will be discussed.

 # 3. Abstract: In order to identify key parameters which influence the efficiency of non-enzymatic template-directed oligonucleotide synthesis, a kinetic study of oligoguanylate synthesis on a polycytidine (poly(C)) template has been performed. This is the first reported study which includes rate data as a function of the concentration of both poly(C) template and activated guanosine 5'-monophosphate-2-methylimidazolide (2-MeImpG) monomer. Rates determined in the range 2 mM $\leq$ [poly(C)] $\leq$ 50 mM and 5 mM $\leq$ [2-MeImpG] $\leq$ 50 mM support a mechanism of template-directed elongation of an oligonucleotide primer by a reacting monomer which is assisted by the presence of two additional downstream template bound 2-MeImpG molecules. These results provide new design principles for the optimization of non-enzymatic polymerizations.

 # 4. Abstract. Polycytidylate, poly(C), serves as a scaffold or template to direct and catalyze the synthesis of long oligoguanylates from guanosine 5'monophosphate 2-methylimidazolide, 2-MeImpG. A mechanistic model proposed to describe the template-directed oligoguanylate synthesis was recently successfully used in correlating kinetic data of the elongation process of dimers and longer oligomers in a wide range of monomer and template concentrations. In this study, we determined rates for dimer, pG3'pG, formation from 2-MeImpG in the absence and in the presence of poly(C) as a function of both monomer and template concentration at pH 8.0 and 23° C. It was found that the order of the dimerization reaction which is two in the absence of poly(C) changes to three in the presence of poly(C). A small, non-stoichiometric concentration of poly(C) expressed in cytidine equivalents, compared to 2-MeImpG, is responsible for the change in the order and evidently the mechanism of reaction. Based on these results, a poly(C)-induced model for dimerization is presented in addition to the poly(C)-directed model; the potential and limitations of both models are discussed.

DETAILED BUDGET: **1/15/98 - 1/14/99**

BUDGET CATEGORY	**AMOUNT**
A. SENIOR PERSONNEL	
Principal Investigator	
Anastassia Kanavarioti	
100% time 11 cal mos	61,865
Co-Investigator	
Claude Bernasconi	
5% time 1 sum mo	631
Assistant III	
Lynn F. Lee	
100% time 2 cal mos	2,106
TOTAL SALARIES AND WAGES	64,602
B. FRINGE BENEFITS	
Kanavarioti cal mos 11.8%	7,300
Bernasconi sum mos 2.9%	18
Lee cal mos 2.7%	57
TOTAL FRINGE BENEFITS	7,375
TOTAL SALARIES, WAGES, AND FRINGE BENEFITS	71,977
C. PERMANENT EQUIPMENT	0
D. TRAVEL	0
E. OTHER DIRECT COSTS	
1. Materials and Supplies	
a. Filters, microelectrodes, buffers, etc.	600
b. HPLC solvents (mainly acetonitrile), consumable replacement parts for the HPLC instrument include: 2 lamps for diode array detector, vials for analysis, etc.	2,137
c. Enzymes and nucleotides	936
d. 2 Guards and 1 HPLC column	538
Total Materials and Supplies	4,211
2. Publication Costs/Page Charges	300
3. Other	
a. Local/laboratory telephone	420
b. Long-distance telephone	300
c. Duplicating	300
d. Postage	70
e. Charges fur running analysis of samples by LC/MS/MS	425
f. Contract with Hewlett Packard for maintenance and servicing of 1090 LC located in Thimann 372 for half a year only	1,914

DETAILED BUDGET: **1/15/98 - 1/14/99**

BUDGET CATEGORY	**AMOUNT**
E. OTHER DIRECT COSTS (continued)	
3. Other (continued)	
g. Memberships and subscriptions to associated journals:	
- American Chemical Society	110
- American Association for the Advancement of Science	110
h. Subscription to Nature	120
Total Other	3,769
TOTAL OTHER DIRECT COSTS	8,280
F. TOTAL DIRECT COSTS	80,257
G. INDIRECT COSTS	
Indirect Cost Base	*80,257*
Off campus research rate of 24.4% of MTDC	19,583
TOTAL INDIRECT COSTS	19,583
H. TOTAL DIRECT AND INDIRECT COSTS	99,840